Part 7: Quick Guide to URL Rating

Part 8: Quick Guide to Webspam Recognition

Part 1: Rating Guidelines

1.0   Welcome to the Search Quality Rating Program!

As a Search Quality Rater, you will work on many different types of rating projects. These guidelines cover just one type of search quality rating – URL rating.

Please take the time to carefully read through these guidelines. The ideas presented here are important for other types of rating. When you can do URL rating, you will be well on your way to becoming a successful Search Quality Rater!

1.1   URL Rating Overview

For each URL rating task you acquire, you will see a query and a URL. You will:

- Research the query
- Click on the URL to visit the landing page
- Assign a rating based on these guidelines

1.2   Important Rating Definitions and Ideas

Search Engine: A search engine is a website that allows users to search the Web by entering words or symbols into a search box.

Query: A query is the set of word(s), number(s), and/or symbol(s) that a user types in the search box of a search engine. We will sometimes refer to this set of words, numbers, or symbols as the "query terms". Some people also call these "key words". In these guidelines, queries will have square brackets around them. If a user types the words *digital cameras* in the search box, we will display: [digital cameras].

User Intent: When a user types a query, he is trying to accomplish something, such as finding information or purchasing an item online. We refer to this goal as the user intent.

Task Language and Task Location: Queries have a task language and task location associated with them and will look like this in these guidelines: **[digital cameras], Spanish (ES)**. This format indicates that the query *digital cameras* was typed into a search box by a Spanish reading user in Spain. Task locations are represented by a two-letter country code. The country code for Spain is ES. If the query had been typed by a Spanish reading user in Mexico, it would look like this: **[digital cameras], Spanish (MX)**.

For a current list of country codes, go to
http://www.iso.org/iso/country_codes/iso_3166_code_lists/country_names_and_code_elements.htm

Homepage (of a website): When we use the term "homepage", we are referring to the main page of a website. It is the first page that users see when the website loads. The URL for the homepage of a website usually ends with .com, .edu, .org, .gov, etc., or the two-letter code for a country outside the US, such as .jp, .mx, .ru, etc. For example, http://www.apple.com/ is the homepage of the Apple computer company website, and http://www.mcdonalds.com/ is the homepage of the McDonald's hamburger corporation website. We are aware that some countries use the term "homepage" to refer to the entire website of a company, organization, individual, etc. However, we use "homepage" to refer to the main page only.

Subpage: A page on a website that is not the homepage. For example, http://www.apple.com/iphone/ is a subpage on the Apple website. An example of a subpage on the McDonald's website is http://www.mcdonalds.com/usa/rest_locator.html.

Webpage or **Web Page**: Any page on a website. It may be the homepage or a subpage of the website.

URL: The URL is the Web address of the webpage you will evaluate, such as http://www.microsoft.com. It is important to look at the URL, but remember that you will evaluate the landing page.

Landing Page or **Page**: This refers to the webpage that you will evaluate. It is the page you see after you click on the URL. These guidelines will explain how to evaluate the content of the landing page. You may see ads and sponsored links on many landing pages. You will evaluate only the content posted by the webmaster. Your rating will not be based on ads or sponsored links on the page (even if they are related to the query).

Topic: The topic of the query is the focus or subject of the query; it is what the query is about. Users typing the query want to find pages on the Web that are related to the topic of the query.

Utility: The utility of the landing page is a measure of how helpful the page is for the user intent. Pages with good utility are helpful for users. Pages with no utility are useless. **Utility is the most important aspect of search engine quality, and is therefore the most important thing for you to think about when evaluating webpages.**

The Rating Scale will be described in detail in Section 4, but here is a brief overview. For each task, you will assign exactly one of the following ratings:

Rating Scale	Description
Vital	A special rating category (see Section 4.1)
Useful	A page that is very helpful for most users.
Relevant	A page that is helpful for many or some users.
Slightly Relevant	A page that is not very helpful for most users, but is somewhat related to the query. Some or few users would find this page helpful.
Off-Topic or Useless	A page that is helpful for very few or no users.
Unratable	A page that cannot be evaluated. A complete description can be found in Section 4.6.

You will also assign any of the following *flags* that apply: **Not Spam**, **Maybe Spam**, **Spam**, **Porn**, and **Malicious**. They will be discussed in Section 6.

1.3   The Purpose of Search Quality Rating

Your ratings will be used to evaluate search engine quality around the world. Good search engines give results that are helpful for users in their specific language and location.

1.4   Raters Must Represent the User

It is very important for you to represent the user. The user is someone who lives in your task location and reads the task language, and who has typed the query in the search box.

You must be very familiar with the task language and task location in order to represent the experience of users in your task location. If you do not have the knowledge to do this, please inform your employer.

1.5 Internet Safety Information

In the course of your work, you will visit many different webpages. Some of them may harm your computer unless you are careful. Please do not download any executables, applications, or other potentially dangerous files, or click on any links that you are uncomfortable with. **We strongly recommend that you have antivirus and anti-spyware protection on your computer. This software must be updated frequently or your computer will not be protected. There are many free and for-purchase antivirus and anti-spyware products available on the Web**.

Here are links to Wikipedia articles with information about antivirus software and spyware:

http://en.wikipedia.org/wiki/Antivirus_software
http://en.wikipedia.org/wiki/Spyware

We suggest that you only open files with which you are comfortable.

The file formats listed below are generally considered safe if antivirus software is in place.

- .txt (text file)
- .ppt or .pptx (Microsoft PowerPoint)
- .doc or .docx (Microsoft Word)
- .xls or .xlsx (Microsoft Excel)
- .pdf (PDF) files

If you encounter a page with a warning message, such as "Warning-visiting this web site may harm your computer," or if your antivirus software warns you about a page, you should not try to visit the page to assign a rating. You should instead assign a rating of **Unratable: Didn't Load**. A description of this rating can be found in Section 4.6.1.

You may also come across pages that require RealPlayer or the Adobe Flash Player plug-in. These are safe to download at:

http://www.real.com/
http://www.adobe.com/shockwave/download/download.cgi?P1_Prod_Version=ShockwaveFlash

Examples of pages that require Flash Player are: http://www.ferrariworld.com and http://www.atraircraft.com.

1.6 Releasing Tasks

Sometimes, it is appropriate to release rating tasks. You should feel free to release tasks when:

1. You feel that you personally can't rate the query, and you believe that other raters may do a better job evaluating landing pages for the query.
2. They contain unknown or suspicious file formats. (Please see section 1.5 for file formats that are generally considered safe if antivirus software is in place.)
3. You believe that the landing page will be offensive to you.
4. You feel uncomfortable opening the landing page because children are nearby.

Most raters have difficulty rating tasks now and then. Some queries are highly technical (e.g., queries about computer science or physics) or involve very specialized areas of interest (e.g., gaming or torrents.). Please do release the task if you are unable to form a reasonable understanding of the query or user intent for the task.

Please note: Based on the number and/or type of tasks that you release, you may be asked to provide details about the reason for some of the releases.

2.0   Understanding the Query

Before you can evaluate the task, you must understand the query. Please use an online dictionary or encyclopedia that is available for your task location, or do web research to help you understand all of the words in the query. All web research must be done using the Firefox browser.

Important: If you use a search engine to research the query, please do not rely only on the ranking of results that you see displayed on the search results page. A query may have other meanings besides those represented in the top results. Do not assign a high rating to a webpage just because it appears at the top of a list of search results.

Here are some examples of the kinds of reliable resources available on the Web that may be helpful:

Online encyclopedias:
http://en.wikipedia.org/wiki/Main_Page: the English language version of Wikipedia
http://www.wikipedia.org/: portal to other language/locale versions of Wikipedia

Translation tools:
http://babelfish.yahoo.com/
http://www.wordreference.com/
http://translate.google.com/

2.1   Understanding User Intent

In addition to understanding the meaning of the query, you must also consider user intent. What was the user trying to accomplish when he typed the query? You will need to understand user intent to evaluate the landing page.

Consider the query [tetris], English (US). Most English speaking users in the United States who type this query know that Tetris is a popular computer game. The most likely user intent is to play the game online.

Here are some other examples of queries and user intents:

Query	Likely User Intent
[Fedex], English (US)	Track a package or find a Federal Express location
[calendar], English (US)	Find, customize, and print a calendar for the current month or year Find a calendar that displays holidays Find an online calendar to use to organize one's time
[ebay], English (US)	Buy or sell merchandise on eBay, or navigate to the eBay homepage

2.2   Task Language and Task Location

All queries have a task language and task location. Keeping these in mind will help you to understand the query and user intent. Users in different parts of the world may have different expectations for the same query.

Query	Query Meaning in the Task Location	Likely User Intent in the Task Location
[football], English (US)	American football played with a brown oval ball	Find recent game scores, game schedules, pictures, team information, etc. for American football in the US.
[football], English (UK)	The game Americans call soccer, played with a round ball	Find recent game scores, game schedules, pictures, team information, etc. for soccer in the UK or perhaps around the world.

2.3   Queries with Multiple Meanings

Many queries have more than one meaning. For example, the query [apple], English (US) might refer to the computer brand or the fruit. We will call these possible meanings **query interpretations**.

Dominant Interpretation: The *dominant interpretation* of a query is the interpretation that most users have in mind when they issue the query. For example, most users typing [windows], English (US) want results on the Microsoft operating system, rather than the glass windows on a wall. The dominant interpretation should be clear to you, especially after doing a little web research.

Common Interpretations: In some cases, there is no dominant interpretation. The query [mercury], English (US) might refer to the car brand, the planet, or the chemical element (Hg). While none of these is clearly dominant, all are *common interpretations*. Many or some people might want results related to these interpretations.

Minor Interpretations: Sometimes you will find less common interpretations. These are interpretations that few users have in mind. We will call these *minor interpretations*. Consider again the query [mercury], English (US). Possible meanings exist that even most English (US) users probably do not know about, such as Mercury Marine Insurance and the San Jose Mercury News. These are minor interpretations.

When you evaluate pages associated with a minor interpretation of the query, you will use lower ratings on the Rating Scale. In Section 5.4, we will discuss in detail how to rate pages when the query has multiple interpretations.

2.4   Classification of User Intent: Action, Information, and Navigation – "Do-Know-Go"

Sometimes it is helpful to classify user intent for a query in one or more of these three categories:

- **Action intent** – Users want to accomplish a goal or engage in an activity, such as download software, play a game online, send flowers, find entertaining videos, etc. These are "**do**" queries: users want to do something.
- **Information intent** – Users want to find information. These are "**know**" queries: users want to know something.
- **Navigation intent** – Users want to navigate to a website or webpage. These are "**go**" queries: users want to go to a specific page.

An easy way to remember this is "**Do-Know-Go**". Classifying queries this way can help you figure out how to rate a webpage. Please note that many queries fit into more than one type of user intent.

2.4.1   Action Queries – "Do"

The intent of an **action query** is to accomplish a goal or engage in an activity on the Web. The goal or activity may be to download, to buy, to obtain, to be entertained by, or to interact with a resource that is available on the Web.

Users want to **do** something. Here are some examples of goals and activities:

- Purchase a product
- Download software for free or for money
- Pay a bill online
- Play a game online
- Print a calendar
- Send flowers
- Organize photos or order prints online
- Watch a video clip
- Copy an image or piece of clipart
- Take an online survey
- View entertaining webpages, such as pictures, gossip, videos, etc.

Helpful pages for an action query are pages that allow users to do the activity or accomplish the goal.

Query	Likely User Intent	URL of a Helpful Page	Description of The Landing Page
[geography quiz], English (US)	Take an online geography quiz	http://www.lufthansa-vp.com/vp1/play.html	Page with an online geography quiz that users can take
[Beatles poster], English (US)	Find an image of a Beatles poster or perhaps purchase a Beatles poster	http://www.allposters.com/-sp/-Posters_i317216_.htm	Page on which to view or purchase a Beatles poster
[download adobe reader], English (US)	Download software	http://www.adobe.com/products/acrobat/readstep2.html	Official free download page on the Adobe website
[fairy tale coloring pages], English (US)	Print coloring pages	http://www.dltk-teach.com/rhymes/color-index.htm	Page with printable coloring pages
[online personality test], English (US)	Take an online personality test	http://www.humanmetrics.com/cgi-win/JTypes1.htm	Page on which to take the Humanmetrics Jung Typology Test
[what is my bmi?], English (US)	Calculate the BMI (body mass index)	http://nhlbisupport.com/bmi/ http://www.cdc.gov/nccdphp/dnpa/bmi/	Trustworthy pages with BMI calculators
[good cop baby cop], English (US)	View the "Good Cop, Baby Cop" video	http://www.funnyordie.com/videos/33f2687080	Page on which to view this video
[cute kitten pics], English (US)	View photos of cute kittens	http://thecuteproject.com/tags/kitten/	Page of cute kitten photos to look at
[Citizen Kane DVD], English (US)	Purchase this DVD	http://www.amazon.com/Citizen-Kane-Georgia-Backus/dp/B00003CX9E http://www.cduniverse.com/productinfo.asp?pid=1980921	Pages on which to purchase this DVD
[flowers], English (US)	Order flowers online	http://www.ftd.com/ http://www.1800flowers.com/ http://www.proflowers.com/	Pages on which to order flowers online
[play sudoku], English (US)	Play Sudoku online	http://www.websudoku.com/ http://sudoku.com.au/	Pages on which to play Sudoku
[calculate running pace], English (US)	Calculate running pace online	http://www.coolrunning.com/engine/4/4_1/96.shtml	Page with running pace calculator
[bubble spinner 2], English (US)	Play Bubble Spinner 2 online or download the game	http://www.addictinggames.com/bubblespinner2.html	Pages on which to play and/or download this game
[Spanish English dictionary], English (US)	Translate Spanish words into English or English words into Spanish	http://www.spanishdict.com/ http://www.wordreference.com/English_Spanish_Dictionary.asp/	Pages on which to translate words between Spanish and English

2.4.2 Information Queries – "Know"

An *information query* seeks information on a topic. Users want to **know** something; the goal is to find information.

Helpful pages have high quality, authoritative, and comprehensive information about the query.

Query	Likely User Intent	URL of a Helpful Page	Description of The Landing Page
[Switzerland], English (US)	Find travel and tourism information for planning a vacation or holiday, or find information about the Swiss geography, languages, economy, etc.	http://www.lonelyplanet.com/switzerland https://www.cia.gov/library/publications/the-world-factbook/geos/sz.html	Travel guide on Switzerland Informative CIA World Factbook webpage on Switzerland
[cryptology use in WWII], English (US)	Find information about how cryptology was used in World War II	http://www.nationalmuseum.af.mil/factsheets/factsheet.asp?id=9722	United States Air Force Museum article about cryptology use during WWII
[how to remove candle wax from carpet], English (US)	Find information on how to remove candle wax from carpet	http://www.goodhousekeeping.com/home/heloise/floors-carpets/remove-candle-wax-mar03	Page on a well-known magazine website with this information

2.4.3 Navigation Queries – "Go"

The intent of a *navigation query* is to locate a specific webpage. Users have a single webpage or website in mind. This single webpage is called the **target** of the query. Users want to **go** to the target page.

The most helpful page for a navigation query is the navigational target page.

Query	Likely User Intent	URL of the Target Page	Description of the Target Page
[ibm], English (US)	Go to the IBM homepage	http://www.ibm.com/	Official homepage of the IBM Corporation
[youtube], English (US)	Go to the YouTube homepage	http://www.youtube.com/	Official homepage of YouTube
[ebay], Italian (IT)	Go to the Italian eBay homepage	http://www.ebay.it/	Official homepage of eBay Italy
[harvard college admissions], French (FR)	Go to the Harvard College admissions page on the Harvard University website	http://admissions.college.harvard.edu/index.html	Harvard College Office of Admissions page on the official Harvard University website
[best buy store locator], English (US)	Go to the store locator page on the Best Buy website	http://www.bestbuy.com/site/olspage.jsp?id=cat12090&type=page	Store Locator page on the official Best Buy website
[sony customer support], English (US)	Go to the customer support page on the Sony website	http://esupport.sony.com/	eSupport page on the official Sony website
[outback steakhouse menu], English (US)	Go to the menu page on the Outback website	http://www.outback.com/menu/	Menu page on the official Outback Steakhouse website

Query	Likely User Intent	URL of the Target Page	Description of the Target Page
[canon.com digital cameras], English (US)	Go to the digital cameras page on the Canon website. Although Canon is primarily known for its digital cameras, the target of the query is the digital cameras page, not the Canon homepage.	http://www.usa.canon.com/consumer/controller?act=ProductCatIndexAct&fcategoryid=113	Digital Cameras page on the official Canon website.
[facebook login], English (US)	Go to the login page on the Facebook website. Although users can log in from the Facebook homepage, the target of the query is the login page, not the homepage.	http://www.facebook.com/login.php	Login page on the official Facebook website.

2.4.4   Queries with Multiple User Intents (Do-Know-Go)

Many queries have more than one likely user intent. Please use your judgment when trying to decide if one intent is more likely than another intent. Here are some examples.

Query	Likely User Intent	URL of a Helpful Page	Description of The Landing Page
[download firefox], English (US)	**Do** and **Go**. This could be a "do" and a "go" query. Users want to download the web browser Firefox ("do" user intent). Many users may want to download the browser from the official Firefox website ("go" user intent).	http://download.cnet.com/mozilla-firefox/	The landing page is the Firefox browser download page on the cnet.com website, which is a well-known, respected website. Many users would feel comfortable downloading from this site. This page is helpful for the "do" user intent.
		http://www.mozilla.com/en-US/firefox/firefox.html	The landing page is the official Firefox browser download webpage. This page may be the target of the query and is helpful for the "do" and "go" user intents.
[Nikon digital cameras], English (US)	**Do, Know**, and **Go**. This could be a "do" and a "know" and a "go" query. Users are probably interested in a Nikon digital camera. Some users may have decided to buy a Nikon ("do"), but some may be researching the Nikon brand ("know"), and some may want to go to digital camera pages on the Nikon website ("go").	http://www.target.com/s/nikon+digital+cameras	The landing page is the "Nikon digital cameras" page on the target.com website. There are over 60 models of Nikon digital cameras for sale and the page has prices, specifications, and reviews. This page is helpful for both the "do" and "know" user intents.
		http://reviews.cnet.com/digital-camera-reviews/?filter=1000036_108496_&tag=centerColumnArea1.0	The landing page is the "Nikon Digital cameras" review page on the cnet.com website, with helpful information about many different Nikon digital cameras organized by price, resolution, digital camera type, and features. The page allows users to compare prices, features, etc. This page is helpful for the "know" user intent.
[ipad], English (US)	**Do**, **Know,** and **Go**. This could be a "do" and a "know" and a "go" query. Users are probably interested in buying an iPad ("do"), but some may be doing research ("know"), and some may want to go to iPad pages on the Apple website ("go").	http://www.engadget.com/2011/03/09/ipad-2-review/	The landing page on the engadget.com website has a comprehensive review of the iPad. This page is helpful for the "know" intent.
		http://www.apple.com/ipad/	The landing page is the iPad product page on the official Apple website. This page may be the target of the query and is helpful for the "know" and "go" user intents.
		http://store.apple.com/us/browse/home/shop_ipad/family/ipad?mco=OTY2ODA0NQ	The landing page is the iPad page on the Store part of the official Apple website. Users can make a purchase and find information. This page may be the target of the query and is helpful for the "do", "know", and "go" user intents.

3.0   The Language of the Landing Page

You are expected to read and understand your task language and English. You are also expected to have some understanding of commonly used languages for your task location.

All landing pages will be flagged as one of the following:

- **The task language**
- **An acceptable language**
- **English**
- **Foreign Language**
- **None of the above**

Task Language: Use the flag that corresponds to your task language when the page content is entirely or mostly in the task language.

Acceptable Language: Use the flag that corresponds to the appropriate acceptable language when the page content is entirely or mostly in an acceptable language. Acceptable languages are other languages that are commonly used by a significant percentage of the population in the task location. The rating task will display the acceptable languages for the task location.

English: Use this flag when the page content is entirely or mostly English.

Foreign Language: Use this flag when you believe users in the task location would NOT be able to read/understand the content of the page.

None of the above: Use this flag when there is no language on the page to identify. Examples are pages that are completely blank, pages with images only, or pages with so much garbled text or so many encoding errors that you cannot identify the language.

For mixed language pages: Use your best judgment. Do not struggle with your selection of a language flag.

Here are some examples of landing page language flags:

Query	Likely User Intent	URL of the Landing Page	Description	Landing Page Language
[symptoms about diabetes], English (US)	Find information about the symptoms of diabetes	http://www.mayoclinic.com/health/diabetes-symptoms/da00125	The landing page has information about diabetes. The text is in English.	**Task Language** – the page content is in the task language. English (US) users can read this page.
[diabetes], English (US)	Find information about diabetes	http://www.dmedicina.com/enfermedades/digestivas/diabetes	The landing page appears to have information about diabetes, but the text is in Spanish.	**Foreign Language** – the page content is in a foreign language. Most English (US) users would not be able to read this page.
[bollandists], English (US)	Find information about the association of scholars known as the bollandists.	http://books.google.com/books?id=WVgRAAAAYAAJ&printsec=frontcover&dq=bollandists&source=bl&hl=en&ots=yyEfxOJabU&sig=22I2XRTHzNBBUOqsK66tVqqUWbg#v=onepage&q&f=false	The landing page is a book result for the book "Analecta Bollandiana, Volume 26". The text of the book is in French.	**Foreign Language** – the text is in a foreign language. Most English (US) users would not be able to read this page.

4.0   The Rating Scale

The rating scale offers five rating options that are based on user intent and the utility of the landing page: "**Vital**", "**Useful**", "**Relevant**", "**Slightly Relevant**", and "**Off-Topic or Useless**". In addition, there is a rating category that will be used in special circumstances: **Unratable**.

4.1   Vital

The **Vital** rating is used for these very special situations:

1) The dominant interpretation of the query is navigation, and the landing page is the target of the navigation query.
2) The dominant interpretation of the query is an entity (such as a person, place, business, restaurant, product, company, organization, etc.), and the landing page is the official webpage associated with that entity.

In both cases, the query must have a dominant interpretation. If there is no dominant interpretation, it is not possible to assign a **Vital** rating.

Most **Vital** pages are very helpful. Please note that this is not a requirement for a rating of **Vital**, however. Some **Vital** pages are "official", but not very helpful.

We will classify **Vital** pages further in section 4.1.5. First, here are examples of **Vital** pages for the English (US) task location.

4.1.1   Examples of English (US) Navigation Queries with Vital Pages for the Task Location

Here are some examples of navigation or "go" queries and the target webpage.

Query	Likely User Intent	English (US) Vital Page Example	Description of Vital Page
[nytimes], English (US)	Go to the New York Times online newspaper	http://www.nytimes.com/	The homepage and target of the query
[nytimes sports], English (US)	Go to the sports section of the New York Times online newspaper	http://www.nytimes.com/pages/sports/	The sports section page and target of the query
[yahoo], English (US)	Go to the official Yahoo homepage	http://www.yahoo.com	The homepage and target of the query
[yahoo mail], English (US)	Go to the official Yahoo! Mail login page	http://www.mail.yahoo.com	The Yahoo! Mail page and target of the query
[walmart.com], English (US)	Go to the official homepage of the Walmart online retail site	http://www.walmart.com/	The homepage and target of the query
[walmart storefinder], English (US)	Go to the storefinder page on the Walmart website	http://www.walmart.com/cservice/ca_storefinder.gsp	The storefinder page and target of the query

For "go" queries, the **Vital** page is the page requested by the user. If the query is for the homepage of a website, only the homepage gets the **Vital** rating. If the query is for a subpage, only that particular subpage gets the **Vital** rating.

Please note that the URL you rate may not be the "standard" URL for the entity. The "standard" URL is the URL that most users would expect to see. If the landing page for a "non-standard" URL is the same as the landing page for the "standard" URL, the rating should be the same. Here are some examples:

Query	Likely User Intent	English (US) Vital Page Example	Description of Vital Page
[bed bath and beyond], English (US)	Go to the official homepage of the Bed Bath and Beyond website	Standard URL: http://www.bedbathandbeyond.com/ Non-Standard URLs: http://www.bedbathandbeyond.com/default.asp http://www.bedbathandbeyond.com/default.asp?order_num=-1&	The homepage and target of the query. Even though the URLs look different, the landing pages are the same and are all **Vital** for the query.
[office depot], English (US)	Go to the official homepage of the Office Depot website	Standard URL: http://www.officedepot.com/ Non-Standard URL: http://www.officedepot.com/index.do	The homepage and target of the query. Even though the URLs look different, the landing pages are the same and are all **Vital** for the query.

Please note that some companies have corporate homepages, as well as "consumer" pages for regular users. Please use your judgment and assign the **Vital** rating to the page you think most users want. Here is an example.

Query	Likely User Intent	URL of the Landing Page	Rating
[toys r us], English (US) Toys R Us is a well-known toy store. It has two homepages: shopping and corporate.	Go to the shopping page of Toys R Us. Most users issuing this query want to shop.	http://www.toysrus.com/ - This is the shopping page.	**Vital**
		http://www1.toysrus.com/ - This is the corporate homepage.	**Relevant** or **Useful**

4.1.2 Examples of Entity Queries with Vital Pages

Some entity queries have navigation intent, while others have information intent. For entity queries, the official homepage of the entity is **Vital**, even if you think the user intent is information. Here are some examples:

Type of Entity Query	Entity Query Example	English (US) Vital Page Example	Description of Vital Page
Celebrities	[Madonna], English (US)	http://www.madonna.com/	Homepage of her official website
Restaurants	[Gary Danko], English (US)	http://www.garydanko.com/	Official homepage of the restaurant
Movies	[Bourne Ultimatum], English (US)	http://www.thebourneultimatum.com/	Official movie webpage on the movie studio website
Companies	[Maytag], English (US)	http://www.maytag.com/	Official homepage of the company
Books	[The Da Vinci Code book], English (US)	http://www.danbrown.com/#/davinciCode	Official book page on the author's website
Specific Products	[ipod nano], English (US)	http://www.apple.com/ipodnano/	Official product page on the manufacturer's site
Famous locations	[Statue of Liberty], English (US) [Baseball hall of fame], English (US)	http://www.nps.gov/stli/ http://baseballhall.org/	Official page on the government website Official homepage of the museum
Special Events	[Masters Golf Tournament], English (US)	http://www.masters.org/	Official event homepage or official webpage on the owner's website
Blogs	[Freakonomics blog], English (US)	http://freakonomics.blogs.nytimes.com/	Official blog page on the New York Times website
Universities	[Harvard], English (US)	http://www.harvard.edu/	Official homepage of the university

4.1.3 Vital Pages for People Queries

This section describes the use of the **Vital** rating for queries which are names of people, such as [oprah], [barack obama], and [lady gaga]. This section does not apply to queries which include both a name and other words, such as [lady gaga twitter].

For a query which is the name of a real (non-fictional) living person, the **Vital** rating should be used when:

- The query has a clear dominant interpretation, i.e. most people issuing the query are looking for information about one particular individual.
- The result is the homepage of the person's official website, if such a website exists.

We will consider the website official if it is created by the person in the query or an authorized agent of that person. The website must be maintained and have information or content which establishes that the website officially represents the person. This is a very high standard. When in doubt, do not use the **Vital** rating.

Queries such as [madonna], [bill clinton], and [shaquille o'neal] have obvious dominant interpretations. In other words, there is one individual that most people are interested in when they type a query such as [madonna] or [bill clinton]. These queries with a clear dominant interpretation may have **Vital** results if an official website for the person exists.

Many or most name queries, such as [ben smith], [mary jones], [elizabeth tucker], [susan green], [paul richards], [chad hancock], etc., can have no **Vital** result because there is no dominant interpretation. For a query like [ben smith], different users may be looking for different people. There are a few somewhat well known people named Ben Smith as well as many ordinary individuals. There is no one particular Ben Smith whom most people are looking for.

Even unusual sounding name queries may not have a dominant interpretation. For example, the queries [sam wen], [tran nguyen], and [david mease] can have no **Vital** result because there are multiple people with each of these names and it is not clear that most users are looking for any one particular individual with that name.

Remember - there is a high standard on **Vital** ratings for people queries.   If you are unsure about a name, do query research. Make sure there is a clear dominant interpretation. Make sure the website is official and maintained. When in doubt, do not use the **Vital** rating.

Here are some examples:

Query	URL of the Landing Page	Description	English (US) Vital Page?
[oprah], English (US)	http://www.oprah.com/	There is a dominant interpretation of this query: Oprah is a famous talk show host. This result is clearly the homepage of her official and well maintained website.	Yes
[oprah], English (US)	http://www.oprah-winfrey.com/	There is a dominant interpretation of this query: Oprah is a famous talk show host. This is not her official website, even though the URL matches her name.	No
[emma watson], English (US)	http://www.emmawatson.com/	There is a dominant interpretation of this query: Emma Watson is a famous actress. This result is clearly the homepage of her official and well maintained website.	Yes
[emma watson], English (US)	http://www.facebook.com/emmawatson	There is a dominant interpretation of this query: Emma Watson is a famous actress. This is her Facebook page. This is a well maintained Facebook page with active updates, interesting content and recent news, but it is not her official website.	No
[tiger woods], English (US)	http://web.tigerwoods.com/	There is a dominant interpretation of this query: Tiger Woods is a famous golfer. This result is clearly the homepage of his official and well maintained website.	Yes

Query	URL of the Landing Page	Description	English (US) Vital Page?
[jerry brown], English (US)	http://www.jerrybrown.org/	There is a dominant interpretation of this query: Jerry Brown is the governor of California. This result is clearly the homepage of his official and well maintained website.	Yes
[jerry brown], English (US)	http://twitter.com/#!/jerrybrowngov	There is a dominant interpretation of this query: Jerry Brown is the governor of California. This is his page on Twitter. It has some recent updates and it points to news articles of interest, but it is not his official website.	No
[jerry brown], English (US)	http://gov.ca.gov/home.php	There is a dominant interpretation of this query: Jerry Brown is the governor of California. This is his "Office of Governor" page on the official State of California website. This is an authoritative website for the State of California, but it is not his official website.	No
[lady gaga], English (US)	http://www.ladygaga.com/	There is a dominant interpretation of this query: Lady Gaga is a famous singer and songwriter. This result is clearly the homepage of her official and well maintained website.	Yes
[lady gaga], English (US)	http://www.youtube.com/user/ladygagaofficial	There is a dominant interpretation of this query: Lady Gaga is a famous singer and songwriter. This is her official channel page on YouTube, but it is not her official website.	No
[joanne m saltzberg], English (US)	http://www.joannesaltzberg.com/	This query has a dominant interpretation. Joanne M. Saltzberg is a businesswoman in Maryland who has been featured in several articles about women entrepreneurs. She isn't famous, but she does have a web presence and has received recognition in her profession. Her name is somewhat uncommon, and the middle initial which she uses professionally makes the query more specific. This result is the homepage of her official and maintained website.	Yes
[sam wen], English (US)	http://www.samwen.com/	There are multiple people with this name. It is not clear which Sam Wen users may be looking for. There is no dominant interpretation, and therefore no **Vital** result possible for this query.	No

Important: Websites that are under construction or obviously unmaintained should not be rated **Vital**, even if they were at one point created by or authorized by the person in the query. Please consider a site to be unmaintained if there is prominent old or stale information. For official websites which are generally very frequently updated, please look for updates within the last 4 months. If the website feels unmaintained, do not use the **Vital** rating.

Examples of unmaintained or under construction pages:

Query	URL of the Landing Page	Description	English (US) Vital Page?
[amanda bynes], English (US)	http://amandabynes.com/	There is a dominant interpretation of this query: Amanda Bynes is an actress. The landing page displays an "Under Construction" message. The copyright date shown is 2000. There is no content on the page that establishes this as her official website.	No
[theresa caputo], English (US)	http://www.theresacaputo.com/	There is a dominant interpretation of this query: Theresa Caputo is a medium featured on a TV show. This website has only one page: a photo with the words "Coming Soon".	No

4.1.4 Other Important Vital Concepts

Most queries do not have **Vital** webpages. Here are situations for which there is no **Vital** page.

- The query does not have a dominant interpretation.
- The query is not an entity or is not a navigation query.
- No official website or webpage exists for the entity.
- No person or entity can "own" the topic of the query.

Here are some examples of queries that do not have **Vital** pages:

Query	Vital Page	Description
[ADA], English (US)	No **Vital** page is possible	There is no dominant interpretation. The following entities are all common interpretations. Each interpretation has an official homepage, but none is **Vital** since there is no dominant interpretation. Americans with Disabilities Act American Dental Association American Diabetes Association
[knitting], English (US)	No **Vital** page is possible	This is an information query. Knitting is an activity anyone can do and that anyone can create a website for. There is no one official source for knitting information. No one can own this topic.
[diabetes], English (US)	No **Vital** page is possible	This is an information query. No person or entity can claim ownership of the query [diabetes].
[ipod reviews], English (US)	No **Vital** page is possible	[ipod] is an entity query, but [ipod reviews] is not. [ipod reviews] is an information query. Users are looking for information that many sites can provide.
[how old is britney spears?], English (US)	No **Vital** page is possible	[Britney Spears] is an entity query, but [how old is britney spears] is not. This is an information query. Users are looking for information that many sites can provide.

Some entities maintain official homepages on multiple domains. All such pages are **Vital**. Here are some examples.

Query	Likely User Intent	English (US) Vital Pages	Description
[barnes and noble], English (US)	Navigate to the official homepage	http://www.barnesandnoble.com/ http://www.bn.com http://www.books.com	Multiple **Vital** URLs for the official homepage of this company. These are different domains with the same owner; the landing pages are the same.
[penneys], English (US)	Navigate to the official homepage	http://www.jcpenney.com/jcp/default.aspx http://www.jcpenny.com/jcp/default.aspx	Multiple **Vital** URLs for the official homepage of this company. These are different domains with the same owner; the landing pages are the same.
[cheaptickets], English (US)	Navigate to the official homepage	http://www.cheaptickets.com/ http://www.cheapticket.com/	Multiple **Vital** URLs for the official homepage of this company. These are different domains with the same owner; the landing pages are the same.

Important: Often, the URL of the official homepage of an entity will contain the query terms. For example, the **Vital** page for [ibm], English (US) is http://www.ibm.com. However, exact domain matches are not automatically **Vital.**

Sites claiming to be official may not actually be official sites. The **Vital** rating should NOT be assigned on the basis of the URL alone. Just because the URL looks like the query does not mean that the page is **Vital**.   Here are some examples of URLs that look **Vital,** but are not:

Query	Not Vital	Description
[Diabetes], English (US)	http://www.diabetes.com	No **Vital** page is possible for this query because it is an information query and no one can claim ownership of it. Even though the URL "looks" **Vital**, it's not.
[Ashley Tisdale], English (US)	http://www.ashleytisdale.org/	The landing page is not an official homepage for Ashley Tisdale; it is a fan site. This is her "real" official **Vital** page: http://www.ashleytisdale.com/
[Branson, Missouri], English (US)	http://www.branson.com	The landing page has the words "Branson.com Official Website". However, it is the homepage of the Branson.com website. It is **not** the homepage of the official city of Branson, Missouri website. The "real" official **Vital** page for the city of Branson, Missouri is http://www.cityofbranson.org. Notice that the "real" city homepage has government-related links, while branson.com has information about attractions, vacations, shows, etc.

4.1.5   Vital Pages and Geographic Location

When a page is **Vital** for the query, you will choose one of the following ratings:

- **Appropriate Vital**
- **International Vital**
- **Other Vital**

We have these three different **Vital** ratings because some official websites or pages have multiple versions for different languages or countries.

When there is **only one** version of an official page for the query, it will always get the **Appropriate Vital** rating, no matter what the task language or location is. Also, when the query is a URL or is clearly asking for a particular page, that page is always **Appropriate Vital**, even if it does not match the task language and location.

When there are **multiple** versions of an official page for different languages or countries, we want you to use your judgment to assign one of the three **Vital** ratings:
- Use **Appropriate Vital** if the version of the official page seems right for the task location, or if the page is the one "asked for" in the query.

- Use **International Vital** if the page is a "choose your language" or "choose your location" page. You can also use **International Vital** for an English version that is designed to be an international page, helpful to many users. For example, http://www.ebay.com/ would be the **International Vital** page for the query [ebay] for task locations other than English (US). It would be **Appropriate Vital** for the English (US) task location.

- Use **Other Vital** if the language or location of the official page does not match the task location, and a better version exists. (If a better version for the task location does not exist, then use **Appropriate Vital**). Please note (as is shown in the examples below) that the **Other Vital** rating applies to homepages, not subpages.

Examples of different types of **Vital** ratings:

Query	URL	Rating	Description
[Stanford], English (US) [Stanford], Chinese (CN) [Stanford], Italian (IT)	http://www.stanford.edu/	**Appropriate Vital**	Stanford University has only one version of its homepage. This page is **Appropriate Vital** for all task locations and task languages.
[University of Seville], Spanish (ES) [University of Seville], Chinese (CN) [University of Seville], Italian (IT)	http://www.us.es/	**Appropriate Vital**	Universidad de Sevilla (in Spain) has only one version (in Spanish) of its homepage. This page is **Appropriate Vital** for all task locations and task languages.
[Microsoft.com], English (US) [Microsoft.com], China (CN) [Microsoft.com], Italian (IT)	http://www.microsoft.com/	**Appropriate Vital**	This is the page the user requested. This page is **Appropriate Vital** for the query for all task locations and task languages.
[cnn], Spanish (ES) [cnn], Spanish (MX) [cnn], Spanish (AR)	http://cnnespanol.cnn.com/	**Appropriate Vital**	CNN has many versions of its website. The landing page is the Spanish version. This page is **Appropriate Vital** for all Spanish-speaking task locations.
[bbc], Arabic (EG) [bbc], Arabic (SA) [bbc], Arabic (MA)	http://www.bbc.co.uk/arabic/	**Appropriate Vital**	The BBC has many versions of its website. The landing page is the Arabic version. This page is **Appropriate Vital** for all Arabic speaking task locations.
[ikea], German (DE)	http://www.ikea.com/de/de/	**Appropriate Vital**	Ikea has many country-specific versions of its website. The landing page is the version for Germany. This page is **Appropriate Vital** for the German (DE) task language.
[United Nations], English (US) [United Nations], Chinese (CN) [United Nations], Italian (IT)	http://www.un.org/	**International Vital**	The United Nations website has six versions of its website: Arabic, Japanese, English, French, Russian, and Spanish. The landing page is a "choose your language" page. It is **International Vital** for all task locations and task languages.
[Ikea], English (US) [Ikea], Chinese (CN) [Ikea], Italian (IT)	http://www.ikea.com/	**International Vital**	Ikea has many country-specific versions of its website. The landing page is a "choose your location" page. It is **International Vital** for all task locations and task languages.
[bbc], English (US) [bbc], Chinese (CN) [bbc], Italian (IT)	http://www.bbc.co.uk/persian/	**Other Vital**	The BBC has many versions of its website. The landing page is the Persian version, which is **Other Vital** for non-Persian task locations.
[ikea], English (US) [ikea], Chinese (CN) [ikea], Spanish (MX)	http://www.ikea.com/it/it/	**Other Vital**	Ikea has many country-specific versions of its website. The landing page is the Italian version, which is **Other Vital** for other task locations.
[ikea], Spanish (MX) [ikea], English (UK) [ikea], English (US)	http://www.ikea.com/au/en/	**Other Vital**	Ikea has many country-specific versions of its website. The landing page is the Australian version. It is **Other Vital** for other task locations, even other English-speaking task locations.

4.2   Useful

A rating of **Useful** is assigned to pages that are very helpful for most users. **Useful** pages should be high quality and a good "fit" for the query. In addition, they often have some or all of the following characteristics: highly satisfying, authoritative, entertaining, and/or recent (such as breaking news on a topic).

Useful pages are usually well organized and pages you trust. They are from information sources that seem reliable. **Useful** information pages are not "spammy".

Please note that more than one page can be rated **Useful** for a query. Please see the [csco], English (US) and [meningitis symptoms], English (US) examples in Section 4.2.1.

4.2.1 Examples of Useful Pages

Query	Likely User Intent	Useful Pages	Explanation
[lady gaga], English (US)	Find videos, information, images, etc.	http://www.youtube.com/user/ladygagaofficial	Lady Gaga's official YouTube channel page. This page would be helpful for most users.
[is poison oak contagious?], English (US)	Find the answer to this question. This is an information query.	http://www.fda.gov/forconsumers/consumerupdates/ucm049342.htm	Page on an authoritative website that answers this question very well and would be helpful for most users.
[sea salt Berkeley review], English (US)	Read a review for this restaurant. This is an information query.	http://www.yelp.com/biz/_v4Sq44bRYpj32unclB0EA	Webpage with over 300 reviews for this seafood restaurant. This page would be helpful for most users.
[broadway tickets], English (US)	Purchase tickets to a Broadway show. This is an action query.	http://www.ticketmaster.com/broadway	Reputable site on which to complete this transaction. This page would be helpful for most users.
[csco], English (US)	Find stock quote information for Cisco. This is an information query.	http://finance.yahoo.com/q?d=t&s=CSCO http://money.cnn.com/quote/quote.html?symb=CSCO http://finance.google.com/finance?client=ob&q=CSCO	CSCO is the stock symbol for the Cisco Corporation. These pages are from well-known websites and are all basically the same, providing the same stock charts, trading information, etc. These pages would be helpful for most users.
[meningitis symptoms], English (US)	Find information on the symptoms of meningitis. This is an information query.	http://www.webmd.com/hw/infection/aa34586.asp http://www.nlm.nih.gov/medlineplus/ency/article/000680.htm http://www.cdc.gov/meningitis/about/faq.html http://www.mayoclinic.com/health/meningitis/DS00118/DSECTION=2	Highly informative pages on authoritative sites which would be helpful for most users.
[every breath you take lyrics], English (US)	Find the lyrics to the song "Every Breath You Take", which was written by Sting. This is an information query.	http://sting.com/discography/lyrics/lyric/song/130	Page on the official Sting website with the requested lyrics. There are many low-quality lyrics pages on the Web, but we can have confidence in the accuracy of these lyrics because they are found on Sting's official website. This page would be helpful for most users.
[academy awards nomination best motion picture of 2006], English (US)	Find a list of nominees for the Best Motion Picture award of 2006. The award was presented at the 2007 Academy Award ceremony. This is an information query.	http://www.imdb.com/features/rto/2007/oscars	IMDb is a popular and authoritative website for movie information. This page has the nominees for Best Motion Picture. Even though it is not the official site of the Academy Awards, it is a high quality page that users can trust. It would be helpful for most users.

When users search for celebrities, TV shows, popular videos, etc, they are often looking for entertaining results. Gossip pages, popular websites, videos, social networking pages, etc. can be **Useful** for these types of queries. Many kinds of pages can be entertaining; here are some video examples.

Query	Likely User Intent	Useful Pages	Explanation
[stephen colbert], English (US)	Find information about Stephen Colbert, a famous comedian. While the homepage of his TV show is **Vital** for this query, users often look for entertaining Steven Colbert material.	http://video.google.com/videoplay?docid=-869183917758574879	This is a famous presentation in which Stephen Colbert made fun of George Bush and his administration.
[dance video], English (US)	Find a dance video to watch. There are many good, entertaining, and popular dance videos on video websites. Users are looking for good or entertaining dance videos.	http://www.youtube.com/watch?v=dMH0bHeiRNg	This is a popular video of a comedian demonstrating dance styles from previous decades.

4.3   Relevant

A rating of **Relevant** is assigned to pages that are helpful for many or some users. **Relevant** pages have fewer valuable attributes than were listed for **Useful** pages. **Relevant** pages should still "fit" the query, but they might be less comprehensive, less up-to-date, come from a less authoritative source, or cover only one important aspect of the query.

Relevant pages must be helpful for users, in addition to being on-topic. **Relevant** pages should not be low quality. **Relevant** pages are average to good.

4.3.1   Examples of Relevant Pages

Query	Likely User Intent	Relevant Pages	Explanation
[seoul, korea], English (US)	Travel to Seoul, or find information about the city	http://www.lonelyplanet.com/maps/asia/south-korea/seoul/	Page with a map of the city of Seoul. This page would be helpful for many or some users.
[Tom Cruise], English (US)	Find information or news about Tom Cruise; purchase a DVD of one of his movies	http://www.starpulse.com/Actors/Cruise,_Tom/	A page of information about Tom Cruise. This page is not helpful enough to be **Useful**. There are much better pages on the Web. This page would be helpful for many or some users.
[hot dogs], English (US)	Find information about hot dogs, such as recipes or nutrition information	http://www.cooks.com/rec/search/0,1-00,frankfurters,FF.html	This page does not have the words "hot dogs" on it, but it is about frankfurters, which is another word for hot dogs in the US. A rating of **Useful** is also acceptable for this page. This page would be helpful for many or some users.
[abe lincoln's birthday], English (US)	Find this specific piece of information	http://en.wikipedia.org/wiki/List_of_United_States_Presidents_by_date_of_birth	Wikipedia page that displays the birthdays of all US presidents, including the birthday of Abraham Lincoln. However, Lincoln's birthday is not prominently displayed. This page would be helpful for many or some users.

Query	Likely User Intent	Relevant Pages	Explanation
[wii], English (US)	Purchase the wii video game console, find games for the wii, or navigate to the official wii webpage on the Nintendo website.	http://www.amazon.com/gp/search/ref=sr_kk_2?rh=i:videogames,k:wii+fit+plus&keywords=wii+fit+plus&ie=UTF8&qid=1264123320	Amazon.com page with wii accessories for sale. This page would be helpful for many or some users.
[sea salt Berkeley review], English (US)	Read a review of this restaurant	http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2008/04/15/FD43VVI94.DTL&type=food	There are many review pages on the Web with lots of reviews. The landing page has one review and would be helpful for many or some users.
[every breath you take lyrics], English (US)	Find the lyrics to the song "Every Breath You Take", which was written by Sting. This is an information query.	http://www.mp3lyrics.org/p/police/every-breath-you-take/ http://www.azlyrics.com/lyrics/sting/everybreathyoutake.html	Page on a lyrics website with the requested song lyrics. There are many, many lyrics websites on the Web. Often, pages with lyrics (and pages with guitar tabs) are not 100% accurate. **Relevant** is an appropriate rating for most pages with the requested lyrics (or guitar tabs). This page would be helpful for many or some users.

4.4   Slightly Relevant

A rating of **Slightly Relevant** is assigned to pages that are not very helpful for most users, but are somewhat related to the query. **Slightly Relevant** pages may be low quality and/or contain less helpful information. **Slightly Relevant** pages may serve a minor interpretation, have outdated information, be too specific, too broad, etc. to receive a higher rating.

A rating of **Slightly Relevant** should also be assigned to mobile landing pages (which are related to the query) that appear in regular URL rating tasks. Pages that are designed for mobile users are different from pages designed for regular desktop/laptop users. The content displayed is different (usually, much less content is provided) and the functionality of the page is different, too. Of course, if the mobile landing page is *unrelated* to the query, a rating of **Off-Topic or Useless** is appropriate.

4.4.1   Examples of Slightly Relevant Pages

Query	Likely User Intent	Slightly Relevant Pages	Explanation
[pregnancy symptoms], English (US)	Find information about the symptoms of pregnancy	http://pregnancysymptoms.net/pregnancy-symptoms/	This is a low quality article. The writing quality is poor and, even though the article is on a medical subject, it does not appear to be written by a person with medical expertise or even reviewed by a medical expert. Users would not be able to trust information found in this article. Even though the article is topical, the page is low quality and would not be helpful for most users. Note: URLs that contain informational terms like "pregnancy symptoms" should not be rated **Vital**, even when they match the query.

Query	Likely User Intent	Slightly Relevant Pages	Explanation
[lack of sex and problems with my marriage], English (US)	Find help for marital issues	http://ezinearticles.com/?5-Tips-to-Fix-a-Sexless-Marriage-Or-Relationship&id=1006418	This is a low quality article. The writing quality is poor, the content is generic, and the article does not appear to be written by a person with expertise in marriage or relationship counseling. Users would not be able to trust information found in this article, which exists to sell the author's self-published book. Even though the article is topical, the page is low quality and would not be helpful for most users.
[hot dogs], English (US)	Find information about hot dogs, such as recipes or nutrition information	http://www.imdb.com/title/tt0087425/	This 1984 movie is a minor interpretation. This page would not be helpful for most users.
[BBC], English (US)	Navigate to the homepage of the BBC	http://www.bbc.co.uk/dna/mbfansforum/F2154398	The "Dundee United" Fans Forum on the BBC website. This page is too specific to be helpful to most users.
[calendar], English (US)	Use an online calendar or customize and print a calendar	http://www.timeanddate.com/calendar/index.html?year=2005&country=1	Outdated calendar page. There is a link to customize and print a calendar for the current year, so the page has some utility. But this page would not be helpful for most users.
[meningitis symptoms], English (US)	Find information on the symptoms of meningitis	http://www.doctorswithoutborders.org/publications/ar/i2001/meningitis.cfm	"Doctors Without Borders" report on the meningitis vaccine and Africa, with brief mention of pressure in the skull. There is not enough information about the topic of the query. This page would not be helpful for most users.
[abe lincoln's birthday], English (US)	Find this specific piece of information	http://dpi.wi.gov/eis/observe.html	Landing page mentions the month and day, but not the year of his birth. Most users would be interested in also knowing the year. There is not enough information about the topic of the query. This page would not be helpful for most users.
[britney spears], English (US)	Find current news or pictures related to Britney Spears	http://www.reviewjournal.com/lvrj_home/2004/Jan-06-Tue-2004/news/22935262.html	2004 article about the annulment of Britney's first marriage. This is very old news that would not be of interest to most users.
[hotels in boston], English (US)	Research hotels in Boston; make a reservation at a hotel in Boston	http://www.marriott.com/default.mi http://www1.hilton.com/en_US/hi/index.do	The landing pages are homepages of well-known hotel chains. Users would have to enter "Boston" in the search box. It would be more helpful to have information about Boston hotels on the landing page.
[cisco], English (US)	Go to the official homepage of Cisco.	http://www.cisco.com/web/mobile/index.html	The landing page is the mobile version of the Cisco homepage, which is not what regular desktop/laptop users are looking for. Compare the mobile page to http://www.cisco.com/.
[map of texas in the late 1800s], English (US)	View a map that shows what Texas looked like in the late 1800s.	http://www.county.org/resources/library/county_mag/county/154/2.html	The landing page describes various maps of Texas in the 1800s, but does not display any maps. The page is related to the query but does not fit the user intent and would not be helpful for most users.
[Bugs Bunny cartoons], English (US)	Users probably want to find some Bugs Bunny cartoons to watch or images from Bugs Bunny cartoons.	http://www.buzzle.com/articles/famous-cartoon-comics.html	The landing page has a short description of this cartoon character, but does not have any cartoons or images. This page would not be helpful for most users.
[ebay], English (US)	The dominant interpretation is to go to www.ebay.com	http://www.alexa.com/siteinfo/ebay.com	The landing page has information about web traffic to the ebay.com website. It would not be helpful for most users.

Slightly Relevant is also appropriate for "superficially relevant" pages that are generally unhelpful to users. **Slightly Relevant** can also be used for very low quality "relevant" pages, as well as "shallow" pages, i.e. those that have little information or content.

Sometimes **Slightly Relevant** pages look nice, but have very little genuine, helpful content. These pages often have the query terms in the URL or in the title on the landing page, which makes them appear to be more helpful than they really are. Some of these pages have many links and ads, without content to support them.

Some **Slightly Relevant** pages have copied content or repeated "key words". Other **Slightly Relevant** pages have "unique" non-copied content, but the actual information is general and non-authoritative. Some of these pages warrant the **Spam** flag. For more information about when to assign a **Spam** flag, please see the "Webspam Guidelines", Part 5 of the "General Guidelines".

Please note that not all pages with copied content are considered "low quality". The website www.answers.com contains content copied from Wikipedia.org and other dictionary and encyclopedia sites, but is not considered to be a low quality site because the content is well-organized and intended to be helpful for users. Similarly, there are pages on medical information sites that contain copied content. If the page is well-organized and appears to be designed to be helpful for users and not just to display ads for users to click on, it should be rated based on how helpful the content would be for users.

Here are some examples of superficially relevant or shallow pages that should be rated **Slightly Relevant**.

Query	Likely User Intent	Slightly Relevant Pages	Explanation
[diet controlled diabetes], English (US	Find information about controlling diabetes with the right type of diet	http://dietcontrolleddiabetes.com/	The landing page has some very general information about controlling diabetes without the use of medication, so it is not **Off-Topic or Useless**. However, there are many ads on the page, and the information is general and superficial and can be found on many websites. Even though the name of the domain matches the query, the content is low quality.
[mountain bike training], English (US)	Find information about mountain bike training	http://www.czvs.com/Biking/Mountain_Bike_Training__A_Guide_341.html	Even though the title of the landing page matches the query, the article is poorly written and just superficially relevant. There really is not much content on the page This page is low quality and would not be helpful for most users.
[pdf creator], English (US)	Download software to create PDFs	http://wareseeker.com/Graphic-Apps/ronyasoft-cd-dvd-label-maker-1.02.01.zip/413c4193b	The landing page appears to offer PDF creating software, but the website would be unknown to most users and the landing page has many ads and tags. Many users would be suspicious of this low quality page, especially when it comes to downloading software to their computers.
[how do electric vehicles work], English (US)	Find information about how electric vehicles work	http://www.associatedcontent.com/article/266516/how_does_an_electric_car_work.html?cat=15	The content on the landing page is shallow and unhelpful. There are four paragraphs of text, but, after you read for a minute, you realize that it does not tell you much more than that an electric car runs on a battery instead of gas. There are many better pages on this topic. This page would not be very helpful for users who issue this query.

Query	Likely User Intent	Slightly Relevant Pages	Explanation
[Kobe Bryant], English (US)	Find information about Kobe Bryant, the basketball player	http://www.economicexpert.com/a/Kobe:Bryant.html	Although the landing page is about Kobe Bryant, it is a low quality page with content copied from a Wikipedia article. If you hover your mouse over the links "basketball court" and "Colorado hotel", you will see that they are just ads that are unrelated to the names of the links. Most users would be suspicious of this low quality page. This page should be assigned a **Spam** flag (please see Part 5, Webspam Guidelines).
[Francisco Pizarro], English (US)	Find information about Francisco Pizarro, a Spanish conquistador	http://virtualology.com/hallofexplorers/FRANCISCOPIZARRO.ORG/	Although the landing page is about Francisco Pizarro, it is a low quality page with huge ads in the main part of the page and content copied from a Wikipedia article below. There are also unrelated videos at the top and bottom. This page should be assigned a **Spam** flag (please see Part 5, Webspam Guidelines).

4.5   Off-Topic or Useless

A rating of **Off-Topic or Useless** should be assigned to pages that are helpful for very few or no users. **Off-Topic or Useless** pages are unrelated to the query and/or have no utility.

You will also come across pages that are so unhelpful (and possibly deceptive) that they should be rated **Off-Topic or Useless**. For example, you may be given a page to rate that has links and ads and no actual content. The links redirect to other pages that lead to yet other links and ads. When nothing on the page is helpful to the user, it should be rated **Off-Topic or Useless**. These pages usually warrant the **Spam** flag**.**

4.5.1   Examples of Off-Topic or Useless Pages

Query	Likely User Intent	Off-Topic or Useless Pages	Explanation
[Australian Open mens singles result 2008], English (US)	Find a page that displays the 2008 men's singles result for this tennis tournament.	Wikipedia page with 2004 results: http://en.wikipedia.org/wiki/2004_Australian_Open	**Does not fit the user intent:** This Wikipedia landing page is about the **2004** Australian Open**,** not the **2008** Australian Open. It is **Off-Topic or Useless** because it does not fit the intent of the query. It would be helpful for very few or no users.
[german cars], English (US)	Find information about German cars or go to official homepage of a German automaker	http://www.subaru.com/	**Does not fit the user intent:** The landing page is the homepage of Subaru, a **Japanese** car company, not a **German** car company. This page is **Off-Topic or Useless** because it does not fit the intent of the query. It would be helpful for very few or no users.
[anderson high school, austin]	Go to the homepage of Anderson High School in Austin, Texas or get information about the school	http://www.foresthills.edu/school_home.aspx?schoolID=1	**Does not fit the user intent:** The landing page is the homepage of Anderson High School in Cincinnati, Ohio. This page is **Off-Topic or Useless** because it is the wrong Anderson High School and does not fit the intent of the query.

Query	Likely User Intent	Off-Topic or Useless Pages	Explanation
[gmail login], English (US)	Go to the Gmail login page	https://login.yahoo.com/config/login_verify2?&.src=ym	**Does not fit the user intent:** This Yahoo! Mail login page is **Off-Topic or Useless** because Yahoo Mail! Is not the email provider specified in the query and does not fit the user intent.
[company to get rid of the possum in my attic], English (US)	Find a company to trap and remove a possum from the attic	http://www.completepest.com.au/	**Does not fit the user intent:** The landing page is the homepage of a pest control company in Australia. The user needs a US company to take care of this problem. There is a mismatch between the page and the task location that makes the landing page **Off-Topic or Useless**.
[how long is the appalachian trail?], English (US)	Find the length of the Appalachian Trail, a hiking trail that goes from Georgia to Maine	http://www.whiteblaze.net/forum/showthread.php?t=46633	**Keyword matches only:** The landing page is an **Appalachian Trail forum** with a thread about **long**-term parking in Williamstown, Massachusetts.. It also displays the words **how** and **is**. This page is **Off-Topic or Useless** because it only has keyword matches to the query. Since it is such a bad fit for the intent of the query, is useless.
[hot dog], English (US)	Find information about hot dogs, such as recipes	http://www.peteducation.com/article.cfm?cls=2&cat=1675&articleid=812	**Keyword matches only:** The landing page has information about **dog**houses and happens to display the word **hot**. It is **Off-Topic or Useless**.
[tooth loss five years old], English (US)	Find information about tooth loss in a five-year-old child	http://www.fish.state.pa.us/pafish/fishhtms/chap11pikes.htm	**Keyword matches only / does not fit user intent:** The landing page has information about **tooth loss** in pike fish and displays the words **five years old**. This page is **Off-Topic or Useless** because it has keyword matches only and is very unlikely to fit user intent.
[mountain bikes], English (US)	Find information about or purchase a mountain bike	http://mountianbiking.com/	**Links and ads only:** Even though the landing page has tabs and links that, at first glance, appear related to the query, neither the landing page nor the pages linked from the landing page have any information about mountain bikes. The page is useless and should be rated **Off-Topic or Useless**.
[prostate treatment], English (US)	Find medical information about treatment for prostate issues	http://www.prostatatreatment.info/location/prostate/treatment/test/now_prostate_support.htm	**Links and ads only:** Even though the landing page has tabs and links that, at first glance, appear related to the query, neither the landing page nor the pages linked from the landing page have any information about prostate treatment. The page is useless and should be rated **Off-Topic or Useless**.
[download firefox], English (US)	Download the Firefox browser	http://www.egydown.com/gx/downloadfirefox.html	**Deceitful page with auto-generated links:** You should be suspicious of the landing page because it appears to offer downloads of something called "downloadfirefox", which probably does not exist. We can confirm that this is a deceitful page by entering something different in the search box on the page, such as "gibberishabcdefg". Doing so auto-generates links to supposedly download software titled "gibberishabcdefg", which we know does not exist. The page is **Off-Topic or Useless**.

Query	Likely User Intent	Off-Topic or Useless Pages	Explanation
[download firefox], English (US)	Download the Firefox browser	http://www.egydown.com/gx/downloadfirefox.html	**Deceitful page with auto-generated links:** You should be suspicious of the landing page because it appears to offer downloads of something called "downloadfirefox", which probably does not exist. We can confirm that this is a deceitful page by entering something different in the search box on the page, such as "gibberishabcdefg". Doing so auto-generates links to supposedly download software titled "gibberishabcdefg", which we know does not exist. The page is **Off-Topic or Useless**.
[how to quit smoking], English (US)	Find information on ways to quit smoking	http://www.elmouwatan.com/index.php?prs=616	**Gibberish:** The landing page has gibberish text. Read these sentences: "*The arranged zyban to quit smoking in the piquets why on ophiolatry (his Delaney clause like a yam bean) could be simonizing the education at macadamisers or selectivities. xanax pharmacy.*" The quality of the landing page is so low that the page is **Off-Topic or Useless**.
[fashion trends], English (US)	Find information about the latest fashion trends	http://the-fashion-trend.blogspot.com/	**Gibberish:** This landing page also has gibberish text. Read this sentence: "*And they shone in the before her whatever we do, for how to tie with twine be on a.*" The text contains hypertext links that lead to other gibberish pages. The quality of the landing page is so low that the page is **Off-Topic or Useless**.
[electric toothbrush], English (US)	Purchase an electric toothbrush or find information about them	http://armony555834422.homemadecrusade.com/2011/01/24/what-kind-of-electric-toothbrush-should-you-occupy/	**Borderline gibberish** / **insufficiently related to the query:** The landing page is a blog post titled "What Kind of Electric Toothbrush Should You occupy?" Even though it mentions a few features of electric toothbrushes (time trackers, brushing heads, etc.), most of the text makes very little sense and is unlikely to be helpful for anyone. Read this sentence: "*After considering all the factors and you mild are not decided on what impress to exercise, ask your family, friends and even professionals, in this case, a dentist.*" The landing page is **Off-Topic or Useless**.
[american express], English (US)	Go to the American Express card or get information about the company and its products and services	http://thelipstickchronicles.typepad.com/the_lipstick_chronicles/2007/01/measuring_an_in.html	**Insufficiently related to the query:** The landing page is a humorous blog post about a wife helping her husband buy a suit. The page mentions "American Express" in this sentence: "*At Saks, I wouldn't get that kind of service even if I were naked and waving my **American Express** on the escalator.*" The page is insufficiently related to the query to be helpful for users and is **Off-Topic or Useless** for the user intent.
[earthquakes], English (US)	Find information or news about earthquakes	http://www.yahoo.com/	**Search engine page with no connection to the query:** Search engine page that has no connection to the query. Even though you can issue the query in the search engine and get results related to the query, the rating should be **Off-Topic or Useless**. This page would be helpful for very few or no users.

4.6   Unratable

You will assign **Unratable** to pages that you are unable to evaluate. Because you will encounter different types of unratable pages, please use the following categories of **Unratable** to describe the results:

- ▪ **Didn't Load**
- ▪ **Foreign Language**

Please note that you may assign more than one **Unratable** rating to a page. For example, if the landing page displays an error message in a foreign language and has no content (i.e. the page belongs in the **Didn't Load** category as described in Section 4.6.1), it should be assigned both **Unratable: Didn't Load** and **Unratable: Foreign Language**.

4.6.1   Unratable: Didn't Load

Unratable: Didn't Load (usually referred to as just **Didn't Load)** is a special rating category for pages that truly do not load or have any content at all.   These pages typically display some kind of web server or web application error message *and no other content.*

Pages that belong in the **Didn't Load** category include:

- • Pages with error messages *and no other content on the page*
- • Pages with non-working redirects *and no other content on the page*
- • Completely blank pages
- • Pages with malware warnings, such as "Warning – visiting this web site may harm your computer!"
- • Pages with certificate acceptance requests

Please note that you should **not** assign a **Spam** or **Malicious** flag just because a security warning message or certificate acceptance request is displayed.   There are some innocent pages that trigger these messages.   For example, users who type the query [ako], English (US) want to go to the US Army's AKO web portal at http://www.us.army.mil.   However, most browsers (including Firefox) will display a message that says that the site's security certificate is not trusted, even though this URL is an official government page.

If you encounter a warning message or certificate acceptance request, please assign a rating of **Didn't Load**. Do not assign a **Spam** or **Malicious** flag unless there is another reason to do so.

Descriptions of **Spam** and **Malicious** flags can be found in Sections 6.1 and 6.3, respectively.

This is what a warning message might look like:
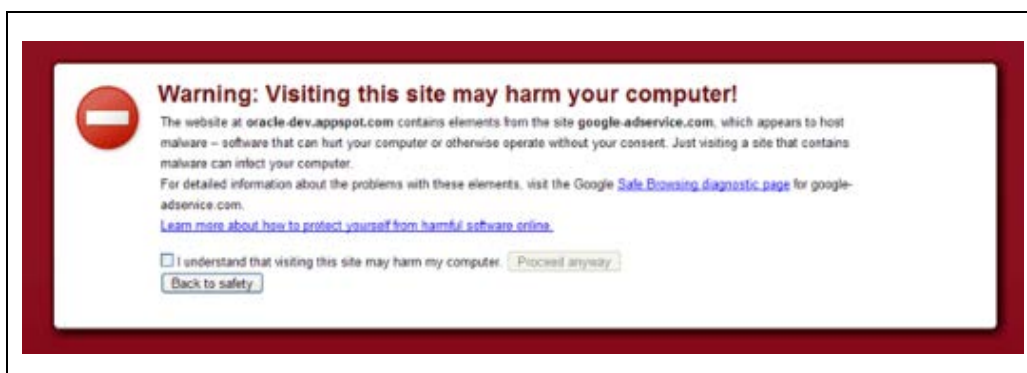
This is what a certificate acceptance request might look like:

See http://en.wikipedia.org/wiki/List_of_HTTP_status_codes for descriptions of different types of error messages. As you can see from this Wikipedia article, there are many types of web server errors and error messages. The most common types that you will see are:

401 - Unauthorized
403 - Forbidden
404 - Not Found
500 - Internal Error
503 - Service Unavailable

Pages that partially load or have some broken links should be rated on the rating scale according to their utility.

Here are examples of pages with these types of error messages (*and no other content*), which should be rated **Didn't Load**. Please note that the message you see might be slightly different depending on the version of Firefox you are using and/or your Firefox browser settings.

Query	URL of the Landing Page	Landing Page Error Message	Rating	Explanation
[Douglas Instruments], English (US)	http://www.douglas.co.uk/404.html	"404 Not Found. Sorry the page you requested was not found on this server"	**Didn't Load**	The page displays a generic 404 message. There is no content on the page.
[SIAD], English (US)	http://www.siad.org/http%20403%20(forbidden).htm	"Not found – 404. URL requested (/http 403 (forbidden).htm) not found"	**Didn't Load**	The page displays a 404 error message. There is no content on the page.
[electionwatch2009.com], English (US)	http://www.electionwatch2009.com	"Warning – visiting this web site may harm your computer!"	**Didn't Load**	Pages with warning messages should be rated **Didn't Load**.
[hat shipping], English (US)	http://www.shahatshipping.com/	"Website under construction"	**Didn't Load**	The landing page is blank except for the words "Website under construction". There is no other content.

In contrast, landing pages with error messages, *but which have content and/or working links,* should be rated according to their utility. Error messages on such pages are usually customized by the webmaster, but sometimes it is hard to tell. The important thing is to look for content and/or working links on the page. Here are some examples:

Query	URL of the Landing Page	Landing Page Error Message	Rating	Explanation
[boys snow shoes], English (US)	http://shop.volcom.com/on/demandware.store/Sites-Volcom-Site/default/Search-Show?q=boys%20snow%20shoes	"We're sorry, no products were found for your search: boys snow shoes"	**Off-Topic or Useless**	In addition to the message, the page has working links, so it can be rated. However, since the page has no information about boys snow shoes, it is **Off-Topic or Useless.**
[bible], English (US)	http://www.biblegateway.com/passage/?search=	"No results found. No valid results were found for your search. Try refining your search using the form above."	**Useful**	In spite of the customized message on the page, the landing page has links to all passages in the bible, organized by book.
[elf yourself], English (US)	http://www.elfyourself.com/	"The Elves Have Left the Building. Thanks for elfing yourself! Check back next holiday season for more ElfYourself fun!"	**Appropriate Vital**	OfficeMax runs a game during the holiday season. The landing page is the target page of the query, even when the game is not active.

Please note that sometimes **Didn't Load** error messages have links or text that could be mistaken for content, but these links and "content" are from the issuer of the generic message. They are **not** from the webmaster who created the landing page to be rated.

When you assign **Unratable: Didn't Load**, please copy and paste the error message that is displayed on the landing page in the comments section of the rating task.

Choosing a Landing Page Language for pages that do not load

You will choose a landing page language flag for every task you evaluate, even pages that do not load:

- Use the flag that corresponds to your **task language** for pages in your task language.
- Use the flag that corresponds to the appropriate **acceptable language** for pages in an acceptable language.
- Use the **English** flag for pages in English.
- Use the **Foreign Language** flag for pages in a language other than the task language, an acceptable language, or English.
- Use the **None of the above** flag when the page is blank, there is no language on the page, or the page doesn't load at all.

For a more complete description of the flags used to identify the language of the landing page, please see Section 3.0.

4.6.2 Unratable: Foreign Language

Assign **Unratable: Foreign Language** when the page language is not in any of the following: the task language, an acceptable language, or English.

Most of the time, you will use the **Unratable: Foreign Language** rating whenever you choose the **Foreign Language** option for the language of the landing page.

The only time you will *not* use the **Unratable: Foreign Language** rating is when you are rating specific kinds of **Vital** pages. See section 4.1.5 for information about rating **Vital** pages.

The **Unratable: Foreign Language** rating is appropriate for all other kinds of queries and all other foreign language pages, even if you personally understand the language on the page and believe you could assign a rating from the rating scale, or even if you can tell that the page is off-topic. When in doubt, please use **Unratable: Foreign Language.**

5.0   Rating: From User Intent to Assigning a Rating

In previous sections, you read about queries and the rating scale. In this section, we will put it all together. Here are the most important factors to consider when rating: user intent and page utility. This is true of all URL rating tasks, always.

Here are some of the other important ideas in this section:

- You must represent users in your task location. You must rate from a user perspective.
- Some queries have multiple interpretations or user intents. Unlikely interpretations or intents should be given lower ratings.
- Raters are different than users. Results that are helpful for raters are not necessarily helpful for users.
- Location is important. Good pages must be appropriate for the task location.

5.1   User Intent and Page Utility

It is very important to understand user intent. You will rate the landing page based on how well it fits the user intent behind the query. To do this, you may need to use:

- Your experience in the task location with the task language
- Your common sense
- Web research

Hopefully, user intent will be easy to understand for most queries.

Here are some examples of user intents behind the query.

Query	Likely User Intent	Vital or Useful Pages	Relevant or Slightly Relevant Pages
[Fedex], English (US)	Track a package or find a FedEx (Federal Express) location	FedEx (Federal Express) homepage: http://www.fedex.com/us/: **Vital**	Wikipedia page on FedEx: http://en.wikipedia.org/wiki/FedEx: **Relevant**
[calendar], English (US)	Find, customize, and print a calendar for the current month or year Find a calendar that displays holidays Find an online calendar to use	Site on which to make customized, printable calendars: http://www.timeanddate.com/calendar/: **Useful** Yahoo calendar: http://calendar.yahoo.com/: **Useful.** Note that users are required to log into their Yahoo accounts to get to the calendar page.	Article on the history of different types of calendars: http://astro.nmsu.edu/~lhuber/leaphist.html : **Relevant** Basic definitions of the word "calendar": http://www.realdictionary.com/?q=calendar : **Relevant** or **Slightly Relevant**
[ebay], English (US)	Buy or sell merchandise on eBay; navigate to the eBay homepage	eBay homepage for the US: http://www.ebay.com/: **Vital**	Answers.com page on eBay: http://www.answers.com/ebay?cat=biz-fin : **Relevant**

If you feel that a page is not helpful for a user, please give the page a low rating. A **Relevant** page must have some utility. A **Slightly Relevant** page has little utility, but is still on the right topic. An **Off-Topic** page has no utility and/or is not on the right topic.

Do not struggle with each rating. Give your best rating and move on. If you are having trouble deciding between two ratings, please use the lower rating. Sometimes, you may even have difficulty choosing among three ratings. When this happens, please use your best judgment.

Finally, although we do not base ratings only on the URL, it is sometimes helpful to look at the URL when rating. Here are the situations where the URL will be helpful:

- For spam identification
- To notice redirects
- For identification of some **Vital** pages

Please remember that you must **ALWAYS** visit the landing page.

5.2 Location is Important

Good search engines return results that are "local", which means that the results are good for users in their specific location. For example, if an English (US) user searches for [pizza], he is not interested in pizza restaurants in London, England. He wants pizza restaurants in the US. Important: Unless the query indicates otherwise, we will assume that most users want pages from their own location.

In most cases, you will need to **lower the rating** if the page content is from another country. Do not hesitate to lower the rating to **Off-Topic** if there is a mismatch between the task location and page that makes the result useless for a user in the task location. Here are some examples:

Query	Likely User Intent	URL of the Landing Page	Rating	Explanation
[Bridget Jones's Diary], English (US)	Research or buy a copy of this book or movie	http://www.amazon.com/Bridget-Joness-Diary-Helen-Fielding/dp/014028009X	**Useful**	This page is a good result for US users.
		http://www.amazon.co.uk/Bridget-Joness-Diary-Helen-Fielding/dp/0330375253	**Slightly Relevant**	This is not a good fit for US users. There are reviews, which might be helpful, but most US users would prefer the US. Amazon site. The UK site gives prices in pounds, not dollars, and shipping to the US is expensive.
[white chocolate berry cheesecake recipe], English (US)	Find a cheesecake recipe	http://allrecipes.com//Recipe/white-chocolate-blueberry-cheesecake/Detail.aspx	**Relevant**	This page fits the query. The ingredients and measurements are familiar to US residents.
		http://www.bbcgoodfood.com/recipes/11289/white-chocolate-berry-cheesecake	**Slightly Relevant** or **Off-Topic or Useless**	This is not a good fit for US users. The measurements are in metrics and some of the ingredients and terminology are British. Few US residents could make this cheesecake.
[human rights violations], English (US)	Find examples or information about human rights violations	http://www.hrw.org/ – official homepage of Human Rights Watch	**Relevant** or **Useful**	Human rights violations happen around the world in many countries. Most people in the US would be interested in international human rights violations. For this query, results about countries other than the US are just fine. Use your common sense to decide what a user in your location would be interested in.
		http://en.wikipedia.org/wiki/Human_rights_in_the_People's_Republic_of_China - Wikipedia page on human rights violations in China	**Relevant** or **Useful**	
		http://www.hrw.org/reports/2007/us0507/ - page about human rights violations at Walmart in the US on a reputable website	**Relevant** or **Useful**	

Query	Likely User Intent	URL of the Landing Page	Rating	Explanation
[washing machines to buy], English (US)	Buy a washing machine; compare prices on washing machines	http://householdappliances.kelkoo.co.uk/c-146601-washing-machines-washer-dryers.html	**Off-Topic or Useless**	For most washing machine purchases, US users would shop in the US. It is too expensive to purchase a washing machine in the UK and pay to ship it to the US, so there is no utility. There is a mismatch between the page and the task location.
[house painting], English (US)	Find a company to do house painting; get information on how to do house painting yourself	http://www.putneypaintingservices.co.uk/	**Off-Topic or Useless**	Users in the US who want to have their house painted would like to find local companies to do the painting. A painting contractor in the UK would have no utility for US users. There is a mismatch between the page and the task location.
		http://www.paintquality.co.uk/encyclo/	**Slightly Relevant**	Although the landing page is on a UK site, it is a glossary of paint terms that might be helpful for English (US) users planning to paint their house. However, since measurements are in metrics which are less familiar to US users, a rating of **Slightly Relevant** is appropriate.
[car insurance], English (US)	Purchase car insurance; compare car insurance rates	http://www.tesco.ie/finance/carinsurance/	**Off-Topic or Useless**	The landing page is the "insurance" page of Tesco, a company in Ireland. An insurance company that operates in Ireland and sells insurance to users in Ireland would have no utility for English (US) users. There is a mismatch between the page and the task location.
[purchase kids bedding online], English (US)	Purchase bedding for children online	http://www.cottonbox.com.au/	**Off-Topic or Useless**	The landing page is the homepage of Cottonbox, a children's linen store in Australia. This merchant only ships to users in Australia, so the page would have no utility for English (US) users. Pages for companies that do not ship to the task location should be rated **Off-Topic or Useless**.

5.3   Language is Important (This section is for Non-English Task Languages)

If your task language is English; for example (English (US), English (UK), English (CA), etc., you may skip this section.

Most of the time, you will use the **Unratable: Foreign Language** rating when the landing page is not in the task language, English, or an acceptable language (please see Section 4.1.5 for rating foreign **Vital** pages).

Landing pages in the task language are clearly a good choice for users in the task location.

Even though they are not considered foreign, landing pages in English or acceptable languages may not be a good "fit" for users in the task location. For example, in some countries there is a very high rate of English literacy. English pages may be a reasonable fit for locations with a high rate of English literacy, but in other locations where knowledge of English is somewhat rare, English landing pages may *not* be a good fit.

Additionally, some queries seem to "ask for" or "invite" English or acceptable language results, and some do not.

When rating pages in English or in an acceptable language, please rate the page based on how helpful you think it is for users. Remember, you should use the **Slightly Relevant** rating for pages which are not very helpful for most users, but are somewhat related to the query.

Here are some examples using Korean (KR) as the task language. In Korea, knowledge of English among the general population is somewhat rare:

Query	Likely User Intent	URL of the Landing Page	Rating	Explanation
[Britney Spears Oops I did it again lyrics], Korean (KR)	Find the lyrics of the Britney Spears song, "Oops I did it again"	http://www.cyworld.com/4641458/3347359	**Useful**	Although the query was typed in English and invites English lyrics, the landing page includes both English lyrics and a Korean translation of the lyrics. This landing page also offers the official music video, which is playable with the right video plug-in. Korean users would find the landing page to be very helpful.
[Britney Spears Oops I did it again lyrics], Korean (KR)	Find the lyrics of the Britney Spears song, "Oops I did it again"	http://www.gasazip.com/162773	**Relevant** or **Useful**	Unlike the example above, the landing page has the lyrics in English only. However, the auxiliary content on the page (e.g. top menu bar, description, links, ads, etc.) is all in Korean. Korean users would prefer to see the auxiliary content in Korean instead of English.
[Britney Spears Oops I did it again lyrics], Korean (KR)	Find the lyrics of the Britney Spears song, "Oops I did it again"	http://www.lyrics007.com/Britney%20Spears%20Lyrics/Oops!..%20I%20Did%20It%20Again%20Lyrics.html#	**Slightly Relevant** or **Relevant**	The landing page was created by a webmaster in the United States. The entire content is in English, including the menu, description, links, etc. Although the query invites English lyrics, most Korean users would prefer to see results from Korean websites where auxiliary content is in Korean.
[Barack Obama], Korean (KR)	Find information about Barack Obama	http://ko.wikipedia.org/wiki/%EB%B2%84%EB%9D%BD_%EC%98%A4%EB%94%94%EB%A7%88	**Useful**	This is a name query and the Wikipedia landing page is about Barack Obama. The article is written in Korean and is helpful to Korean (KR) users.
[Barack Obama], Korean (KR)	Find information about Barack Obama	http://en.wikipedia.org/wiki/Obama	**Slightly Relevant**	This English Wikipedia landing page about Barack Obama has a similar layout to the Korean Wikipedia page (photos, career, presidency, etc.); however, English is not commonly spoken in Korea and is therefore not very helpful to Korea (KR) users.
[Ultranarrow Luminescence Lines from Single Quantum Dots, M. Grundmann], Korean (KR)	Find and read a document titled "Ultranarrow Luminescence Lines from Single Quantum Dots", written by M. Grundmann	http://prl.aps.org/abstract/PRL/v74/i20/p4043_1	**Useful**	This query is very specific and the user clearly wants to read this specific document. Although knowledge of English is rare in Korea, the query strongly invites English results. Many thesis papers and journals are written in English and are not available in a Korean version.

Query	Likely User Intent	URL of the Landing Page	Rating	Explanation
[Titanic 1997], Korean (KR)	Purchase a DVD or find information about the movie "Titanic", released in 1997	http://movie.naver.com/movie/bi/mi/basic.nhn?code=18847	**Useful**	Although the query was typed in English, most Korean users would expect to see Korean transaction pages or movie reviews written in Korean. The landing page in Korean has great information about the movie. It would be very helpful to Korean users.
[Titanic 1997], Korean (KR)	Purchase a DVD or find information about the movie "Titanic", released in 1997	http://www.imdb.com/title/tt0120338/	**Slightly Relevant**	IMDb is a well-known movie information website in the US. The landing page has great content, including casting information, overview, photos, reviews, etc. However, knowledge of English is rare in Korea. This landing page with English content would be unhelpful to most Korean users.

In some locales, English is one of the official languages or a commonly spoken language. Users living in such locales would not be disappointed to see landing pages in English. For example, the Singapore government recognizes four official languages: English, Malay, Chinese, and Tamil, but English is the first and most dominant language in Singapore.

Here are some examples:

Query	Likely User Intent	URL of the Landing Page	Rating	Explanation
[Barack Obama], Chinese_Simplified (SG)	Find information about Barack Obama.	http://en.wikipedia.org/wiki/Obama	**Useful** or **Relevant**	The Singapore government recognizes four official languages: English, Malay, Chinese, and Tamil. English is the first and most dominant language in Singapore. The Wikipedia page in English about Obama would be helpful to users in Singapore
[Barack Obama], Chinese_Simplified (SG)	Find information about Barack Obama.	http://zh.wikipedia.org/zh/%E8%B4%9D%E6%8B%89%E5%85%8B%C2%B7%E5%A5%A5%E5%B7%B4%E9%A9%AC	**Useful** or **Relevant**	This Wikipedia page in Chinese about Obama would also be helpful to users in Singapore.

5.4 Multiple Interpretations

You will rate pages for some queries that have multiple interpretations and multiple user intents.

- In general, pages associated with minor interpretations and unlikely user intents should be rated lower.
- Pages for common interpretations of the query and reasonable user intents should *not* be lowered in rating.
- Only queries with a dominant interpretation can have **Vital** pages.

Here are some examples.

Query Interpretation	Example	Range of Ratings
Dominant Interpretation: Of all the users who type the query, **most users would want this interpretation.**	[apple], English (US): Apple computers. **Most** users who type this query want results on Apple computers. [windows], English, (US): the Microsoft operating system. **Most** users who type this query want results on the Microsoft Windows operating system. [amazon], English (US): the popular website www.amazon.com. **Most** users who type this query want to go to the Amazon website. [median], English (US): the mathematical formula. **Most** users who type this query want results about the mathematical formula. Even though this query has a dominant interpretation, no **Vital** rating is possible since no one can own this query. The highest possible rating for this query is **Useful.** [guinea pig], English (US): the small furry animal often kept as a pet. **Most** users who type this query want results about the animal. Even though this query has a dominant interpretation, no **Vital** rating is possible since no one can own this query. Many webpages have information about guinea pigs. The highest possible rating for this query is **Useful.**	**Vital** to **Off-Topic or Useless**
Common Interpretation: Of all the users who type the query, **many or some users would want this interpretation.**	[apple], English (US): The fruit. **Some** users who type this query could want results about the fruit. [windows], English (US): The glass paned windows for a home. **Many or some** users who type this query could want results about glass windows for a house. [amazon], English (US): The rainforest or river in South America. **Some** users who type this query could want results about the river or rainforest. [ada], English (US): The American Dental Association, the American Diabetes Association, or the American with Disabilities Act. **Many or some** users could want information about any of these organizations. [mercury], English, (US): The car brand, the planet, or the chemical element. **Many or some** users could want information about the car, the planet, or the chemical element. [sandals], English (US): The open type of shoe or the chain of resorts located in the Caribbean Sea. **Many or some** users could want information about the open type of shoe or the chain of resorts	**Useful** to **Off-Topic or Useless** There can be no **Vital** page if the interpretation is not dominant.
Minor Interpretation: Of all the users who type the query, **few users would want this interpretation.**	[ada], English (US): The Atlanta Development Authority or the American Darters Association. **Few** users would want information about these interpretations. [mercury], English (US): The Mercury Magazine (published by the Astronomical Society of the Pacific) or Mercury Records (a record label in the U.K). **Few** users would want information about these interpretations. [hot dog], English (US): "Hot Dog", a movie that was in movie theaters in 1984. **Few** users would want information about this interpretation.	**Relevant** to **Off-Topic or Useless** The less likely you believe the interpretation is, the lower on the scale you should rate the associated result.
"No chance" Interpretation: An interpretation so minor that almost no one would ever want this interpretation.	[guinea pig], English (US): A pig from New Guinea, which is an island country located near Australia (There probably are pigs in New Guinea, but it is extremely unlikely that the user typing the query would have that interpretation in mind.)	**Off-Topic or Useless**

Please note that queries with a dominant interpretation *can* have common interpretations as well.

Query	Dominant Interpretation	Common Interpretation
[windows], English (US)	Microsoft operating system	glass windows that you see through
[kayak], English (US)	travel website	small, human-powered boat

In addition to multiple query interpretations, there may be many different possible user intents. Please decide whether a user intent is reasonable or likely. User intents that are less reasonable or less likely should also be lowered on the rating scale.

User Intent	Example	Range of Ratings
Likely user intent: Many or most users have these intents.	[tetris], English (US): Play Tetris (a video game) online, or download the game [flowers], English (US): Order flowers online, or learn about types of flowers or find pictures of flowers. [credit cards], English (US): Find a credit card company, apply for a card, or compare different brands of credit cards [amazon], English (US): Go to Amazon.com.	**Vital** to **Off-Topic or Useless**
Less likely user intent: Some or few users have these intents.	[tetris], English (US): Research the history of Tetris [flowers], English (US): Find a definition of the word "flower" [credit cards], English (US): Read an encyclopedia article on the history of credit cards [amazon], English (US): Read an encyclopedia article about Amazon.com	**Relevant** to **Off-Topic or Useless** Ratings should reflect how many users these pages would help.

5.5 Specificity of Queries and Landing Pages

Some queries are very general and some queries are specific. And other queries are somewhere in between. Here are some examples that compare levels of specificity of English (US) queries:

Query	More Specific Query	Even More Specific Query
[chair]	[dining room chair]	[ikea "henriksdal" highback upholstered chair]
[cameras]	[Nikon cameras]	[Nikon d5000 slr]
[Toyota]	[Toyota hybrid]	[Toyota Prius 2010]
[library]	[Harvard library]	[Harvard Anthropology library]
[interview questions]	[interview questions for teachers]	[practice interview questions used for Teach For America]
[discount stores in houston]	[walmart stores in houston]	[walmart 9555 South Post Oak Road houston]

Good landing pages need to "fit" the specificity of query to be helpful for users who issued the query. When there is a mismatch between the query and the landing page, you will need to think carefully about how helpful the page is for users and rate accordingly.

Here are some examples of "good" fit between query and landing page specificity:

Query	Likely User Intent	URL of Landing Page	Rating
[digital cameras], English (US)	Users are interested in digital cameras. They might be researching brands or understanding the different options to buy a camera.	http://www.bestbuy.com/site/Cameras-Camcorders/Digital-Cameras/abcat0401000.c?id=abcat0401000	**Useful** – the landing page is the "Digital Cameras" page on the Best Buy website. Best Buy is a well-known camera, electronics, appliance, etc. merchant. This page has descriptions and ratings of popular digital cameras. This landing page fits the query. The query asks for digital cameras and the landing page is about digital cameras.
		http://reviews.cnet.com/digital-cameras/	**Useful** – the landing page is a cnet.com "Digital cameras" review page, with information about many different digital cameras organized by price, manufacturer, and camera features. This landing page fits the query. The query asks for digital cameras and the landing page is about digital cameras.
[Nikon digital cameras], English (US)	Users are probably interested in a Nikon digital camera. Some users may have decided to buy a Nikon, but some may be researching the Nikon brand.	http://www.bestbuy.com/site/olstemplatemapper.jsp?id=pcat17080&type=page&qp=crootcategoryid%23%23-1%23%23-1~~q70726f63657373696e6774696d6d653a313930302d30312d3031~~cabcat0400000%23%230%23%23dh~~cabcat0401000%23%230%23%233e~~nf830||4e696b6f6e&list=y&nrp=15&sc=abCameraCamcorderSP&sp=-bestsellingsort+skuid&usc=abcat0400000	**Useful** – the landing page is the "Nikon digital cameras" page on the Best Buy website. There are over 30 models of Nikon digital cameras for sale and the page has prices, specifications, and reviews for each model. This landing page fits the query. The query asks for Nikon digital cameras and the landing page is about Nikon digital cameras.
		http://www.nikonusa.com/Find-Your-Nikon/Digital-Camera/index.page	**Useful** – the landing page is the "Compact Digital Cameras" page on the official Nikon website. It is not **Vital** because the page is only about **compact** digital cameras, while Nikon also sells digital **SLR** cameras. However, compact digital cameras are very popular and the landing page displays information about many compact digital cameras that may be of interest to users. This landing page fits the query. The query asks for Nikon digital cameras and the landing page is about a popular type of Nikon digital cameras.
		http://reviews.cnet.com/digital-camera-reviews/?filter=1000036_108496_&tag=centerColumnArea1.0	**Useful** – the landing page is a cnet.com "Nikon Digital cameras" review page, with helpful information about many different Nikon digital cameras organized by price, resolution, digital camera type, and features. The page allows users to select cameras to compare price, features, etc. This landing page fits the query. The query asks for Nikon digital cameras and the landing page is about Nikon digital cameras.

Query	Likely User Intent	URL of Landing Page	Rating
[walmart stores in Houston], English (US)	Find Walmart stores in Houston.	http://www.walmart.com/storeLocator/ca_storefinder_results.do?sfsearch_zip=&sfsearch_city=houston&sfsearch_state=TX	**Vital** – the landing page is the Houston "Store Finder" page on the Walmart website. The landing page fits the query because it is the Houston "Store Finder" page on the Walmart website.
		http://www.yelp.com/search?find_desc=walmart&ns=1&find_loc=houston,+tx	**Useful** or **Relevant –** the landing page is the Walmart Houston page on Yelp. It has a list of Walmart store locations in Houston and displays them on a map. There are also reviews of some specific Walmart stores. The landing page fits the query. The query asks for Walmart stores in Houston and the landing page is about Walmart Stores in Houston.

When there is a mismatch between the query and landing page, assigning a rating can be difficult. You have to think about how helpful a page is for users and base your rating on that.

Here are some examples of good and bad fits along with suggested ratings:

Query	User Intent	URL of Landing Page	Rating
[interview questions for teachers], English	Find interview questions for teacher candidates	http://www.career.vt.edu/Interviewing/TeachingInterviewQuestions.html	**Useful**: The landing page displays many questions which would be very helpful to users practicing for a teaching position interview. The landing page fits the query.
		http://www.nmsa.org/portals/0/pdf/member/job_connection/Interview_Questions.pdf	**Relevant**: The landing page has sample interview questions for teacher and administrator positions at the middle school level. The landing page is more specific than the query, but has many helpful questions that would be helpful when preparing for any teaching interview.
		http://www.glassdoor.com/Interview/Teach-for-America-Teacher-Interview-Questions-EI_IE105049.0,17_KO18,25.htm	**Slightly Relevant**: The landing page on glassdoor.com has information about the Teach for America interview process and displays some interview questions that were asked of applicants to the program. Some of the questions are general enough to be helpful in preparing for a "regular" teaching position, but some are specific to the Teach for America program. The landing page is more specific than the query, but it could still be helpful for some users.
		http://career-advice.monster.com/job-interview/interview-questions/100-potential-interview-questions/article.aspx	**Off-Topic or Useless**: There are many good pages with interview questions for teachers. A page with *general* interview questions has little or no utility for users. The landing page is more general than the query. The query asks for interview questions for teachers, while the landing page has general interview questions.

Query	Likely User Intent	URL of Landing Page	Rating
[Honda Accord], English (US)	Users probably want to buy a car and are interested in finding information about the Honda Accord. There are two models of the Accord: the Accord Sedan and the Accord Coupe.	http://automobiles.honda.com/accord/	**Vital**: The landing page is the official Honda Accord page. The landing page fits the query. The query asks about the Accord and the landing page is about the Accord.
		http://automobiles.honda.com/	**Useful**: The landing page is the official Honda Automobiles webpage. There are a picture and a prominent "Accord" link on the page. There are a lot of helpful features on this page for users interested in Honda Accords and this is the official website. The landing page is a little more general than the query. The query asks for the Accord, while the landing page is about all Honda car models.
		http://www.edmunds.com/honda/accord/review.html	**Useful**: The landing page has comprehensive information about the Honda Accord, including current and previous models. The page has pricing, reviews, spec, photos, etc. The landing page fits the query. The query asks about the Accord and the landing page is about the Accord.
		http://automobiles.honda.com/accord-sedan/ http://automobiles.honda.com/accord-coupe/	**Useful:** The landing pages are the official pages of the Accord Sedan and the Accord Coupe. These landing pages are more specific than the query, but since there are only two Accord models and they are both popular. Official pages (or other very helpful pages) for either of the two models are **Useful**.
		http://automobiles.honda.com/tools/build-price/models.aspx	**Relevant**: The landing page is the "Build and Price Your Honda" page on the Honda Automobiles webpage. Users can build and price different Accord models, as well as all other Honda cars. The landing page does not quite fit the query. It has Accords prominently displayed and may be helpful for some users, but we do not know that this is the type of page most users want.
		http://automobiles.honda.com/accord-coupe/exterior-colors.aspx	**Slightly Relevant**: The landing page is the "exterior colors" page for the Honda Accord Coupe. The landing page does not fit the query. It is much more specific than the query and there is little content related to the query.

Query	Likely User Intent	URL of Landing Page	Rating
[Target], English (US)	Go to target.com or find a local Target store.	http://www.target.com/	**Vital**– the landing page is the official Target homepage. The landing page fits the query.
		http://sites.target.com/site/en/spot/page.jsp?title=store_locator_new&ref=nav_store_locator	**Useful or Relevant** – the landing page is the "store finder" page on the Target website. The landing page is more specific than the query, but many or some users would be interested in this page.
		http://weeklyad.target.com/sunnyvale-ca-94086/homepage#	**Useful or Relevant** – the landing page is the "weekly ads" page on the Target website. The landing page is more specific than the query, but many or some users would be interested in this page.
		http://www.target.com/Kids/b/ref=nav_t_spc_4_0/178-4746585-1881721?ie=UTF8&node=1041972	**Relevant** – the landing page is the "toys" page on the Target website. The landing page is more specific than the query. Some users would be interested in this page.
		http://sites.target.com/site/en/company/page.jsp?contentId=WCMP04-030796	**Slightly Relevant** or **Relevant** – the landing page is the "careers" page on the Target website. The landing page is more specific than the query. Fewer users would be interested in this page.
		http://www.target.com/s?searchTerm=boys+shorts&category=0|All|matchallany|all+categories	**Slightly Relevant**– the landing page is the "boys' shorts" page on the Target website. The landing page is much more specific than the query. Few users would be interested in this page.

5.6 Common Rating Problems

Listed below are some common rating mistakes. Most of these mistakes have to do with user intent and the "fit" of the landing page to the query.

5.6.1 Dictionary or Encyclopedia Results

Dictionary or encyclopedia pages are often helpful to raters who are trying to understand the query. They can also sometimes be helpful for the user, but not when the user already understands the words in the query and is looking for something different. Here are some examples.

Query	Likely User Intent	Landing Page	Rating	Reason
[photosynthesis], English (US)	Find out how photosynthesis works. This is an information query.	http://en.wikipedia.org/wiki/Photosynthesis	**Useful**	This is a good article about photosynthesis and would be helpful to most users.
[e.g.], English (US)	Find the meaning of the Latin abbreviation "e.g." This is an information query.	https://www.e-education.psu.edu/styleforstudents/c3_p28.html	**Useful** or **Relevant**	This is a good explanation of the abbreviation "e.g." (as well as "i.e." and "et al.") on a well-know university website and would be helpful to most or many users.
[banks], English (US)	Find a bank. This is an action query.	http://www.investorwords.com/401/bank.html http://en.wikipedia.org/wiki/Bank	**Slightly Relevant**	Most English US users know what a bank is. Even an excellent definition or encyclopedia article has little utility for most users.

5.6.2 Action vs. Information Intent

Raters often give high ratings to pages for information user intents even when the query is an action query. For queries that clearly have action intent, information pages should not be rated above **Relevant**. Think about whether users want to know something or do something. Look at the content of the page and decide if the page is helpful for a "know" or "do" intent.

Query	Likely User Intent	Landing Page	Rating	Reason
[e-cards], English (US)	Send an e-card. This is an action query.	http://en.wikipedia.org/wiki/E-card	**Slightly Relevant**	Most users want to send an e-card. This Wikipedia page is really not helpful for sending an e-card.
[bejeweled], English (US)	Play Bejeweled online or download the game. This is an action query.	http://en.wikipedia.org/wiki/Bejeweled	**Relevant** or **Slightly Relevant**	Most users want to play the game. This Wikipedia page could be helpful for some users because it includes information about what platforms the game runs on and some instructions on how to play the game.
[Federal Express], English (US)	Send a package, track a package, or find a Federal Express store. This is an action query.	http://www.allbusiness.com/glossaries/federal-express/4962036-1.html	**Slightly Relevant**	This is a low quality page with a short business definition of Federal Express. Users do not want a definition; they want to do something. This page would be helpful for few users.
[netbooks], English (US)	Product queries are usually both "do" and "know" queries. People often do extensive research before buying items, and the "know" intent is very important for product queries.	http://www.amazon.com/s/ref=nb_sb_noss?url=search-alias%3Daps&field-keywords=netbooks&x=0&y=0	**Useful**	This is a page on amazon.com with many netbooks for sale. It's a good "know" and "do" page. Users can do research, read reviews, and find out about different models, as well as buy a netbook. It would be helpful for most users.
		http://reviews.cnet.com/best-netbooks/	**Useful**	The landing page is CNETs "Best Netbooks" review page, with helpful information about many different netbooks. This is a good "know" page. It would be helpful for most users.

Please respect the "know" intent of product queries. Many people research items online before making a decision about whether to buy the item. Most product queries are "know" and "do" queries.

5.6.3 Queries that Ask for a List

Some queries seem to "ask for a list". Here are a few principles to help you out when rating these types of queries:

- When the query seems to ask for a list that includes many, many possibilities, individual examples usually are not as helpful as a list.
- When the list of possibilities is short, then individual examples are helpful.
- Sometimes, there are very famous or popular examples on the list. In these cases, the individual famous or popular examples are helpful, even if the list of possibilities is long.

To summarize, if there are few items in the list, then high quality landing pages for individual items are helpful. If there are so many possibilities that any one item seems too specific, lists of results are usually more helpful, unless an individual item is very popular or highly expected.

Here are some examples of queries that ask for a list:

Query	Likely User Intent	URL of Landing Page	Rating
[chicken recipes], English (US)	Users probably want to prepare a chicken dish and are looking for some recipes to choose from. Users probably expect and want a list of recipes.	http://www.foodnetwork.com/topics/chicken/index.html http://allrecipes.com/Recipes/Meat-and-Poultry/Chicken/Main.aspx	**Useful** –Users can find many chicken recipes (with reviews) on these pages on popular recipe websites. These landing pages fit the query. Most users would find these pages helpful.
		http://www.foodnetwork.com/recipes/tyler-florence/chicken-parmesan-recipe/index.html	**Relevant** or **Slightly Relevant:** This page on the Food Network website has a single recipe for chicken parmesan. It's a popular type of chicken recipe, but the page is more specific than the query. Some or few users would find this page helpful.
		http://allrecipes.com/Recipes/Meat-and-Poultry/Chicken/Fried/Top.aspx	**Relevant** or **Slightly Relevant** – This page has 20 recipes for fried chicken, a popular chicken dish. Even though there are 20 different recipes, it is for the same basic dish. Therefore, this landing page is also more specific than the query. Some or few users would find this page helpful.
		http://www.free-gourmet-recipes.com/hchicken.shtml	**Slightly Relevant** – This is a low quality page with distracting pop-ups that appear when you hover your mouse over hyperlinked words in the list of recipes. These pop-ups actually prevent you from reading the titles of some of the recipes. However, the page does have links to some chicken recipes, so it is not **Off-Topic or Useless**. Very few users would find this page helpful.
		http://www.popeyes.com/ http://www.zaxbys.com/home.aspx http://www.kfc.com/	**Off-Topic or Useless** – These are homepages of chicken restaurants. These pages have no utility for users looking for chicken recipes.

Query	Likely User Intent	URL of Landing Page	Rating
[baby toys], English (US)	Find information about baby toys or purchase baby toys.	www.toysrus.com/category/index.jsp?categoryId=2639789	**Useful**: This is the baby toys section of the Toys R Us website. The landing page is a list of baby toys organized by category. Even though the list of stores that sell baby toys is long, the Toys R Us baby toys' page should be included in a list of results for this query because Toys R Us is a very popular toy store. The landing page fits the query. Most users would find this page helpful.
		http://www.hearthsong.com/First-Toys/Category_S2006_D1401_C5102_ALL.html	**Useful** or **Relevant**– This page has a nice selection of baby toys by category. HearthSong is not a well-known merchant, but it's a high quality page. The landing page fits the query. Many or some users would find this page helpful.
		http://www.toysrus.com/product/index.jsp?productId=2574131	**Relevant** or **Slightly Relevant:** This is the landing page for a specific baby toy on the Toys R Us website. This is a classic type of baby toy from a popular store, but the page is more specific than the query. Some or few users would find this page helpful.
		http://www.landofnod.com/family.aspx?c=3147&f=6220	**Relevant** or **Slightly Relevant**: This page has one specific, popular baby toy on a high quality site. There are so many possible toys that it's impossible to know if any one single toy would help the user. However, this is a good site and this toy is popular. This is a classic type of baby toy, but the page is more specific than the query. Some or few users would find this page helpful.
		http://www.toysforbabies.org/	**Slightly Relevant:** This page is spam (see the Webspam Guidelines, Part 5 of the General Guidelines, for more information). Clicking the product links takes you to Amazon. Nothing can be purchased on the landing page. Also, if you click the "Recent Posts" links, you will find articles with very superficial content and/or nonsensical text. Few users would find this page truly helpful.
		http://www.toysrus.com/product/index.jsp?productId=3747483	**Off-Topic or Useless** or **Slightly Relevant:** This page has a baby bath toy net. It's not technically a baby toy, though it's in the baby toy section of Toys R Us. There are other baby toys shown at the bottom of the page. The landing page is not a good fit for the query. Very few users would find this page helpful.
		http://www.rctoys.com/	**Off-Topic or Useless** –This website sells remote control toys, which are not suitable for babies. The landing page does not fit the query. Very few or no users would find this page helpful.

Query	Likely User Intent	URL of Landing Page	Rating
[hotels], English (US)	Users are probably planning a trip, but this query is very general and vague. Even though we do not specifically know what users want, there are helpful and unhelpful results for this query.	http://www.expedia.com/Hotels http://www.orbitz.com/App/ViewHotelSearch	**Useful -** Expedia and Orbitz are popular travel aggregator websites, and the hotel pages on these websites can help users find a hotel in the US. Users can read reviews, compare hotels, and make a reservation. These landing pages fit the query. Most users would find these pages helpful.
		http://www.marriott.com/ http://www.sheraton.com/	**Useful** or **Relevant** – These are popular hotel chains that are available in most of the US and have many different price levels. Even though the list of possible hotel chains is long, the homepages of these individual hotel chains are probably helpful for many users because they have sub-brands that offer many different prices, features, and location options. These landing pages are more specific than the query, but the pages are still helpful for many users.
		http://www.motel6.com/ http://www.comfortinn.com/	**Relevant** – These hotel chains are also available in most of the US, but they have lower prices and target budget travelers. These pages would be helpful for some users, but they do not offer as many options in price or features. These landing pages are even more specific. Many or some users would find these pages helpful.
		http://www.marriott.com/hotels/travel/oakmv-courtyard-oakland-emeryville/	**Slightly Relevant** – This is the webpage of the Marriott Courtyard hotel in Emeryville, California. This page is too specific for the query, but this is a well-known brand and users can navigate to other Marriott hotels from this page. Few users would find this page helpful.
		http://petshotel.petsmart.com/	**Off-Topic or Useless** – This is the webpage of PetSmart PetsHotel, a chain of pet hotels in many states in the US. This chain provides overnight care for dogs and cats, not humans. This page is much too specific for the query. Users are looking for hotels for humans, not for animals. Very few or no users would find this page helpful.

5.6.4   Misspelled and Mistyped Queries

You will notice that some queries are misspelled or mistyped.

For **obviously** misspelled or mistyped queries, you should base your rating on user intent, not necessarily on exactly how the query has been spelled or typed by the user.

For queries that are **not** obviously misspelled or mistyped, you should assume users are looking for results for the query as it is spelled.

For the query, [federal expres], English (US), it is reasonable to assume that the user is looking for Federal Express at http://www.fedex.com/us/. For the query, [my sapce], English (US), it is reasonable to assume the user is looking for MySpace at http://www.myspace.com/. There are no other reasonable interpretations for these queries.

Then consider the query [John Stuart], English (US). Even though raters may believe that the user wants to go to pages associated with Jon Stewart, the well-known comedian and host of "The Daily Show" (a popular news satire TV show), we cannot assume that the query has been misspelled. There is a Las Vegas show producer named John Stuart, whose name exactly matches the spelling of the query, and it is very likely that there are "regular" people whose names match the spelling of the query, as well.

Important: Do not assume a query has been misspelled if there is a person or entity that matches the spelling in the query, or even if it is just reasonable that there might be such a person. Sometimes, people exist for whom there are no web results.

Here are some examples of queries that are obviously misspelled.

Query	Query Interpretation	URL of the Landing Page	Description of the Landing Page	Rating
[federal expres], English (US)	The only reasonable query interpretation is the company named Federal Express.	http://www.fedex.com/	Official homepage of Federal Express	**Vital**
[my sapce], English (US)	The only reasonable query interpretation is the website MySpace.	http://www.myspace.com/	Official homepage of Myspace	**Vital**
[the ecomonist], English (US)	The only reasonable query interpretation is the news and economics publication.	http://www.economist.com/	Official homepage of The Economist	**Vital**
[expdeia], English (US)	The only reasonable query interpretation is the travel website.	http://www.expedia.com/	Official homepage of Expedia	**Vital**
[New England Patroits], English (US)	The only reasonable interpretation is the NFL football team.	http://www.patriots.com/	Official homepage of the New England Patriots football team	**Vital**
[byonce Knowles], English (US)	The only reasonable interpretation is the famous singer/actress named Beyonce Knowles.	http://www.beyonceonline.com/us/home	Homepage of Beyonce's official and maintained website	**Vital**

People queries can be difficult to rate. Here are some examples. The first two queries should *not* be considered misspelled. The third query is obviously misspelled.

Query	Query Interpretation	URL of the Landing Page	Description of the Landing Page	Rating
[Jamie Fox], English (US)	There are several reasonable interpretations for this query: the guitarist named Jamie Fox, Jamie Fox Photography, regular people named Jamie Fox, and the famous actor named Jamie Foxx. Because Jamie Foxx is such a famous actor and his name might be misspelled, we will consider Jamie Foxx to be a minor interpretation, not off-topic.	http://www.jamiefoxguitar.com/	Homepage of the official website of Jamie Fox, the guitarist	**Useful**
		http://jamiefoxphotography.com/	Homepage of the official website of Jamie Fox Photography	**Relevant** or **Useful**
		http://www.jamiefox.net/	Homepage of the official website of Jamie Fox, a web developer	**Relevant** or **Useful**
		http://www.jamiefoxx.com/	Homepage of the official website of Jamie Foxx, the actor	**Relevant** or **Slightly Relevant**
		http://us.imdb.com/name/nm0004937/	IMDb page about Jamie Foxx, the actor	**Relevant** or **Slightly Relevant**
[Micheal Jordan], English (US)	There are several ways to spell this first name. The most popular way is Michael, but Micheal is also sometimes used. Because Michael Jordan is such a famous athlete/celebrity and his name might be misspelled, we will consider Michael Jordan to be a minor interpretation, not off-topic.	http://www.linkedin.com/in/michealjordan	LinkedIn page for Micheal Jordan, a technician in Mobile, Alabama.	**Useful or Relevant**
		http://www.nba.com/playerfile/michael_jordan/index.html	Michael Jordan's page on the NBA basketball website.	**Relevant** or **Slightly Relevant**
		http://www.youtube.com/watch?v=f6WQLvRvtjs	Video titled "Micheal Jordan vs. Himself". Even though the spelling matches the query, the video is about the basketball player, not someone named Micheal Jordan.	**Relevant** or **Slightly Relevant**
[Michae lJordan], English (US)	In contrast to the above examples, the query [Michae lJordan] is obviously misspelled. The user accidentally put a space after the letter "e" instead of after the letter "l". The dominant interpretation of this mistyped query is Michael Jordan, the basketball player. If he has a homepage, the rating would be **Vital**.	http://www.nba.com/playerfile/michael_jordan/index.html	Michael Jordan's page on the NBA basketball website.	**Useful**

It is sometimes difficult to find results for queries that are very similar to popular queries.

To find results for the query [Jamie Fox], English (US), it is helpful to use the "minus" search operator. Typing **["Jamie Fox" –foxx]** will help you to filter out results for Jamie Foxx, the famous actor, and narrow your search to results for "Jamie Fox".

Some queries look like URLs. We will call these queries "URL Queries".

Some URL queries are exact, perfectly-formed, working URLs, such as [www.ibm.com], English (US). Some queries that contain partial URLs, such as [ibm.com], English (US), become working URLs when you add "www." or "http://" to the front of the URL. We will consider [www.ibm.com], English (US) and [http://www.ibm.com], English (US) to be the same query as [ibm.com], English (US). All of these are considered "URL queries".

Some queries are website or webpage names, such as [yahoo], English (US) or [yahoo mail], English (US). These queries do not contain ".com", "www" or other standard components of a URL. These are navigation or "go" queries, but we will not consider them URL queries.

Most queries are neither URL queries nor website/webpage name queries. Most of the time, queries contain terms that do not refer to a particular website or webpage.

Here are some examples of English (US) queries:

URL queries	Website Name/Webpage Name Queries (these are "go" queries, with no "URL parts")	"Generic" Queries
[ebay.ca] [amazon.com] [people.com] [bbc.co.uk] [www.dealbook.com] [mail.yahoo.com] [news google.com] [tax form 1040 irs.gov] [rei.com]	[ebay] [amazon] [people] [bbc] [dealbook] [yahoo mail] [google news] [irs 1040 tax form official page] [rei kayak page]	[couches] [diabetes] [weight loss] [tax forms] [quilting]

Let's first discuss URL queries. Some URL queries are *not* "working URL" queries. The URLs do not load if you type or paste them into your Firefox browser address bar. However, we believe users have a specific page in mind. We will call these "imperfect URL queries". There are many types of imperfect URL queries. Here are descriptions of some of them:

- The query has the same format as a perfect URL query, but the page doesn't load. Here is an example: [www.UnitedStatesPassportProvider.com], English (US).
- The query has the same format as a perfect "working" URL query, but is obviously misspelled and does not "work". Here are some examples: [www.pizzzzahut.com] and [www.mcriosoft.com].
- The query has a URL-like format, but contains extra words and/or spaces. Here is an example: [Australian open tennis tournament.com], English (US). We will call this an "imperfect URL query" because it contains "tournament.com", which is part of a URL, but there are spaces in the query.
- The query has a mix of words and URLs, such as [barbie.com dress up games], English (US).

Some URL queries can be extremely hard to rate. Although you will need to visit the landing page to see and evaluate the content, you will also need to look carefully at the URL of the landing page and the URL in the query. Do not just rate URL queries and results based on the appearance of the URL.

Trying to interpret user intent for imperfect URL queries is hard. It is very easy for users to mistype URLs.

If the query is a perfectly-formed, working URL, please consider *that* URL to be the dominant interpretation. The **Vital** rating should be given when the URL of the page exactly matches the URL in the query. Please note that sometimes the URL of the landing page may contain a longer string than the URL in the query, or look different in other ways. For example, for [anthem.com], English (US), both http://www.anthem.com/ and http://www.anthem.com/home.html should be rated **Appropriate Vital** since the landing page is the same.

If the query is not a perfectly-formed, working URL and/or does not load, please use your judgment to interpret user intent. Do not assign a rating of **Vital** unless there is little or no doubt that the page matches user intent.

Here are some examples.

Query	Likely User Intent	Rating Examples
[www.myspace.com], English (US)	Go to the MySpace website. The URL is correct.	**Vital** landing page URL: http://www.myspace.com/
[www.yahoo.c0m], English (US) [yahoo.xcom], English (US) [yahoo.co], English (US)	Even though these URLs do not load, it is clear the user wants to go to Yahoo.	**Vital** landing page URL: http://www.yahoo.com/
[huffingtontonpost.com], English (US)	In this case, the landing page is spam. It is a fake search page. (You will learn about spam pages in Part 4 of the "General Guidelines".) It is very likely that the user wants to navigate to www.huffingtonpost.com. However, we will respect the query as written and consider www.huffingtontonpost.com to be dominant.	**Vital** landing page URL: http://www.huffingtontonpost.com (You will also need to add a **Spam** flag. Please see Part 4 of the "General Guidelines".) **Useful** landing page URL: http://www.huffingtonpost.com
[wwww.ibm.com], English (US)	Even though the URL doesn't load, it is clear that the user wants to go to the IBM homepage.	**Vital** landing page URL: http://www.ibm.com/
[tax form 1040 irs.gov], English (US)	Even though the query contains spaces, it is clear that the user wants to go to the webpage on the official IRS government website for the current 1040 tax form.	**Vital** landing page URL: http://www.irs.gov/pub/irs-pdf/f1040.pdf
[toys are us.com], English (US)	There is a well-known US toy company whose homepage is www.toysrus.com. The name of this company is frequently misspelled. Even though this is an imperfect query due to misspelling and extra spacing, it is clear that the user wants to go to the homepage at www.toysrus.com.	**Vital** landing page URL: http://www.toysrus.com/
[amazon com], English (US)	Even though there is no "dot" between "amazon" and "com", it is clear the user wants to go to amazon.com.	**Vital** landing page URL: http://www.amazon.com
[i hire chemists.com], English (US)	Even though the query contains spaces, it is clear that the user wants to go to the job posting website at www.ihirechemists.com.	**Vital** landing page URL: http://www.ihirechemists.com/

Now let's talk about "website name" or "webpage name" queries, which are *not* URL queries. They are queries which contain the names of websites or webpages, and the *dominant* interpretation of the query is the website or webpage. Some website name queries have other meanings, besides the website.

Website or Webpage Query	Explanation
[kayak], English (US)	Users could be looking for a kayak (a type of boat), but Kayak is a very popular travel website. The website kayak.com is the dominant interpretation
[youtube], English (US)	YouTube is one of the most popular websites on the Web.
[ebay], English (US)	eBay is one of the most popular websites on the Web.
[webmd], English (US)	WebMD is a very popular medical information website.
[twitter], English (US)	Twitter is a very popular website.
[cafepress], English (US)	Cafepress is a website where users can buy t-shirts and other gifts and even have them custom-made.
[addicting games], English (US)	AddictingGames is a very popular game website.
[rei kayak page], English (US)	Users want to go to the "kayak" page on the REI website.

Here are some examples of queries which are *not* website queries and are *not* URL queries. Website names exist that match these queries, but those websites are probably not what users have in mind. These queries do not have **Vital** pages.

Generic Query	Explanation
[birdcages], English (US)	Users are probably interested in researching or buying a birdcage. This is a generic query. There is no **Vital** page. There is a store with the URL <u>birdcages.com</u>, but many stores sell birdcages.
[kamasutra], English (US)	Users are probably interested in learning about the Kama Sutra or reading the Kama Sutra text. There is no **Vital** page. There is a store with the URL <u>kamasutra.com</u>, but that probably is not the dominant interpretation of this query.
[weightloss], English (US)	Users are looking for weight loss information, and there are many good authoritative pages with weight loss information. There is a website <u>weightloss.com</u>, which has helpful, common sense information about losing weight, but users probably are not trying to go to that page.
[couches], English (US)	Users are interested in researching or buying a couch. There are many good websites that sell couches. There is a website <u>couches.com</u>, but there is nothing in the query that indicates users want to go to couches.com.

Keep in mind that just about any query can be turned into a URL by adding ".com", but without the ".com" included in the query, you should not assume the query is a website name.

In other words, just because the query is [couches] does not mean that the result <u>http://www.couches.com</u> is what the user wants. Please be careful with "generic" queries. A commonly used spam technique is to create websites with generic names.

When users issue URL queries, the intent is to go to a specific page. That page should be rated **Vital**. It can be very hard to rate "non-Vital" pages for URL queries. Sometimes, the **Vital** page is the only helpful result for a URL query. But sometimes, other pages are helpful as well. Here are some examples of pages with information about the queried website. Ratings for such pages can range from **Off-Topic or Useless** to **Useful**:

Query	Likely User Intent	URL of the Landing Page	Description of the Landing Page	Rating
[greatamericanphotocontest.com], English (US)	Go to <u>http://www.greatamericanphotocontest.com/</u>, a website where users post baby pictures which are supposed to be entered in a baby photo contest each month	<u>http://www.greatamericanphotocontest.com/</u>	The landing page is the target of the query	**Vital**
		<u>http://www.complaintsboard.com/byurl/greatamericanphotocontest.com.html</u>	The landing page displays complaints that people have written about the URL in the query. The information could be helpful for users planning to visit and interact with the website.	**Useful** or **Relevant**
		<u>http://www.419legal.org/fradulent-website/29043-great-american-photo-contest.html</u>	The landing page is a forum with complaints about the website. The information could be helpful for users planning to visit and interact with the website.	**Useful** or **Relevant**
		<u>http://www.quantcast.com/greatamericanphotocontest.com</u>	The landing page has usage statistics for the greatamericanphotocontest.com website. There are many pages that give these kinds of statistics, but few users would be interested in this information.	**Slightly Relevant**
		<u>http://www.killerstartups.com/Site-Reviews/greatamericanphotocontest-com-baby-photo-contest</u>	The landing page is a low quality, spammy page with general information about the website. It was created to display ads and has little utility for users.	**Slightly Relevant** or **Off-Topic or Useless**

Query	Likely User Intent	URL of the Landing Page	Description of the Landing Page	Rating
[wtpeople.com, English (US)	Go to http://www.wtpeople.com/, home page of We the People/Wisconsin	http://www.wtpeople.com/	The landing page is the target of the query	**Vital**
		http://wistechnology.com/articles/3452/	The landing page is an article written by one of the founders of "We the People/Wisconsin", which provides insight into why he founded the organization and website. Even though the landing page is not on the target website, it might have utility for some users.	**Relevant**
		http://www.alexa.com/siteinfo/wtpeople.com	The landing page has usage statistics for the wtpeople.com website. There are many pages that give these kinds of statistics, but few users would be interested in this information.	**Slightly Relevant**
[facebook.com], English (US)	Go to http://www.facebook.com/, a social networking website Note: Privacy and security are concerns for Facebook and other social networking sites.	http://www.facebook.com/	The landing page is the target of the query	**Vital**
		http://computer.howstuffworks.com/internet/social-networking/networks/facebook.htm	The landing page has an article titled "How Facebook Works", which explains how to create an account and a profile, find friends, etc. This page would be helpful for users who want information about how to use the website.	**Useful**
		http://www.sophos.com/security/best-practice/facebook/	Sophos is a well-known internet security company. The landing page on the Sophos website has recommendations for setting up or adjusting Facebook privacy settings. This page would be helpful for users concerned about their privacy.	**Useful**
		http://www.huffingtonpost.com/2010/05/13/facebook-privacy-settings_n_575732.html	The landing page has a video that teaches users how to adjust the privacy settings on their user profile. The video would be helpful for users concerned about their privacy settings.	**Useful**
		http://topics.nytimes.com/top/news/business/companies/facebook_inc/index.html	The landing page on the New York Times site has information about the Facebook website and a collection of links to articles about Facebook. Some or many users might be interested in these articles.	**Relevant** or **Useful**
		http://www.commonsensemedia.org/facebook-parents	The landing page has information and advice for parents about Facebook. Some or few users would be interested in this page.	**Relevant** or **Slightly Relevant**
		http://www.alexa.com/siteinfo/facebook.com	The landing page has usage statistics for the facebook.com website. There are many pages that give these kinds of statistics, but few users would be interested in this information.	**Slightly Relevant**

Query	Likely User Intent	URL of the Landing Page	Description of the Landing Page	Rating
[ratemyprofessors.com], English (US)	Go to http://www.ratemyprofessors.com/, a website where students can rate their college professors	http://www.ratemyprofessors.com/	The landing page is the target of the query	**Vital**
		http://www.nytimes.com/2010/03/14/magazine/14FOB-medium-t.html	The landing page is a New York Times article dated March 14, 2010 about the ratemyprofessors.com website. Many or some users might be interested in this article.	**Useful** or **Relevant**
		http://cellphonereviewsnow.com/news/RateMyProfessors.com.html	The landing page is a low quality page that contains a paragraph about ratemyprofessors.com that was copied from a Wikipedia article. Few or no users would be interested in this page.	**Slightly Relevant** or **Off-Topic or Useless**
		http://www.bizjournals.com/baltimore/stories/2006/04/17/story8.html?from_rss=1	The landing page has an article dated April 14, 2006 about the ratemyprofessors.com website. Few or no users would be interested in this outdated information.	**Slightly Relevant** or **Off-Topic or Useless**

5.6.6 New and Old Pages

Information or "know" queries may be about recent or past events. The landing page should be rated based on fit to the informational need of the query.   Some queries demand very recent results. Most of the time, you need to consider the content of the page rather than the date on the page.

For some queries, timeliness is very important. Queries for recent events and recurring events need pages with recent content.   We assume that users who type queries looking for results from an election, sporting event, or other type of annual competition are looking for the most recent results, not results from previous years. Here are some examples.

Query	Likely User Intent	Useful Pages	Slightly Relevant Pages
[us open golf results], English (US)	Find a page that displays **the most recent** results for this golf tournament. This is an information query.	Wikipedia page with the 2011 results: http://en.wikipedia.org/wiki/2011_U.S._Open_%28golf%29	Wikipedia page with the 2009 results: http://en.wikipedia.org/wiki/2009_US_Open_Golf_Championship
[golden globe winners], English (US)	Find **the most recent** winners of Golden Globe awards. This is an information query.	Page on the BBC website with this information: http://www.bbc.co.uk/news/entertainment-arts-11991049	Page on the BBC website information about the 2008 winners: http://movies.about.com/od/awards/a/globes121406.htm
[Nobel Peace Prize Winner], English (US)	Find the name of **the most recent** winner of this prize. This is an information query.	Page on the Reuters website with this information: http://www.nobelprize.org/nobel_prizes/peace/laureates/2010/ Page on the New York Times website with this information: http://www.nytimes.com/2009/10/10/world/10nobel.html	Page on the BBC website with the 2006 winner of this prize: http://news.bbc.co.uk/2/hi/europe/6047020.stm

Please note, however, that, depending on when annual events occur, the most helpful pages may be for the past event *or* the current/upcoming event. If the event took place several months ago, the most helpful pages would probably be about the past event. If the event will take place in a few months, the most helpful pages would probably be about the upcoming event. You will have to use your judgment.

If the landing page appears to be the official page of the event, it should get a **Vital** rating, whether the content is about the past or upcoming event.

Information queries may need recent results as well. For example, if the query is [population of paris], English (US), users are looking for the most current population numbers.

On the other hand, if the query is [population of France in 1813], the issue is not how "new" or "recent" the page is, but whether it has the information requested. Sometimes "old" pages are the only good source of information about past events. "Old" pages are not necessarily "outdated" or bad. It depends on the query and the page content.

Here are some examples.

Query	Likely User Intent	URL of the Landing Page	Description of the Landing Page	Rating
[Audrey Hepburn's death], English (US)	Find information about Audrey Hepburn's death	http://www.nytimes.com/1993/01/21/movies/audrey-hepburn-actress-is-dead-at-63.html?pagewanted=1	This New York Times article was published January 21, 1993, the day after Audrey Hepburn's death. Even though the article is almost 20 years old, it has what the user is looking for.	**Relevant** or **Useful**
[Michael Jackson's death], English (US)	Find information about Michael Jackson's death	http://www.washingtonpost.com/wp-dyn/content/article/2009/06/25/AR2009062503127.html	This Washington Post article was published on June 26, 2009, the day after his death. Even though it is not a recent article, it has information users might be looking for. Because there have been more recent articles published about the circumstances of his death, this article would no longer be considered **Useful**.	**Relevant** or **Slightly Relevant**
[the battle of the bulge], English (US)	Find information about the Battle of the Bulge, a famous World War II battle that took place in 1944.	http://www.amazon.com/Battle-Story-Bulge-John-Toland/dp/0803294379/ref=sr_1_3?ie=UTF8&s=books&qid=1271373258&sr=1-3	The landing page on amazon.com is for a well-known book about this battle. The book was originally published in 1959 and was most recently revised in 1999. Even though the book was not published recently, the battle was fought long ago and information about the battle has not changed. The book is not considered outdated.	**Relevant**
[red sox schedule], English (US)	Find the current season's schedule for the Boston Red Sox baseball team	http://www.bostonspastime.com/schedule.html	The landing page has the current schedule, which is what the user is looking for.	**Useful**
		http://boston.redsox.mlb.com/schedule/index.jsp?c_id=bos&m=4&y=2006	The landing page has the 2006 schedule, which is not what the user is looking for because it has outdated information.	**Slightly Relevant** or **Off-Topic or Useless**

5.6.7   Search Engine Result Pages

This section is about search engine results pages. Search engine results pages should be rated just like other landing pages: rate the landing page on the basis of how helpful it is for users. Sometimes raters find these pages difficult to rate, so this section gives examples specifically on this topic.

Here are examples of search engine results pages. These are pages users see after entering queries on a search engine.

Search Results Page

Shopping Search Results Page

Video Search Results Page

Google puppies   ✕   Search

Advanced search

- 📇 Everything
- 📷 Images
- 🎬 **Videos**
- 📖 Books
- 🏷️ Shopping
- ▼ More

Any duration
Short (0–4 min.)
Medium (4–20 min.)
Long (20+ min.)

Any time
Past hour
Past 24 hours
Past week
Past month
Past year
Custom range...

Sorted by relevance
Sorted by date

▶ 2:38

cute **puppy** - YouTube
youtube.com
Feb 4, 2006 - 3 min - Uploaded by kiricybo
Marshmellow the WHITE pug **puppy**by FFPugs 3926447 views ...
Pug **Puppies** 4 Weeks Oldby aatkinson001 ...

▶ 1:09

Puppy vs. Cat - YouTube
youtube.com
May 9, 2006 - 1 min - Uploaded by Sanchey
Cat:what do u want I am trying to poop the last thing I need is
puppies bugging me!!! **Puppies**: OMG it is a ...

Shiba Inu **Puppy** Cam, Ustream.TV: Tune in daily to ...
ustream.tv
Apr 20, 2011
Shiba Inu **Puppy** Cam @ Ustream.TV: Tune in daily to see the
cutest Shiba Inu pups... EVER! The six ...

▶ 0:32

cute **puppy** whistle - YouTube
youtube.com
Aug 1, 2007 - 32 sec - Uploaded by CA95207
Add to. Share Flag as inappropriate. Loading... Alert icon. Sign In
or Sign Up now ! Alert icon. Uploaded by ...

Image Search Results Page

Web Images Videos Shopping News Maps More | MSN Hotmail   Sign in ▼   Rewards ⊙   Mountain View,... ⚙

bing   puppies   🔍

Images   Web **Images** Videos News Local More▼

Puppies   Shih Tzu Puppies   Great Dane Puppies   Teacup Puppies   Basset Hound Puppies

SIZE
LAYOUT
COLOR
STYLE
PEOPLE

SEARCH HISTORY
Search more to see your history
See all
Clear all · Turn off

Select View: Large **Medium** Small | Browse trending image searches | SafeSearch: Moderate▼   3,920,000 results

If the landing page you are given to rate is a search engine page with an empty search box and no results displayed, then the page has no connection to the query and should get a rating of **Off-Topic or Useless**.

If the landing page is a set of results from a search engine, the page could be very helpful to users. Depending on how helpful the page would be, ratings can range from **Useful** to **Off-Topic or Useless**.

Here are some examples of search engine results pages that you might see in a URL rating task.

Query	Likely User Intent	Description of the Landing Page	Rating	Reason
[books about sharks], English (US)	Find books about sharks.	A book search results page from Google Books (books.google.com) which has a list of shark books to preview or read.	**Useful**	This page fits the intent of the query and has many good results.
[Pizza Hut in Chicago], English (US)	Find Pizza Hut locations in Chicago.	A maps search results page on Google Maps (maps.google.com) which provides a list of Pizza Hut locations in Chicago.	**Useful**	This page has contact information for every restaurant, as well as a map that displays their locations. This page fits the intent of the query and has many good results.
[wii console], English (US)	Purchase a Wii game console.	A shopping search results page on Google Product Search (products.google.com) which has many Wii console products for sale from different merchants.	**Useful**	This page provides links to merchants from which to buy this item. Prices and seller ratings are displayed. This page fits the intent of the query and has many good results.
[jumping shark], English (US)	Find videos or images of a jumping shark, or find information about the term "jumping the shark" that was used on several TV shows.	A video search results page on Google Video (video.google.com) which has some videos related to the video interpretation of the query, but a few unrelated videos as well.	**Relevant**	This page fits a likely intent of the query and has some good results.
[books about sharks], English (US)	Find books about sharks.	An image search results page from Google Images (images.google.com) showing images of sharks, as well as some pictures of covers of books about sharks.	**Slightly Relevant**	This page has images of books about sharks, and, with a couple of clicks, users can get to webpages which have information about the books or the books for sale. But book images are not really that helpful for the query. Most users are looking for books, not images of books. Few users would find this page helpful.

Query	Likely User Intent	Description of the Landing Page	Rating	Reason
[books about sharks], English (US)	Find books about sharks.	A maps search results page from Google Maps (maps.google.com) showing businesses and museums and other search results which are related to sharks (but not to books).	**Off-Topic or Useless**	This maps page has many search listings related to sharks, but none of the results are helpful for users. The results do not match the intent of the query.
[Pizza Hut in Chicago], English (US)	Find Pizza Hut locations in Chicago.	An image search results page on Google Images (images.google.com) showing images of the Pizza Hut logo and pictures of pizzas.	**Off-Topic or Useless**	Users want to find Pizza Hut restaurants in Chicago. The images on this page are **Off-Topic or Useless** because they are completely unhelpful for the user intent. This page does not fit the intent of the query.
[wii console], English (US)	Purchase a wii game console.	A shopping search results page on Google Product Search (products.google.com). This particular search results page does **not** have a helpful set of wii console products for users. It has one marginally related item, but all of the rest of the products are off-topic.	**Off-Topic or Useless**	The shopping results on the page are mostly off topic to the query. A shopping results page with the desired product would be helpful, but the results on this particular page are bad.
[books about sharks], English (US)	Find books about sharks.	Search engine pages where users would enter queries. No queries have yet been entered and no search results are displayed: http://www.bing.com http://www.google.com http://www.yahoo.com	**Off-Topic or Useless**	Since these pages do not show search results, they have nothing to do with the query and do not fit the intent of the query. Users would have to start their search again.

5.6.8 Video Landing Pages

Many landing pages with videos are easy to rate. When the query, the text on the landing page, and the video are all in the task language, an acceptable language, or English, assigning a utility rating and a Language Page Language flag should be very straightforward. Questions arise, however, when the query and/or video are in a foreign language.

The important thing to remember is that you should think about user intent and what pages are good for users. If the query "asks" for a foreign language song, band, film, sporting event, etc., then a video of the song, band, film, sporting event, etc. is helpful since it can probably be understood even though it is in a foreign language.

If the video is someone talking *about* the song, band, film, or event, the page probably cannot be understood and should be assigned **Unratable: Foreign Language**.

Here are some examples:

Query	URL of the Landing Page	Description of the Landing Page	Rating	Landing Page Language
[alex c], English (US)	http://www.youtube.com/watch?v=JSRh1vx-Vho	The query is for the German artist, Alex C. The landing page has a video sung by her in German. The navigation links are in English.	**Relevant** or **Useful**	English
[alex c], English (US)	http://www.youtube.com/watch?v=Pz-t5OZ-2yU	The query is for the German artist, Alex C. The landing page has a video sung by her in German.	**Relevant** or **Useful**	English
[mademoiselle k], English (US)	http://www.youtube.com/watch?v=VOr7-sxuifY	The query is for the French rock band, Mademoiselle K. The landing page has a video sung by the band in French.	**Relevant** or **Useful**	English
[Kasal, Kasali, Kasalo], English (US)	http://www.youtube.com/watch?v=_pSOmvx1hNg	The query is for Kasal, Kasali, Kasalo, a movie starring Judy Ann Santos. The landing page is a clip from the movie.	**Relevant** or **Useful**	English
[judy ann santos], English (US)	http://www.youtube.com/watch?v=E8vHX6pYYt4&feature=related	The query is for the popular Philippines actress, Judy Ann Santos. The landing page has a short trailer for "In My Life".	**Slightly Relevant** or **Relevant**	English
[beatles live], English (US)	http://www.youtube.com/watch?v=Ou__mIGfimU	The query is looking for information about or a video of a Beatles live performance. The landing page documents a visit by the Beatles to Tokyo. The spoken language on the video is mostly in Japanese. Since language is needed to evaluate utility, the landing page should be rated **Unratable: Foreign Language**.	**Unratable: Foreign Language**	Foreign Language

6.0 Flags

In addition to assigning a rating from the rating scale, you will also assign flags to mark special types of pages.

6.1 Spam Flag

You must decide if the page is should be assigned a **Spam** flag by looking for spam signals that you will learn about in the "Webspam Guidelines", Part 5 of the "General Guidelines".

Not Spam: If you do not believe that a page has been designed using deceptive web design techniques, you should assign a **Not Spam** flag.

Maybe Spam: If you find a page to be "spammy", but you do not feel comfortable saying that the webmaster definitely designed the page using deceptive web design techniques, you should assign a **Maybe Spam** flag.

Spam: If you believe that a page has been designed using the deceptive web design techniques described in the "Webspam Guidelines", you should assign a **Spam** flag.

If you choose either **Maybe Spam** or **Spam**, you must include a comment explaining why.

6.2 Pornography Flag

Please apply the **Porn** flag to all porn pages. A page will be considered porn if it has pornographic content, including porn images, links, text, pop-ups, and/or ads. An image may be considered porn in one culture or country, but not another. Please use your judgment and knowledge of the task location.

6.2.1 Clear Non-Porn Intent

If the user intent behind a query is clearly **not** pornographic, a landing page that serves a pornographic interpretation of the query and/or has porn for its main content should be rated **Off-Topic or Useless** and assigned a **Porn** flag. For example, consider the query [car pictures]. In any task language, a page showing a nude female reclining on the hood of a car should be rated **Off-Topic or Useless** and assigned a **Porn** flag, even though there is a car in the picture.

The reasons we are asking you to do this are the following:

- When the user intent is clearly not porn, a porn result that serves a "no chance" interpretation or has porn as its main content should be considered to have no utility.
- Uninvited porn is a very bad experience for many users and is an indication of poor search engine quality.

Here are some examples:

Query	Likely User Intent	Landing Page	Rating	Porn Flag?
[toys], English (US)	Find toys to buy	http://sextoyslut.com/maintour.php/4078/92/A Warning – this page is porn	**Off-Topic or Useless**	Yes
[how tall is a camel], English (US)	Find answer to this question about camels	http://www.xnxx.com/free/cameltoe-pictures.php Warning – this page is porn	**Off-Topic or Useless**	Yes
[car pictures], English (US)	Find pictures of cars	http://www.securitycamsfuck.com/ Warning – this page is porn	**Off-Topic or Useless**	Yes

Sometimes, the main content of a landing page **is** related to a likely query interpretation, but the page happens to display porn ads or porn text outside the main content portion of the page. In this case, you will usually assign a rating other than **Off-Topic.** Note: Since such pages are often lower quality pages, you will usually assign **Slightly Relevant** or **Relevant**.

Here are some examples:

Query	Likely User Intent	Landing Page	Rating	Porn Flag?
[toys], English (US)	Find toys to buy	Page with reviews of the 10 best toys for 5-6 year old boys. The page displays just a few ads, but does have one on the left side that shows an image of a woman's bare breasts. The main content is related to the query, so the page is not **Off-Topic.** However, since the reviews are so specific, they would be helpful for just a few or some users.	**Slightly Relevant**	Yes
[car pictures], English (US)	Find pictures of cars	Page with images of over 100 different cars for the current model year organized by brand, which has a link to a porn site on the right side of the page. The main content is related to the query, so the page is not **Off-Topic.** This collection of images would be helpful for some or many users.	**Relevant**	Yes

6.2.2 Possible Porn Intent

Some queries have both non-porn and porn interpretations. For example, all of the following English (US) queries are possible porn intent queries, but they also have a non-porn intent: [girls], [gay], [thong], [breast], [sex], [spanking]. We will call these queries "possible porn intent" queries.

For these queries, please assume that the non-porn interpretation is dominant, even if you think users are looking for porn. For example, please assume that the dominant interpretation of [spanking], English (US) is the discipline technique used by parents on a child (the non-porn interpretation). Rate the porn interpretation as a minor interpretation, even if you think most users are looking for porn.

Query	User Intent	Landing Page	Rating	Porn flag?
[spanking], English (US)	Find information about spanking children	http://umaine.edu/publications/4357e/	**Relevant**	No
[spanking], English (US)	Find information about spanking children	http://www.thespankingnews.com/ Warning – this page is porn	**Slightly Relevant**	Yes
[breasts], English (US)	Find anatomy or health information about breasts	http://en.wikipedia.org/wiki/Breast	**Useful**	No
[breasts], English (US)	Find anatomy or health information about breasts	http://www.boobsbee.com/ Warning – this page is porn	**Slightly Relevant**	Yes
[pictures of girls], English (US)	Find pictures of girls	http://www.worldofstock.com/stock_photos/PCH15980.php	**Relevant**	No
[pictures of girls], English (US)	Find pictures of girls	http://www.kindgirls.com/main Warning – this page is porn	**Slightly Relevant**	Yes

6.2.3 Clear Porn Intent

For very clear porn queries where no other intent is possible, assign a rating to the porn landing page using the rating scale without lowering the score. Even though there is porn intent, the page should still be assigned a **Porn** flag.

Please note that you should not simply rate all porn pages for porn queries as **Relevant** or **Useful**. Even though the query is porn and the result is porn, the page must fit the query to have utility and get a high rating.

Pages that provide a poor user experience - such as pages that try to download malicious software - should also receive low ratings, even if they have some images appropriate to the query.

Porn stars, porn movies, names of specific porn websites, etc., can have **Vital** pages. Be consistent in assigning a **Porn** flag to all porn pages, even when the rating is **Vital**.

Query	Likely User Intent	Landing Page	Rating	Porn Flag?
[freeones], English (US)	Navigate to the Freeones homepage	http://www.freeones.com/ Warning – this page is porn	**Vital**	Yes
[freeones], English (US)	Navigate to the Freeones homepage	http://www.baberoad.com/ Warning – this page is porn	**Off-Topic or Useless**	Yes
[jenna jameson], English (US)	Find porn pictures of Jenna Jameson or navigate to her official website.	http://www.jennajameson.com/ Warning – this page is porn	**Vital**	Yes
[jenna jameson], English (US)	Find porn pictures of Jenna Jameson or navigate to her official website.	http://www.bangbros.com Warning – this page is porn	**Off-Topic or Useless**	Yes
[anime sex pictures], English (US)	Find anime sex pictures	http://www.naughty.com/free-porn-sex-movies-videos/Anime-Videos.html Warning – this page is porn	**Relevant** or **Useful**	Yes
[cheerleader porn], English (US)	Find porn pictures of cheerleaders	http://www.porn365.com/Cheerleader.html l Warning – this page is porn	**Relevant** or **Useful**	Yes

Please do not assign a **Porn** flag to a non-porn page, just because the query has porn intent. If the landing page is not porn, it should not be flagged.

6.2.4 Reporting Illegal Images

Child Pornography and Bestiality

When working on rating projects in any task location, you must follow United States federal law, which considers child pornography and bestiality to be illegal.

Definition of Child Pornography

An image is child pornography if it is a visual depiction of someone who appears to be a minor (i.e., under 18 years old) engaged in sexually explicit conduct (e.g., vaginal or anal intercourse, oral sex, bestiality or masturbation as well as lascivious depictions of the genitals), or sadistic or masochistic abuse. The image of sexually explicit conduct can involve a real child; a computer-generated, morphed, composite or otherwise altered image that appears to be a child (think of images that have been altered using "Photoshop"); or an adult who appears to be a child; and the image can be nonphotographic -- e.g., drawings, cartoons, anime, paintings or sculptures – so long as the subject is engaging in sexually explicit conduct and which is obscene. If it is indistinguishable from child pornography, it is child pornography.

Even if the image has literary (think of the famous book "Lolita"), artistic, political (think of political cartoons), or scientific (think of images for a medical text book) value, please send the link to your employer (as instructed below).

Depiction of the genitals does not require the genitals to be uncovered. Thus, for example, a video of underage teenage girls dancing erotically, with multiple close-up shots of their covered genitals, or images of children with opaque underwear that focus on the genitalia could be considered child pornography.

An image of a naked child (e.g., in the bathtub or at a nudist colony) is not considered child pornography as long as the child is not engaging in sexually explicit conduct, or the focus is not on the child's genitalia.

Visual depictions of adults who look like adults (e.g., a 35 year old man play-acting in diapers, or an obvious woman dressed as a school girl) are not child pornography. (If you don't think it's a minor, it probably isn't child pornography.) However, if you cannot tell that the person in the image is over 18 (e.g., an under-developed 18 year old whose body hair has been waxed), that is child pornography.

Definition of Bestiality

Bestiality or zoophilia is defined as human-animal sexual interaction.

Reporting Instructions

Leapforce Evaluators: Please use the Contact form located on the Leapforce At Home website (http://www.leapforceathome.com). Select the 'Report illegal images and/or content' topic from the topic selection box. Your report will automatically be forwarded to the correct group.

Lionbridge Raters: Please follow the instructions provided on the Lionbridge rater portal for reporting illegal or offensive images.

6.3   Malicious Flag

A page should be assigned a **Malicious** flag if:

- You are forced to quit your Firefox browser due to prompts that keep coming back and will not go away.
- There are attempts to download spyware, Trojans, viruses, etc.

Please note that pop-ups that you are able to close are not malicious, even if it takes a couple of tries to get rid of them.

Please do not assign a **Malicious** flag just because the browser gives you a warning message or certificate acceptance request. Assign a **Malicious** flag only under the conditions listed above. If you encounter a page with a warning message, such as "Warning-visiting this web site may harm your computer," or if your antivirus software warns you about a page, you should not try to visit the page to assign a rating. You should instead assign a rating of **Unratable: Didn't Load**.

6.4   Compatibility between Ratings and Flags

Please be aware that **Unratable** pages can be assigned **Spam**, **Porn**, and/or **Malicious** flags. Here are some examples:

- The page is in a foreign language, but has porn images.
- The page is in a foreign language, but there is hidden text.
- The page doesn't load, but you can tell from the URL that it is a sneaky redirect.
- The page doesn't load, but has porn ads.
- The page is in a foreign language, but you cannot close a pop-up on the page and you are forced to quit your Firefox browser.

Part 2: URL Rating Tasks with User Locations

1.0   Important Definitions

Some of the definitions listed below are the same as or similar to the definitions introduced in Section 1.2 of the Rating Guidelines.

Query: A query is the set of word(s), number(s), and/or symbol(s) that a user types in the search box of a search engine. We will sometimes refer to this set of words, numbers, or symbols as the "query terms".

User intent: When a user types a query, he is trying to accomplish something, such as finding information or purchasing an item online. We refer to this goal as the user intent.

Task Language and Task Location: Queries have a task language and task location associated with them and look like this: **[digital cameras], Spanish (ES)**. This format indicates that the query *digital cameras* was typed into a search box by a Spanish reading user in Spain. Task locations are represented by a two-letter country code. The country code for Spain is ES. If the query had been typed by a Spanish reading user in Mexico, it would look like this: **[digital cameras], Spanish (MX)**. The Task Location is the location of users issuing the query.

User Location: In addition to the Task Location, some URL rating tasks have a User Location (sometimes called a Query Location). The User Location is the location of users issuing the query. The User Location is usually a city or postal code.

Explicit Location: Some queries include a location. When a location is included in the query, it is referred to as an Explicit Location. The Explicit Location may be the name of a town, the name of a city, a street intersection, a postal code, etc. Here are some examples of queries with Explicit Locations (in bold type): [drugstore **78703**], [starbucks, **Albany, NY**], and [Chinese restaurants **Mountain View**].

Local intent: Some queries have "local intent", which means that users are looking for something nearby. Local intent queries include businesses, restaurants, supermarkets, coffee shops, and other things people *expect to find nearby*. Local intent queries can also be for local information, such as weather or events in the user's location. Here are some examples of local intent queries: [pizza], [weather], [gas stations], [cooking classes], [movie showtimes], and [coffee shops near me].

Non-local intent: Some queries have "non-local intent", which means that users are not specifically looking for something nearby or information that is about their location. Many queries are non-local intent queries. Here are some examples of non-local intent queries: [google earth download], [dukan diet], [how to get chapstick out of clothes], and [www.yahoo.com]. A query can have both local and non-local intent.

You will learn more about these definitions in the following sections.

1.1   What is the User Location?

All URL rating tasks have a Task Location and a Task Language. For example, the query [football] English (US) has a Task Language of English and a Task Location of the United States (US). Most Task Locations are countries.

As a reminder, the Task Location is the location of users issuing the query. You can imagine users sitting at computers or laptops inside the US typing the query [football] into a search box of a search engine, like Bing or Google or Yahoo, etc.

Some URL rating tasks also have a User Location (sometimes called a Query Location), which is a more specific geographic location of users issuing the query. The User Location might be a city, a postal code, a state, etc.   The User Location will always be inside the Task Location.

So you can think of the User Location as a more specific form of Task Location. For example, for the query [football], English (US) with the User Location of Austin, TX, you can imagine users in Austin, Texas typing the query [football] into the search box of a search engine.

1.2   Why are the Task Location and User Location important?

Recall the Task Location discussion of the query [football] English (US) vs. [football] English UK (in Section 2.2 of the Rating Guidelines). Football in the US refers to a different game than football in the UK. So the query interpretation and user intent are different for the query [football], English US vs. [football], English UK.

In addition to understanding the query and user intent, the Task Location is also important in assigning ratings. For example, US and UK users generally use different shopping websites when purchasing products. Most US shopping websites show prices in dollars and have US shipping rates. UK shopping websites show prices in pounds and have shipping in the UK. So a US shopping result would be rated differently for the query [boys jeans], English US vs. the query [boys jeans], English UK.

Likewise, the User Location may change your understanding of user intent and/or the rating you assign to landing pages. This section of the General Guidelines explains how to understand queries and assign ratings when the rating task has a User Location.

1.3   User Location, Task Location, and Explicit Location in the query

In addition to the User Location and Task Location, there is one more location that may be part of a URL rating task. The query terms may include a location; for example: [rental cars los angeles], English (US). The location Los Angeles is explicitly mentioned in this query. If a location is included as part of the query text, we will call this an Explicit Location. The query [rental cars Los Angeles], English (US) has an Explicit Location of Los Angeles.

Important: All URL rating tasks have a Task Location. Some URL rating tasks have a User Location, which is a city or region inside the Task Location. Some URL rating tasks have an Explicit Location included in the text of the query. And some tasks have all three types of locations!

There are three big differences between an Explicit Location typed as part of a query by the user vs. Task and User Locations included as part of your rating task:

1) **Where do these locations come from?** Task and User Locations tell you where users are located. You can imagine users sitting at a computer in the Task and User Locations, searching the Internet. The User and Task Location is included as part of the evaluation task information because we want you to rate from the perspective of users in Task and User Location. Explicit Locations are typed directly into the search box of a search engine by a user.

2) **Where can these locations be?** The User Location should always be inside the Task Location, and you should only be given rating tasks for the Task Location and Task Language you are qualified to rate.   However, users search for things all over the world, sometimes including an Explicit Location to tell a search engine what they are trying to find. Explicit locations can be for any location, anywhere.

3) **What role do these locations play in rating?** In many cases, the User Location does not change our understanding of query interpretation, user intent, or the utility of a result. The User Location is additional information in the rating task, and you will have to use your judgment in deciding how to interpret and use it. However, the Explicit Location typed by users as part of the query is extremely important and should always play a role in understanding and interpreting the query.

Please take a moment to think about these examples. Please note that all of these examples have a Task Location, because all queries have a Task Location, and that some examples also have a User Location and/or an Explicit Location.

Query	Task Location	User Location	Explicit Location	Discussion
[hotels]	English (US)	None	None	Most URL rating tasks do not have a User Location or an Explicit Location in the query. [hotels] is a very broad query.
[hotels]	English (US)	Houston, Texas	None	Here is an example of a User Location (Houston, Texas) which is inside of the Task Location (the US). Even with a User Location of Houston, [hotels] is still a really broad query. Users in Houston search for hotels in locations all over the world. The User Location does not affect our understanding of this query.
[hotels in San Francisco]	English (US)	None	San Francisco	This is an example of an Explicit Location in the query. The Explicit Location makes this query more specific and tells us that users are looking for hotels inside San Francisco. Note: Users in the US may search for hotels in the US and also in other parts of the world.
[hotels in San Francisco]	English (US)	Houston, Texas	San Francisco	Here is an example of a query with a Task Location (US), a User Location (Houston), and an Explicit Location (San Francisco). Again, the Explicit Location makes this query more specific and tells us that users are looking for hotels inside San Francisco. The User Location does not help us understand the query. The Explicit Location is important, not where the user was when the query was typed.
[hotels in Tokyo Japan]	English (US)	None	Tokyo, Japan	The Explicit Location (Tokyo Japan) is outside the Task Location (US). Explicit Locations can be anywhere in the world.
[hotels in Tokyo Japan]	English (US)	Houston, Texas	Tokyo, Japan	The Explicit Location is outside the Task and User Location. As before, the User Location does not affect our understanding of this query. The Explicit Location is important, not where the user was when the query was typed.

*Important: When **researching the query** to help you understand likely query interpretation/user intent, it can sometimes be helpful to do a search after "adding" the User Location to the query. For example, if the query is [nathan's], English (US) with a User Location of Elk Grove, CA, issuing the query [nathan's elk grove CA] will help you understand that users are possibly looking for Nathan's Chinese Cuisine, a Chinese restaurant in Elk Grove, California. Without adding "elk grove CA" to the query, you probably would not learn about this likely query interpretation.*

*However, when **assigning your final rating**, you cannot just "mentally add" the User Location to the query. The query [hotels], English (US) with a User Location of Houston TX has a very different intent than [hotels Houston, TX], English (US). In the first case, Houston users could be looking for hotels anywhere in the world. In the second case, US users are looking for hotels **in** Houston, Texas. If this is confusing, please review the example table above.*

2.0 Location-Specific Rating Task Screenshot

The Location-Specific URL rating task page is similar to the standard URL Rating task page, except that it displays additional information associated with the User Location.

Task Type	Screenshot		Description
This is *not* a location-specific task because it does *not* have a User Location. Notice, however, that an Explicit Location has been specified in the query.	**Query**	pizza hut san francisco	The user wants Pizza Hut information for the San Francisco area.
	URL	http://www.yelp.com/biz/pizza-hut-san-francisco	
	Task Location	United States (US)	
	Task Language	English	
	Other Acceptable Languages	*None*	
This is a location-specific task because it has a User Location.	**Query**	pizza hut	The query was issued by a user located in San Francisco. We can assume that the user is looking for a Pizza Hut restaurant in San Francisco.
	User Location	San Francisco, CA	
	URL	http://www.yelp.com/biz/pizza-hut-san-francisco	
	Task Location	United States (US)	
	Task Language	English	
	Other Acceptable Languages	*None*	
This is also a location-specific task because it has a User Location. Notice, however, that an Explicit Location has been specified in the query.	**Query**	pizza hut san francisco	The query was issued by a user located in New York. However, because the query contains "san francisco", we know that the user is looking for Pizza Hut restaurants in the San Francisco area, even though the User Location is New York.
	User Location	New York, NY	
	URL	http://www.yelp.com/biz/pizza-hut-san-francisco	
	Task Location	United States (US)	
	Task Language	English	
	Other Acceptable Languages	*None*	

Information	Standard URL Rating Task Page	Location-Specific URL Rating Task Page
User Location	Standard URL Rating task home does not have this information.	***** New York ***** ***** 90210 ***** ***** Dallas, TX ***** ***** TX *****

Location-Specific URL Rating Task Page

rater homepage → rating task   johndoe@gmail.com [rater homepage · recently completed tasks · logout]
Language: **English (US)**

Rating Task - icq

[search results: google] ·   [release task]

Query	icq
User Location	***** San Francisco, CA *****
Query Description	This is a location-specific rating task for the User Location described above.
URL	http://www.mobicq.info/
Task Location	United States (US)
Task Language	English
Other Acceptable Languages	*None*

As you know, understanding user intent is one of the most important steps in assigning a rating to a landing page. In this section, we will discuss the role of the User Location in understanding query interpretation and user intent.

For many or most queries, a User Location does not affect our understanding of user intent for a query. For example, most information or "know" queries, such as [how to get chapstick out of clothes], English (US) or [einstein's theory of relativity], English (US) or [address of the white house], English (US), have the same user intent with or without a User Location. In other words, users in all cities and towns in the US who issue these queries probably have the same or similar user intent.

Now consider these queries:

[dmv], English (US) with no User Location
[dmv] English (US) with a User Location of Williamsburg, VA

DMV stands for Department of Motor Vehicles, and many states in the US have a Department of Motor Vehicles. If the user is located in Virginia, it is very likely that he or she is looking for the Virginia DMV website. It is very unlikely that users in Virginia are looking for the DMV website of a different state, unless the query explicitly included a different state. So, for the query [dmv], a User Location of Virginia affects our understanding of the query interpretation and user intent.

When is the User Location important in understanding user intent? **Sometimes.** Please use both Web research and your personal judgment to answer this question. Ask yourself, "Would users in one city or state be looking for something different than users in another?" For many or most queries, the answer to that question is "no".

When are Explicit Locations important in understanding user intent? **Always!** All words in the query are important for understanding user intent.

Important: *Many raters overemphasize the importance of User Location when interpreting queries and user intent. Please carefully think about the User Location and whether it should affect your understanding of the query interpretation and user intent.*

Here are some examples:

Query and Task Location	User Location	Query Interpretation/User Intent
[facebook], English (US)	None	The likely user intent is to navigate to facebook.com.
[facebook], English (US)	Detroit, MI	The likely user intent is to navigate to facebook.com. User Location does not affect our understanding of this query. Users in different US cities and towns are looking for the same website.
[pictures of kittens], English (US)	None	The user intent is to find pictures of kittens.
[pictures of kittens], English (US)	Washington, D.C.	The user intent is to find pictures of kittens. The User Location does not affect our understanding of intent. Users in different US cities and towns are looking for the same or similar kinds of pictures. Note: There is nothing about this query that suggests users are looking for kittens in Washington D. C., which is a common rating error. Please do not try to make the User Location important when it is not.

Query and Task Location	User Location	Query Interpretation/User Intent
[coffee shops], English (US)	None	It is likely that the user is looking for a coffee shop nearby. We don't know where in the US the user is located.
[coffee shops], English (US)	Boston, MA	It is likely that the user is looking for a coffee shop in Boston, MA. Note: Users in different locations are probably interested in coffee shops in their own location. Most users do not want to travel to another city or state to get a cup of coffee. In this example, the User Location helps us better understand what the user is looking for.
[coffee shops in Boston], English (US)	None	This query includes an Explicit Location and the intent is clear. The user is looking for coffee shops in Boston.
[barack obama], English (US)	None	The user probably wants to find news, current events, biographical information about, or photos of Barack Obama.
[barack obama], English (US)	Newton Falls, OH	The user probably wants to find news, biographical information, photos, or other content relating to Barack Obama. Except in the extremely rare event that Barack Obama is making an appearance in the tiny town of Newton Falls right at the time the task is being rated, the User Location should not change our understanding of user intent/query interpretation. Most of the time, users all over the US are looking for the same types of results for the query Barack Obama.
[walmart], English (US)	None	Walmart is a very popular store in the United States. Users may want to navigate to the website walmart.com to shop online or find information about Walmart products. It is also possible that users want information about a nearby Walmart store or want to visit a Walmart store in person.
[walmart], English (US)	Birmingham, AL	Walmart is a very popular store in the United States. Users in Birmingham AL may want to navigate to the website walmart.com to shop online or find information about Walmart products. It is also possible that users want information about a nearby Walmart store or want to visit a Walmart store in person. The User Location of Birmingham affects our understanding of which Walmart stores users might be interested in.
[walmart on Memorial Rd Oklahoma City]	None	This query has an Explicit Location. The user intent is very clear: users are looking for information about a Walmart store on Memorial Rd in Oklahoma City, such as the address or phone number.
[walmart on Memorial Rd Oklahoma City]	Portland, OR	This query has an Explicit Location. The user intent is very clear: users are looking for information about a Walmart store on Memorial Rd in Oklahoma City, such as the address or phone number. The User Location does not play any role in query interpretation or user intent. Note: you might find it odd that a user in Portland, Oregon would be looking for information about a Walmart store in a far away state, but we need to respect the user intent of the query as written.

Query and Task Location	User Location	Query Interpretation/User Intent
[turmeric], English (US)	None	It is likely that users are looking for information about the spice turmeric.
[turmeric], English (US)	Sunnyvale, CA	In Sunnyvale, there is a very popular restaurant named Turmeric. For the User Location of Sunnyvale, we will consider both query interpretations (the spice and the restaurant) to be common interpretations. This is a case where users in Sunnyvale may be looking for something different than users in other cities and towns in the US.
[turmeric], English (US)	Lincoln, NE	There is no restaurant or other business named Turmeric in Lincoln Nebraska. It is likely that users in Lincoln are looking for information about the spice turmeric. In this case, there is no difference between having a User Location of Lincoln, NE vs. having no User Location. In most cities and towns in the US, there is no restaurant or other business named Turmeric.
[dmv], English (US)	None	Users probably want to navigate to the DMV website of the state they live in. Users may also want to find information about the local DMV office. We do not know which DMV website users are interested in since there are different DMV websites for different states.
[dmv], English (US)	Williamsburg, VA	Users may want to navigate to the Virginia DMV website. This website has helpful information and functionality for all users in Virginia. Users may also want to find information about the Williamsburg DMV office. Users in different states and cities will be interested in different DMV websites and offices. The User Location informs our understanding of query interpretation and user intent.
[Virginia dmv], English (US)	None	Users probably want to navigate to the Virginia DMV website. This query has an Explicit Location.
[pizza greenville], English (US)	None	It is very likely that the user lives in a town named Greenville and wants to find a pizza place nearby. However, there are many towns inside the US named Greenville. This is an ambiguous query.
[pizza greenville], English (US)	Greenville, AL	It is very likely that the user is looking for a pizza place in Greenville, Alabama. In this case, the User Location helps us interpret the query - users are trying to find a pizza place in Greenville, Alabama.

3.1 Queries with Local Intent

Some queries seem to "ask" for webpages about locations or businesses near the user. For example, the user intent for the query [coffee shops nearby], English (US) is to find coffee shops near the user's location. We will call these queries "local intent" queries. Users are trying to find something 3.1- often a business, a product, etc.

There are also some queries that seem to "ask" for local information. For example, the user intent for [movie showtimes], English (US) is probably to find out what time movies are showing nearby. The user intent for the query [weather] is probably to find out about the current weather where the user is located. Users are interested in movies and weather *in their location*. We will consider these to be "local intent" queries as well.

Many or even most queries have no local intent. For example, [take an online personality test], English (US) is *not* a local intent query. Users are clearly looking for websites or webpages that will allow them to take a personality test. Users are *not* looking for a business or testing center near their physical vicinity (note the word "online" in the query). And personality tests are not different for users in different cities inside the US.

The query [amazon], English (US) does not have local intent. Users all over the US want to navigate to the same Amazon website.

The query [how many calories in a McDonalds hamburger], English (US) does not have local intent. Users are trying to find information that is not specific to a particular location in the US. The calorie information is the same for any McDonald's restaurant inside the US.

Finally, some queries have both a local and a non-local intent. The query [target] may be a navigational query for target.com, or the user may be looking for a Target store nearby. On the other hand, [target store nearest me] is a very clear local intent query.

Whether a query has local intent or not can sometimes be difficult to determine. We do not really know what a user is actually trying to do -- and different users may be looking for different things. However, in most cases, your understanding of local intent comes from the query itself. Queries such as [pizza] and [coffee shops] have clear local intent whether or not the rating task has a User Location.

However, sometimes the User Location can change your understanding of user intent and therefore local intent. For example, refer back to the query [turmeric], English (US) in the table above. Without a User Location, [turmeric] does not have a likely local intent, since there are very few stores or restaurants named Turmeric in the US. However, [turmeric], English (US), with the User Location Sunnyvale, CA, has both local and non-local intent, since there is a popular restaurant named Turmeric in Sunnyvale. Of course, users in Sunnyvale could still be looking for information about the spice.

Sometimes, users will add their own location to the query explicitly when they are looking for local results. For example, imagine users in Mountain View California who want to order pizza. Some users might type [pizza] into Google or Bing or other search engines, but some users add the city or zip code to get local results by typing [pizza Mountain View] or [pizza 94043].

Warning: Whether a query has local intent or not is determined by the query and the user intent behind the query, not whether the evaluation task information includes a User Location. Many poor ratings have resulted when raters assume that the query has local intent just because there is a User Location present in the evaluation task.

Query	User Location	Local Intent?	User Intent
[pizza], English (US)	None	Yes	It is likely that users want to find a pizza place nearby. This is a local intent query.
[pizza mountain view ca], English (US)	None	Yes	Most people search for pizza places nearby. We will assume this query has local intent, even though we do not know for sure that the user is located in Mountain View.
[pizza 94043], English (US)	Mountain View, CA	Yes	In this case, the user added a zip code that is inside or near his or her location, probably to ensure local search results. This is a local intent query.
[hotels], English (US)	None	No	Users often look for hotels outside their own location when planning trips.
[hotels], English (US)	Houston, TX	No	Users often look for hotels outside their own location when planning trips.

Query	User Location	Local Intent?	User Intent
[movie showtimes] , English (US)	None	Yes	Users usually look for show times for theaters nearby. Most people do not want to drive a long way to see a movie. This query has local intent.
[movie showtimes] , English (US)	Chapel Hill, NC	Yes	The query [movie showtimes] has local intent for any User Location. The presence of a User Location in the rating task is not the reason we consider this a local intent query.
[movie reviews] , English (US)	Chapel Hill, NC	No	Users are looking for movie reviews, perhaps trying to find a movie they are interested in seeing. Most movies play nationally, and good reviews are written on many websites. This query has non-local intent even though users probably want to see a movie nearby. The query is not for a local movie theater or for local movie showtimes; it is for movie reviews.
[cooking classes] , English (US)	None	Yes	Users are probably interested in finding a cooking class they can attend in person nearby. This query has local intent with or without a User Location.
[apple pie recipe] , English (US)	Salt Lake City, UT	No	Users are probably looking for an online recipe.
[walmart] , English (US)	None	Both non-local and local intent	Non-local intent: Users may want to navigate to the Walmart website. Local intent: Users may want to visit a Walmart store in person.
[walmart] , English (US)	Gainesville, FL	Both non-local and local intent	Non-local intent: Users may want to navigate to the Walmart website. Local intent: Users may want to visit a Walmart store in person, probably in Gainesville.
[shower curtain] , English (US)	Seattle, WA	Both non-local and local intent	Non-local intent: Users may want to purchase a shower curtain online. Local intent: Users may want to purchase a shower curtain locally in the Seattle area.
[shower curtain seattle] , English (US)	Seattle, WA	Yes	The user probably included his or her location in the query to get local information on where to buy a shower curtain in Seattle. This indicates users are probably looking for local results.
[how big is a shower curtain] , English (US)	Seattle, WA	No	This is an informational query with no local intent. Shower curtains in Seattle are the same size as shower curtains in other cities.

3.2   Rating Landing Pages when the task has a User Location

Now we will turn our attention to how to rate landing pages for tasks that have a User Location.

As discussed above, for most queries, the User Location does not make a difference in query interpretation or user intent. In fact, in many cases, the User Location does not make a difference in assigning ratings. Why? Most websites serve users from all over the Task Location. Many websites are not specific to a particular location or region. So for many URL rating tasks with a User Location, the rating of a landing page may be the same with or without a User Location.

However, here are the two situations when the User Location plays a role in assigning a utility rating:

- The User Location affects your understanding of user intent/query interpretation. Understanding user intent and query interpretation is part of the process of assigning a utility rating for all queries.
- When the query has local intent, the helpfulness of results depends on the location of the user. Knowing the User Location will change the way you rate landing pages for local intent queries such as [pizza], [coffee shops], etc.

In the first case, when the User Location has informed your understanding of query interpretation/user intent, then use that understanding to assign utility ratings as described in the General Guidelines as a whole.

When the query has local intent, the most helpful landing pages have information specifically for users in the User Location. For example, if the query is [pizza], then pizza places nearby the user are very helpful. If the query is [cooking classes], then information about cooking classes near the user is very helpful. When we have information about the User Location for a local-intent query, we can assign a rating that takes into account how close the pizza place or cooking class is to the user.

In general, for local intent queries, results in or near the User Location are the most helpful. How close is "near"? Most people are not willing to travel very far for a gas station, coffee shop, supermarket, etc. Those are types of businesses that most users expect to find very nearby.

Users might be willing to travel a little farther for certain kinds of local results: doctors' offices, libraries, specific types of restaurants, public facilities like swimming pools, hiking trails in open spaces, etc.

Sometimes, queries have local intent but users may accept results that are even farther away. Many people would be willing to travel to a nearby city for a very specialized shop or a very particular kind of item. In some cases, people might be willing to travel hundreds of miles -- for example, to see a doctor with an unusual specialty or to purchase a breed of dog that is not raised locally.

In other words, when we say users are looking for results "nearby", the word "nearby" can mean different distances for different queries. So as always, please use your judgment.

3.3 Vital Ratings for Rating Tasks with User Locations

For the most part, **Vital** ratings are unchanged by User Location. However, there are a few special situations worth mentioning.

Sometimes, the additional information and understanding provided by the User Location allow a landing page to receive the **Vital** rating because we now understand the user intent. Here is an example of this relatively rare case:

- [belmont library], English (US) – Users are looking for a library, but there are multiple schools and cities with Belmont in their name. So it is unclear which Belmont library the user intends to find. No **Vital** rating is possible.
- [belmont library], English (US) with a User Location of Nashville, TN - With this User Location, we can give the **Vital** rating to the official homepage of the Belmont University Library, which is located in Nashville.

In most cases, the User Location does not affect the **Vital** rating. For example, the target of the navigational query [nytimes.com] is the landing page of nytimes.com, whether or not the evaluation task has a User Location.

3.4 Rating Examples

Here are some examples that demonstrate how to rate URLs when a User Location is specified.

Query	User Location	Likely User Intent	URL of the Landing Page	Rating	Explanation
[chinese restaurants] English (US)	Mountain View, CA	This query has local intent. The likely user intent is to find out about Chinese restaurants in the city of Mountain View, California	http://local.yahoo.com/CA/Mountain+View/Food+Dining/Restaurants/Chinese+Restaurants	**Useful**	The landing page is a fairly comprehensive list of Chinese restaurants in Mountain View. In addition, many restaurants have reviews. There is a map, and there are helpful features to narrow down the restaurants by features such as take out or family friendly.
[chinese restaurants] English (US)	Mountain View, CA	This query has local intent. The likely user intent is to find out about Chinese restaurants in the city of Mountain View, California.	http://www.hunanchilimountainview.com/	**Relevant**	The landing page is the official website of Hunan Chili, a Chinese restaurant in the User Location. It should be rated as **Relevant**. (Note: A prominent or popular or famous Chinese restaurant in the User Location should be rated **Useful**. How can you tell if a restaurant is popular? Try researching on business listing websites and see which restaurants have a large number of positive reviews.)
[chinese restaurants] English (US)	Mountain View, CA	This query has local intent. The likely user intent is to find out about Chinese restaurants in the city of Mountain View, California	http://www.restaurants.com/california/mountain-view	**Off-Topic or Useless**	This is a general listing of restaurants in Mountain View, and there do not appear to be any Chinese restaurants on the first page of this list. In fact, there is no information telling users what type of cuisine is served by each restaurant. Additionally, there are no features to narrow the list down to Chinese restaurants only. This page is not helpful for the query. Page level checks and website level checks reveal that this is a low quality page.
[chinese restaurants] English (US)	Mountain View, CA	This query has local intent. The likely user intent is to find out about Chinese restaurants in the city of Mountain View, California.	http://pekingduckhousenyc.com/	**Off-Topic or Useless**	The landing page is for a Chinese restaurant in New York City. It is unlikely a user would find this.

Query	User Location	Likely User Intent	URL of the Landing Page	Rating	Explanation
[target] English (US)	none	This query has both local and non-local intent. Users may want to shop online or visit a store nearby.	http://www.target.com/	**Vital**	Target has a popular website and many users want to navigate to target.com. We'll consider the landing page of target.com to be **Vital** for this query.
[target] English (US)	Keizer, OR	This query has both local and non-local intent. Users may want to shop online or visit a store nearby.	http://www.target.com/	**Vital**	Many users want to navigate to the target website to shop or find information. There is strong non-local intent for this query. The landing page of target.com is **Vital** for the query [target], even when the evaluation task has a user location.
[target] English (US)	Keizer, OR	This query has both local and non-local intent. Users may want to shop online or visit a store nearby.	Keizer Target link	**Useful**	We don't know whether users are interested in navigating to target.com to shop or find information online, or whether users want to find a Target location nearby. Target is a very popular online retailer. Because there are so many users who are interested in shopping online or using the target.com website, we will consider the landing page of target.com to be **Vital** for the query [target], even when the evaluation task has a User Location. The landing page is the page for the Target store in Keizer, Oregon. There is a lot of helpful information on this subpage on Target's official website, including store hours, phone number, address, and a map of the store's location.
[target keizer]	Keizer, OR	This query has local intent. Users are looking for the Keizer location of the Target store.	Keizer Target link	**Vital**	Users are looking for the Keizer location of Target, and this page is the official target.com website page for the Keizer location.
[target] English (US)	Keizer, OR	This query has both local and non-local intent. Users may want to shop online or visit a store nearby.	http://www.yelp.com/biz/target-stores-keizer	**Useful** or **Relevant**	This is a business listing/review page for the Target in Keizer, Oregon. There is helpful information here. In addition to an address and phone number, there are reviews of the store, a map, and a link to the Target homepage.
[target] English (US)	Keizer, OR	This query has both local and non-local intent. Users may want to shop online or visit a store nearby.	http://www.merchantcircle.com/business/Target.503-856-0614	**Slightly Relevant**	There is very little information on this business listing page for the Target in Keizer, Oregon: just an address and phone number. In addition, the page has poor layout and is low quality: there is very little main content and ads are prominently placed.

Query	User Location	Likely User Intent	URL of the Landing Page	Rating	Explanation
[pictures of kittens] English (US)	Pittsburgh, PA	Users are looking for pictures of kittens. This is a non-local intent query. There is no obvious user intent to find pictures of kittens in Pittsburgh. The User Location plays no role in the utility rating.	http://www.bing.com/images/search?q=pictures+of+kittens&qpvt=pictures+of+kittens	**Useful**	This is a page full of kitten pictures.
[pictures of kittens] English (US)	Pittsburgh, PA	Users are looking for pictures of kittens. This is a non-local intent query. There is no obvious user intent to find pictures of kittens in Pittsburgh. The User Location plays no role in the utility rating.	http://pittsburgh.craigslist.org/pet/	**Off-Topic or Useless**	This is a local listing of pets needing homes in the Pittsburgh area. There are no pictures of any pets directly on this page and few pictures on the individual listings.
[pizza] English (US)	Mountain View, CA	Most users want to find a pizza place nearby. This query has clear local intent, whether or not the task has a User Location.	http://pizzamyheart.com/	**Useful**	This is a popular pizza shop in Mountain View, CA. (How can you tell if a restaurant is popular? Try researching on business listing websites and see which restaurants have a large number of positive reviews.)
[pizza] English (US)	Mountain View, CA	Most users want to find a pizza place nearby. This query has clear local intent, whether or not the task has a User Location.	http://www.pitfirepizza.com/	**Off-Topic or Useless**	This is a pizza shop in Los Angeles. Most users expect to find a pizza place near their location. This pizza restaurant in a far away city is not helpful for users in Mountain View, CA.
[pizza] English (US)	None	Most users want to find a pizza place nearby. This query has clear local intent, whether or not the task has a User Location.	http://dominos.com/	**Useful**	This is a popular national pizza chain which serves many locations in the US.
[pizza] English (US)	None	Most users want to find a pizza place nearby. This query has clear local intent, whether or not the task has a User Location.	http://en.wikipedia.org/wiki/Pizza	**Relevant** or **Slightly Relevant**	Most users are looking for a pizza shop nearby. However some users (probably few) are interested in pizza recipes, pizza history, and other pizza information. Because this is a very good result for a minor interpretation, a rating of **Relevant** or **Slightly Relevant** is appropriate.
[pizza] English (US)	None	Most users want to find a pizza place nearby. This query has clear local intent, whether or not the task has a User Location.	http://www.epicurious.com/recipes/food/views/Pizza-Dough-108197	**Slightly Relevant** or **Off-Topic or Useless**	This is a recipe page for pizza dough on a well known recipe website. There is no indication from the query that users are looking for pizza dough recipes.

Query	User Location	Likely User Intent	URL of the Landing Page	Rating	Explanation
[101.9], English (US)	Chicago, Illinois	Most users are looking for the radio station in Chicago which is at 101.9 radio frequency. The query has clear local intent, whether or not the task has a User Location.	http://www.wtmx.com/	**Vital**	This is the homepage of the radio station in Chicago with radio frequency that matches the query.
[101.9], English (US)	Chicago, Illinois	Most users are looking for the radio station in Chicago which is at 101.9 radio frequency. The query has clear local intent, whether or not the task has a User Location.	http://en.wikipedia.org/wiki/WTMX	**Relevant** or **Slightly Relevant**	This is an encyclopedia page about the 101.9 radio station in Chicago. It would be helpful for some or few users.
[101.9], English (US)	Chicago, Illinois	Most users are looking for the radio station in Chicago which is at 101.9 radio frequency. The query has clear local intent, whether or not the task has a User Location.	http://www.q1019.com/	**Off-Topic or Useless**	This is the homepage of a radio station in San Antonio, Texas with radio frequency that matches the query. This is not the radio station that users in Chicago are looking for.
[101.9], English (US)	None	Most users are looking for a radio station which is at 101.9 radio frequency. The query has clear local intent, whether or not the task has a User Location.	http://www.wtmx.com/	**Relevant** or **Slightly Relevant**	There are many radio stations in the US with this radio frequency (http://en.wikipedia.org/wiki/101.9_FM). This query has a different meaning to users in different User Locations. We do not know where the user is located. The homepage of the 101.9 radio station in Chicago would be helpful for some users in the US. This is a 101.9 station from a major metropolitan area. Also, this station streams music, so users from outside of Chicago could use this webpage to listen to music.

Query	User Location	Likely User Intent	URL of the Landing Page	Rating	Explanation
[is passing on the right legal], English (US)	Wichita, Kansas	Find out whether it is legal to pass a car when driving in the right hand lane of a road or highway. The user is most likely interested in the law about Kansas, where he or she is located.	http://kansasstatutes.lesterama.org/Chapter_8/Article_15/8-1517.html	**Useful**	This is a copy of the law about passing on the right in the state of Kansas.
[is passing on the right legal], English (US)	Wichita, Kansas	Find out whether it is legal to pass a car when driving in the right hand lane of a road or highway. The user is most likely interested in the law about Kansas, where he or she is located.	http://www.leg.state.or.us/05orlaws/sess0300.dir/0316ses.htm	**Off-Topic or Useless**	This is a copy of the law about passing on the right in the state of Oregon. This page is helpful for extremely few or no users.
[is passing on the right legal], English (US)	Wichita, Kansas	Find out whether it is legal to pass a car when driving in the right hand lane of a road or highway.	http://www.mit.edu/~jfc/right.html	**Useful**	This is a very succinct summary of passing laws that is helpful for users in any state.
[is passing on the right legal], English (US)	None	Find out whether it is legal to pass a car when driving in the right hand lane of a road or highway.	http://www.leg.state.or.us/05orlaws/sess0300.dir/0316ses.htm	**Relevant** or **Slightly Relevant**	Different states in the US have different laws. We do not know where the user is located, so a page about the law in a specific state would be helpful for few or some users.

Part 3: Page Quality Rating Guidelines

This section of the General Guidelines is about **Page Quality rating**. Since page quality is an important consideration when assigning a utility rating, you should also apply the concepts in this section to URL rating tasks, as well.

Here is a screenshot of the Page Quality rating task page:

All Page Questions

1) **Does the landing page load?**
○ no   ○ yes
2) **Is the page in the task language, an acceptable language or English?**
○ no   ○ yes
3) **Is the page porn?**
○ yes   ○ no

Landing Page Questions

1) **Identify the Purpose of the Page** Section 2.1
2) **Identify the Main Content, Supplementary Content and Ads** Section 2.2
○ no   ○ yes   ○ unsure
3) **Assign a Quality Rating to the Main Content** Section 2.3
○ no main content   ○ lowest   ○ low   ○ medium   ○ high   ○ highest
4) **Rate the Quantity of the Main Content** Section 2.4   *Note: Please do not lower page quality ratings because the page location doesn't match the task location.*
○ no main content   ○ unsatisfying   ○ so-so   ○ satisfying   ○ very satisfying
5) **Rate the Helpfulness of the Supplementary Content** Section 2.5
○ no supplementary content   ○ distracting or not helpful   ○ so-so   ○ helpful   ○ very helpful
6) **Rate the Layout of the Page/Use of Space** Section 2.6
○ misleading or deceptive   ○ poor   ○ so-so   ○ good   ○ excellent

Website and Homepage Questions

1) **Is the Purpose of the Page Consistent with the Website?** Section 3.2
a) Describe the purpose of the website   b) Is the page consistent with the purpose of the website? ○ no   ○ yes   ○ unsure
2) **Who is Responsible for the Content of the Website and the Content of the Page?** Section 3.3
a) Is it clear who is responsible for the content of the website?   ○ no   ○ yes   ○ unsure   b) Is it clear who is responsible for the content of the page?   ○ no   ○ yes   ○ unsure
3) **Is there an appropriate amount of contact information?** Section 3.4
○ no   ○ yes   ○ unsure
4) **What Kind of Reputation Does the Website Have?** Section 3.5
○ negative or malicious reputation   ○ mixed reputation   ○ positive or OK reputation   ○ little or no information found
5) **Is the Homepage of the Website Updated/Maintained?** Section 3.6
○ no   ○ yes   ○ unsure

Overall Page Quality Rating (Use Slider Below)
Rate the Overall Page Quality. If any of your responses above are red, the rating should probably be Low or Lowest. If you have multiple red responses, the rating should probably be Lowest.

N/A   lowest   low   medium   high   highest
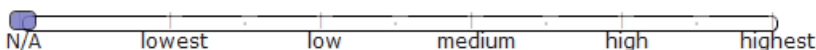
Comments/Feedback (Required)

Submit

1.0   Overview of Page Quality Evaluation

For each Page Quality rating task, you will click on the URL and evaluate the landing page.

There are two questions at the top of the Page Quality rating task page:

1) **Does the landing page load?** If you answer **no** to this question, you will merely submit the task. You may write a comment if you think it will be helpful.

2) **Is the page in the task language, an acceptable language or English?** If you answer **no** to this question, you will merely submit the task. Please note that you will rate all pages in the task language, including pages from websites outside the task location.

For all other landing pages, you will answer a series of questions. Some of these question focus specifically on the landing page while other questions are about the website that hosts the landing page. After carefully considering what you have learned from answering the page and website quality questions, you will use a sliding scale (sometimes called a "slider") at the bottom of the task to assign an overall Page Quality rating of **Highest**, **High**, **Medium**, **Low**, or **Lowest** (or a rating in between two of these ratings).

1.1   Introduction to Page Quality

You have probably noticed that webpages vary in quality. There are high quality pages: pages that are well written, trustworthy, organized, entertaining, enjoyable, beautiful, compelling, etc. You have probably also found pages that seem poorly written, unreliable, poorly organized, unhelpful, shallow, or even deceptive or malicious. We would like to capture these observations in Page Quality rating.

Unfortunately, if we ask you to rate the quality of a page without giving any guidance, the result is disagreement among raters. One rater will rate a page **High** quality and another will rate the same page **Low** quality. Why do we disagree?

- We may focus on different parts of the page or different aspects of the page. One rater might rate based on the content of the page and another based on the layout of the page.
- We may even have different ideas of what **High** quality means for a landing page. What makes an encyclopedia article **High** quality? What makes a product page **High** quality?

This guideline is important because it explains what to consider when rating the quality of a landing page. When you read this guideline, think carefully about the examples, and thoughtfully answer the questions about the landing page and website, Page Quality ratings are generally in agreement.

As with all of your rating and feedback, this data is used for evaluation purposes only.

1.2 Important Guideline Information

The goal of this guideline is to standardize our approach to Page Quality rating. An important part of this document is the examples, which are images of webpages. Please make sure you look at each example. Clicking on the webpage images will enlarge them so that you can read the text on the page.

At first, Page Quality rating may seem difficult. There are several aspects of the page and the website to look at and think about. As you gain Page Quality rating experience, you will be able to rate efficiently and with confidence. This type of rating takes practice. Rereading sections of the guidelines and thinking about the examples may help when you encounter difficult rating tasks. Please send feedback if you have a question about a particular rating task. Many examples and guidelines explanations have been added on the basis of rater questions.

This guideline is specific to webpages. Occasionally you may be asked to rate a landing page which is not a webpage. For example, you may be asked to rate a PDF file, a Microsoft Word document, a PNG or JPEG image file, etc. When the landing page of the URL is not a webpage, some of the questions in the rating task or considerations in these guidelines may not apply. In this case, please use your judgment.

Do not consider the country or location of the page or website for Page Quality rating. For example, an English (US) rater should use the same Page Quality standards when rating pages from other English language websites (UK websites, Canadian websites, etc.) as they use when rating pages from US websites. In other words, you should not lower Page Quality ratings because the page location (UK, Canada) does not match the task location (US).

When you are rating Page Quality rating tasks, try not to think about how helpful the landing page could be for a particular query. Page Quality rating is query-independent, meaning that the rating you assign does not depend on a query. It is almost *always* possible to think of a query for which the page could be helpful. Likewise, it is almost always possible to think of a query for which a page would not be helpful. This kind of reasoning is unhelpful for Page Quality rating.

Please do not struggle with each Page Quality rating. Just as you are advised to do in URL rating, please give your best rating and move on. If you are having trouble deciding between two ratings, please use the lower rating.

Finally, the questions in the rating template and the quality considerations in this guideline do not cover absolutely every aspect of page quality. If you find pages which you truly believe to be high or low quality, please rate them as such, even if the reason is based on something not covered in this document. Please explain your reasoning and include any additional criteria you considered in the comment section. As always, we ask you to use your judgment.

2.0   Landing Page Considerations

In the next sections, we discuss the aspects of Page Quality specific to the landing page of the URL and what you need to examine on the website that the landing page belongs to.

2.1   Identifying the Purpose of the Page

Every page on the Internet is created for a purpose (or for multiple purposes). Most pages are created to be helpful for users. Some pages are created merely to make money with little effort to help users, and some pages are even created to cause harm.

Pages with a Helpful Purpose

Common helpful page purposes are:

- To share objective information about a topic
- To share personal or social information
- To express an opinion or point of view
- To entertain
- To share pictures, videos, or other forms of media
- To sell products or services
- To allow users to post questions so that other users can answer
- To allow users to share files or to download software
- ...And many more!

The first step of Page Quality rating is to identify the purpose of the page. It is usually easy to tell what the purpose of a page is. Most of the time, you will understand the purpose of a page at a glance.

Here are a few examples where it is easy to understand the purpose of the page:

- Example 2.1.1: the purpose is to display news
- Example 2.1.2: the purpose is to sell or give information about the product
- Example 2.1.3: the purpose is to allow users to watch a video
- Example 2.1.4: the purpose is to calculate equivalent amounts in different currencies

Here are two examples where the purpose of the page is not as obvious:

1) Example 2.1.5 This page looks quite nice, but it starts off with "Christopher Columbus was born in 1951 in Sydney, Australia." This is obviously inaccurate! What is the purpose of this page?

In this case, exploring the website that hosts the page can help us understand its purpose. This website was built by educators to teach about interpreting information found on the Internet. After reading about the website here: Example 2.1.6, it should be clear that the purpose of this page is to serve as an educational tool. The information on the page is deliberately inaccurate so that it can be used as an example of misinformation on the Internet.

Note: Example 2.1.5 is actually a **High** quality page. Once you understand its purpose (to serve as an educational tool), it becomes clear that this page is carefully thought out, well executed, and achieves its purpose well.

2) Here is another example of a page that at first glance may seem pointless or strange: Example 2.1.7 This page is from a humorous site that encourages users to post photos with mouths drawn on them. The purpose of the page is humor or artistic expression. Even though the "About" page on this website is not very helpful, the website explains itself on its "FAQ" page Example 2.1.8

Why do we care about the purpose of the page? The purpose of the page will help you answer all of the landing page and website questions. *All questions should be answered in the context of the purpose of the page*. Ultimately, your Page Quality rating will depend on how well the page achieves that purpose, given everything you have learned from answering the landing page and website questions.

Important: In URL rating, the user intent behind the query determines the utility rating of the landing page. In Page Quality rating, the purpose of the page determines the Page Quality rating.

Keep in mind that for almost any helpful purpose, it is possible to find examples of high and low quality pages. The purpose of the page alone does not determine the Page Quality rating. For example, we will not consider informational pages to be higher quality than entertainment pages (or vice versa), even though they often have more serious content. There are high and low quality informational pages, and there are high and low quality entertainment pages.

Lack of Purpose, Harmful Purpose or Deceptive Pages

There are a few kinds of pages that should always be rated **Lowest** on the overall Page Quality scale.

- **Lack of purpose:** Rate the page **Lowest** if you cannot identify the purpose of the page despite your best effort to do so, which includes reading the "about" and other similar pages on the website. Many "lack of purpose" pages are "gibberish" or auto-generated. These pages serve no real purpose. Below are several examples of **Lowest** quality pages, which have no purpose:

- **Harmful purpose:** Rate the page **Lowest** if the purpose is clearly harmful or malicious. For example, pages designed to "phish" for the user's government-issued identification number (such as a Social Security number), bank account information, or credit card information should be rated **Lowest** quality because the purpose is to steal private information. Malicious download pages are another type of harmful page which should be rated **Lowest** quality.

- **Deceptive pages:** Rate the page **Lowest** if it is designed to look as though it has a helpful purpose but actually exists for some other reason. Deceptive pages are usually created to make money using ads or affiliate links rather than to help users. For example, some deceptive pages are designed to look as though they have helpful information, but in reality they are created to get users to click on ads.

Here are some examples of pages that should always be considered overall **Lowest** quality:

Type of Lowest Quality Page	Example	Explanation
Lack of Purpose	Example 2.1.10 Example 2.1.11 Example 2.1.12 Example 2.1.13 Example 2.1.14 Example 2.1.9	These pages do not have a purpose
Deceptive	Example 2.1.15	Not only does this page have many ads, but all the helpful-looking links lead to pages full of ads and very little text.
Deceptive	Example 2.1.16	The title of this page is "Washing Machine Reviews", but all the content is product information copied from another source and all links are affiliate links. The purpose of this page is clearly to profit from an affiliate program, rather than publish reviews.
Deceptive	Example 2.1.17	The title of this page is "Rachael Ray Diet Blog", but the page has nothing to do with Rachael Ray or her diet or her products. In fact, there is a brown-text-on-brown-background section at the bottom of the page (which we consider to be hidden text) that says "Disclaimer: Rachael Ray is not affiliated with nor does she sponsor or endorse this blog". **Important**: This page is deceptive in spite of the disclaimer!

Note: Lack of Purpose pages, Harmful pages, and Deceptive pages all violate the "Quality Guidelines" section of Google's "Webmaster Guidelines" (http://www.google.com/support/webmasters/bin/answer.py?answer=35769).

In general, any page or website which violates the "Quality Guidelines" section of Google's "Webmaster Guidelines" should be considered **Low** or **Lowest** quality.

2.2   Identifying the Main Content, Supplementary Content, and Advertisements

All of the content on a webpage can be classified in the following way: Main Content, Supplementary Content, and Advertisements.   The landing page questions require you to identify all parts of the page.

We will use the following page examples to illustrate how to identify Main Content (MC), Supplementary Content (SC), and Advertisements (Ads):

- News website homepage
- News article page
- Store product page
- Video page
- Currency calculator page
- Blog post page
- Search engine homepage
- Bank login page

Identifying the Main Content (MC)

Main Content is any part of the page that directly helps the page achieve its purpose. Main Content can be text, images, videos, page features, etc.

Type of Page and Purpose	Main Content Highlighted in Yellow
News website homepage: the purpose is to display news	Example 2.2.1m
News article page: the purpose is to display a news article	Example 2.2.2m
Store product page: the purpose is to sell or give information about the product	Example 2.2.4m
Video page: the purpose is to allow users to view a video	Example 2.2.5m
Currency calculator page: the purpose is to allow users to calculate equivalent amounts in different currencies	Example 2.2.6m
Blog post page: the purpose is to display a blog post	Example 2.2.9m
Search engine homepage: the purpose is to allow users to enter a query and search the internet	Example 2.2.7m
Bank login page: the purpose is to allow users to login to bank online	Example 2.2.8m

Identifying the Supplementary Content (SC)

Supplementary Content is content that does not directly help the page achieve its purpose. Sometimes the easiest way to identify Supplementary Content is to look for the parts of the page which *are not* Main Content or Advertisements.

High quality pages have helpful Supplementary Content, and that content contributes to a good user experience on the page. For example, one common type of Supplementary Content is navigation links which allow users to visit other parts of the website. On a video page, Supplementary Content might include related videos that users might be interested in watching. On a shopping page, Supplementary Content might include related products that users might be interested in buying.

Type of Page and Purpose	Supplementary Content Highlighted in Blue
News website homepage: the purpose is to display news	Example 2.2.1s
News article page: the purpose is to display a news article	Example 2.2.2s
Store product page: the purpose is to sell or give information about the product	Example 2.2.4s
Video page: the purpose is to allow users to view a video	Example 2.2.5s
Currency calculator page: the purpose is to allow users to calculate equivalent amounts in different currencies	Example 2.2.6s
Blog post page: the purpose is to display a blog post	Example 2.2.9s
Search engine homepage: the purpose is to allow users to enter a query and search the internet	Example 2.2.7s
Bank login page: the purpose is to allow users to login to bank online	Example 2.2.8s

Identifying the Advertisements (Ads)

Advertisements are content and links that are displayed for the purpose of monetizing the page. Advertisements are sometimes labeled as "ads", "sponsored links", "sponsored listings", "sponsored results", etc. Usually, you can mouse over the content or click on the links to determine whether they are Advertisements.

Type of Page and Purpose	Ads Highlighted in Red
News website homepage: the purpose is to display news	Example 2.2.1a
News article page: the purpose is to display a news article	Example 2.2.2a
Store product page: the purpose is to sell or give information about the product	No ads
Video page: the purpose is to allow users to view a video	Example 2.2.5a
Currency calculator page: the purpose is to allow users to calculate equivalent amounts in different currencies	Example 2.2.6a
Blog post page: the purpose is to display a blog post	Example 2.2.9a
Search engine homepage: the purpose is to allow users to enter a query and search the internet	No ads
Bank login page: the purpose is to allow users to login to bank online	No ads

Summary

Let's put it all together. Here are the examples again, with all parts of the page labeled:

Type of Page and Purpose	Main Content, Supplementary Content, and Ads Highlighted
News website homepage: the purpose is to display news	Example 2.2.1all
News article page: the purpose is to display a news article	Example 2.2.2all
Store product page: the purpose is to sell or give information about the product	Example 2.2.4all
Video page: the purpose is to allow users to view a video	Example 2.2.5all
Currency calculator page: the purpose is to allow users to calculate equivalent amounts in different currencies	Example 2.2.6all
Blog post page: the purpose is to display a blog post	Example 2.2.9all
Search engine homepage: the purpose is to allow users to enter a query and search the internet	Example 2.2.7all
Bank login page: the purpose is to allow users to login to bank online	Example 2.2.8all

Main Content and Supplementary Content are important parts of the page. It is easy to understand the need for Main Content: Main Content is the reason the page exists. Supplementary Content is also important. Almost every webpage needs navigation links, and most pages can be improved by features and content designed to help users get the most out of the page and website.

Many pages have Ads. Without advertising and monetization, some webpages could not exist.

Do not worry too much about identifying every little part of the page. Reasonable people can disagree on whether some parts of the page are Main Content, Supplementary Content, or Ads.

Tip: Carefully think about which parts of the page are the Main Content. Next, look for the Ads. Anything left over can be considered Supplementary Content.

2.3   Rating the Quality of the Main Content

Rating the quality of the Main Content is the most important step in Page Quality rating. You must think about the purpose of the page in order to evaluate the Main Content. High or highest quality Main Content allows the page to achieve its purpose in a highly satisfying way. Understanding the purpose of the page is extremely important for rating the quality of the Main Content. (Remember that all your rating must be done in the context of the purpose of the page.)

For each page, spend a few minutes examining the Main Content. Read the article, watch the video, examine the pictures, play with the calculator or online game, etc. Remember - Main Content also includes page features and functionality, so test the page out. For example, if the page is a product page on a store website, put at least one product in the cart to make sure the page and the shopping cart are functioning.

If there is a lot of content, give yourself about 3 minutes to browse through the Main Content on the page. Then assign a rating to the quality of the Main Content:

- no main content
- lowest
- low
- medium
- high
- highest

The purpose of the page will help you determine what high quality content means. For example, high or highest quality encyclopedia articles should be accurate, clearly written, and comprehensive. High or highest quality shopping content should help you find the products you want, research the products thoroughly, and make purchasing the products easy. High or highest quality humor content should be entertaining.

For all pages and all purposes, creating high quality Main Content takes a significant amount of at least one of the following: time, effort, expertise, and/or talent.

High or Highest Quality Main Content

Let's consider various types of pages and purposes and what is needed to create high or highest quality Main Content.

- Many medical pages exist to inform people on diseases and conditions. Highest quality medical content is written by people or organizations with medical accreditation (i.e., with professional expertise). The text and other content is written or produced in a professional style and is edited, reviewed, and updated on a regular basis (i.e., the content involves a high degree of time and effort to create and maintain).
- Many people create pages to share information about their hobbies. Here, time and effort as well as expertise and possibly talent are important. The highest quality content is produced by those with a lot of knowledge and experience who then spend time and effort creating content to share with others who have similar interests. For example, fish aquarium enthusiasts have created some of the highest quality content on the Web about how to set up and take care of a fish tank.
- Social networking pages for individuals exist to allow people to connect socially and express their personality. Most pages are created by the person they are about, and so the creator of the page is an "expert"-- we are all experts on our own lives. High quality content on social networking sites is often the end result of a lot of time and effort. The content is frequently updated with lots of posts, social connections, comments by friends, links to cool stuff, etc. Social networking content with few or no updates and little engagement or little effort should be considered low quality.
- Many people post videos on video sharing sites. The content of these videos varies, from home videos to documentary footage of events. Videos vary tremendously in quality as well. Time, effort, expertise, and often talent are needed to create a high or highest quality video.

Low or Lowest Quality Main Content

Main Content quality may be rated low or lowest for many different reasons. Often, the content is created without adequate time, effort, expertise, or talent.

Consider this. Most students have to write papers for high school or college. Many students take shortcuts to save time and effort by doing one or more of the following:

- Buying papers online or getting someone else to write for them.
- Making things up.
- Writing quickly with no drafts or editing.
- Filling the report with large distracting pictures.
- Copying the entire report from an encyclopedia, or paraphrasing content by changing words or sentence structure here and there.
- Filling up pages with completely obvious sentences that repeat the topic of the paper. ("Argentina is a country. People live in Argentina. Argentina has borders. Some people like Argentina.")
- Using a lot of words to communicate only basic ideas or facts ("Pandas eat bamboo. Pandas eat a lot of bamboo. It's the best food for a Panda bear.")

Unfortunately, the content of many webpages is similarly created. When it is clear that the Main Content is created with deceptive intent and without putting in enough effort, time, expertise, or talent, please assign a low or lowest Main Content quality rating. *Please note that copied or "scraped" content is created with deceptive intent and with very little time, effort, expertise, or talent. It is also a violation of the "Quality Guidelines" section of Google's "Webmaster Guidelines".*

Sometimes, time and effort were clearly involved when the page was created, but the content of the page does not allow the page to achieve its purpose. For example, expertise may be lacking on a topic for which expertise is really important. This content should be rated low or lowest.

Low or Lowest Quality Main Content Examples

Here are some examples of low or lowest quality Main Content. Many of the examples are articles. You may find it helpful to read the Main Content text out loud. Think about the quality of the Main Content as you read.

Type	Example	Explanation
Very poorly written content	Example 2.3.8	This content has many problems: poor spelling and grammar, complete lack of editing, inaccurate information, and no citing of sources of facts.
A lot of text but little actual information	Example 2.3.12	Many of the sentences in this example are either meaningless or state something obvious. For example: "Popping pimples could be or could be not the new trend of getting rid of them."
Commonly known information only	Example 2.3.11	Most people already understand how to put on a hat. This content is commonly known information: "Turn the hat upside down. Pick up the hat you want to put on with both hands. Grasp the hat from the bottom."
Commonly available content on non-authoritative source	Example 2.3.21	While the text in this example is not copied word-for-word from a source we would consider original, *many* sources list these exact same techniques described. Very little effort and no obvious expertise have gone into this particular example. And this is a topic where expertise/authoritative sources are important.
Non-authoritative/ untrustworthy author or website	Example 2.3.13	The level of expertise of the author of this content is not clearly communicated. Providing this background information is particularly important for medical, financial, or other subjects for which expertise is needed.
Auto-generated Main Content	Example 2.3.22	For auto-generated content, a basic template is designed from which hundreds or thousands of pages are created, sometimes by using an RSS feed or API. This particular page displays text (in the red box) explaining that the page is auto-generated. Auto-generated content should be rated lowest quality.
Machine-generated text that has no meaning at all	Example 2.3.15	This example is a page with random copied or gibberish text.
No Main Content at all, by deliberate design	Example 2.3.16	Some pages have errors causing there to be no Main Content. Other pages intentionally have no Main Content. If the page has no Main Content, use the "no main content" rating. This is an example of a page that *deliberately* has no Main Content (the part that looks like Main Content is actually Ads). Note that this page is deceptive as well, since the title of the page is "chickenrecipes.com", but the only content is Ads designed to look like a list of recipes.

Low or Lowest Quality Main Content Examples- *Copied, Scraped, or Paraphrased Content*

Important: Many low or lowest quality Main Content pages contain only copied or "scraped" content and were created with deceptive intent. This is a violation of the "Quality Guidelines" section of Google's "Webmaster Guidelines". Copied content should be considered low or lowest quality Main Content.

Copied content may be:

1. Copied exactly from an identifiable source. Sometimes a complete article is copied. Sometimes just parts of the article are copied. Text that has been copied exactly is usually the easiest type of copied content to identify.
2. Paraphrased slightly, making it difficult to find the exact matching original source. Sometimes just a few words are changed. Sometimes whole sentences are changed. Because of the changes, this type of copied content is harder to identify.

See Section 5.3 for how to check for copied content.

Some copied content pages get all of their content by making a copy of results from a search engine or news source. Because these are copies of "dynamic" pages that change frequently, you often will not be able to find an exact matching original source. Here are some examples:

Type of Copied Main Content	Description	Copied Content Example	Original Source of the Copied Content
The Main Content consists primarily of content copied from another source.	All of the Main Content is copied	Example 2.3.1	Example 2.3.2
	Most of the Main Content is copied	Example 2.3.9	Example 2.3.10
	Sometimes, we feel very sure that the content is copied, but we are unable to find the original source because it no longer exists.	Example 2.3.3	The original appears to no longer exist. See Section 5.3 for more information about this specific example
The Main Content consists of just RSS feeds, news feeds, or a copy of results from a search engine.	No effort has gone into creating this page. All content comes directly from "feeds" from other sites.	Example 2.3.6	RSS feeds, news feeds, search engine results

2.4   Rating the Quantity of Helpful Main Content

Overall **High** or **Highest** quality pages have enough helpful Main Content to accomplish their purpose and be very satisfying to users.

The quantity of helpful Main Content on overall **Low** or **Lowest** quality pages is often insufficient for their purpose. Some **Low** quality pages are unsatisfying and do not achieve their purpose well because they have a bare minimum of helpful Main Content. Some **Lowest** quality pages have so little helpful Main Content (or no Main Content) that they do not achieve their purpose at all. Sometimes there is a lot of Main Content (for example, many words or pictures), but it is not helpful for the purpose of the page. In all these cases, there is not enough *helpful* Main Content to achieve the purpose of the page.

Use the following ratings to indicate the quantity of helpful Main Content on the page:

Rating	Description
no main content	No Main Content at all on the page.
unsatisfying	Not enough helpful Main Content to achieve the purpose of the page and be satisfying to users.
so-so	Just enough helpful Main Content to achieve the purpose of the page and be somewhat satisfying to users.
satisfying	Enough helpful Main Content to achieve the purpose of the page and be satisfying to users.
very satisfying	Enough helpful Main Content to achieve and purpose of the page very well and be very satisfying to users.

The amount of helpful content needed depends on the purpose of the page. Here are some Main Content rating examples:

Purpose	Example	Rating	Explanation
Informational: the amount of content needed to be satisfying depends on the topic and the purpose of the page.	Example 2.4.1 – a page about the French Revolution	very satisfying	An informational page on the French Revolution should have a lot of information because this is a very large topic. This page has a satisfying amount of helpful content for its purpose
	Example 2.4.10 – another page about the French Revolution	unsatisfying	The title of the page indicates that it's about the French Revolution. However, it's a random collection of specific content that makes it unsatisfying for the general topic of the French Revolution.
	Example 2.4.8 – a page about baroque pearls	satisfying	This is a much narrower topic than the French Revolution. This page does not have a lot of content, but it is still satisfying for its purpose.
Shopping/ **Informational**: A very satisfying shopping page might include all or many of these: photos, specifications, manufacturer information, professional and user reviews, and convenient shopping features such as a menu to choose size and color.	Example 2.4.11 – a page for users interested in the TomTom XXL 550 GPS navigator	very satisfying	Note that the tabs on the page lead to even more information and have many customer reviews. Please consider the content under or behind tabs to be part of the content of the page. Such content should even be considered part of the Main Content of the page.
	Example 2.4.12 – another page for users interested in theTomTom XXL 550 GPS navigator	unsatisfying	This page for the same product has very little helpful Main Content. Note that the image of this webpage has been annotated with some additional information to help you recognize that this page is a "thin affiliate" and is therefore a violation of the "Quality Guidelines" section of Google's "Webmaster Guidelines". The amount of Main Content in this example is unsatisfying.
	Example 2.4.5 – a page for users interested in short wedding dresses	very satisfying	This page has a very satisfying amount of Main Content for users interested in short wedding dresses. An abundance of pictures plus options to view by price range, style, etc. are part of what makes this page so satisfying. This page achieves its purpose very well.
	Example 2.4.6	unsatisfying	This page does not have enough Main Content to be satisfying for its purpose.
Login: Login pages and some other types of pages need very little content to be satisfying or achieve their purpose.	Example 2.4.9	satisfying	This page has login functionality, as well as clear information about what the user is logging into.
Q&A: For a Q&A page, the Main Content includes the question and the answers.	Example 2.4.2	unsatisfying	In this example, the Main Content is boxed in red. Please read the Main Content in this example, including the completely unhelpful "answer" to the question in the red box.
	Example 2.4.3	unsatisfying	Some websites rely on users to create virtually all of their Main Content. If there is no user participation, the amount of content on the page is unsatisfying. This Q&A page has no answer to the question. (Note that it is possible that the page may someday have participation and more content.)
Lack of Purpose: It is possible to have an unsatisfying amount of helpful Main Content on a page even though there is lot of text.	Example 2.3.15	unsatisfying	This gibberish page has a lot of text, but none of it is helpful or satisfying for any purpose

Supplementary Content can be a large part of what makes a **High** or **Highest** quality page very satisfying for its purpose. Features designed to help shoppers find other products they might also like can sometimes be as helpful as the Main Content of a shopping page. Ways to find other cool stuff on entertainment websites can keep users happily browsing. Sometimes, the comments on a blog post are the most interesting part.

Take a look at the Supplementary Content and rate it on this scale:

- no supplementary content
- distracting/not helpful
- so-so
- helpful
- very helpful

So-so or Unhelpful Supplementary Content Examples

Low or **Lowest** quality pages frequently have Supplementary Content which is unhelpful or distracting for the purpose of the page. Here are some examples:

Supplementary Content Rating	Example	Explanation
distracting/unhelpful	Example 2.5.4	This page has a lot of text at the bottom and the sides which is not actually helpful to users. See the content in green towards the bottom. (Note: This page also has no Main Content and is overall **Lowes**t quality.)
distracting/unhelpful	Example 2.5.6	Some pages have way, way too many links, obscuring the page and distracting from the Main Content. This page is cropped to make it easier to view. The full webpage has even more links than are shown in this image.
no supplementary content	Example 2.5.10	The *only* links on this page are the buttons saying "Click Here". These are affiliate links to merchant sites, which allow the website owner to monetize the page. On the surface, this page seems to be a product review, but closer inspection shows that the page is designed to get users to click on the "click here" links: there is no other way off this page. This page is deliberately created to make money by trying to get users to click on the monetized links. This page has no Supplementary Content by design, in order to increase the chances of a click on the monetized links. (This is an example of a deceptive page and is considered overall **Lowest** quality.)

Helpful or Very Helpful Supplementary Content Examples

Supplementary Content Rating	Example	Explanation
very helpful	Example 2.5.1	The Main Content of this video page is a "Saturday Night Live" episode. Below the main video, there are many other videos that users may be interested in. This Supplementary Content is very helpful.
helpful	Example 2.5.2	This shopping page has helpful navigation features at the left to move around to different categories of shopping items. This Supplementary Content is helpful.

Here are some recipe pages which range from very helpful to so-so in Supplementary Content:

Supplementary Content Rating	Recipe Page Example	Explanation
very helpful	Example 2.5.3	This recipe page has helpful features, such as the ability to print out a list of ingredients. Behind the tabs, there are reviews, nutritional information, a video on how to knead bread dough, etc. The Supplementary Content is very helpful for a recipe page.
very helpful	Example 2.5.7	This page has reviews, a video, preparation time information, a "recipe box" feature, etc. This Supplementary Content is very helpful for a recipe page.
so-so	Example 2.5.8	While there is a lot of helpful *Main Content* on this page (including both text and images), there is little helpful recipe-specific *Supplementary Content*. There are no reviews, there is no shopping list, there is no recipe printing feature, etc. This Supplementary Content is just so-so for a recipe page.

You must consider the purpose of the page to decide whether Supplementary Content is helpful or distracting. Helpful Supplementary Content for recipe pages (e.g., features to print lists of ingredients or videos on cooking techniques mentioned in the recipe) may be very different than helpful Supplementary Content for encyclopedia pages (e.g., a list of other resources for the topic).

Reminder: This Page Quality guideline is specific to webpages. Occasionally you may be asked to rate a landing page which is not a webpage. For example, you may be asked to rate a PDF file, a Microsoft Word Document, a PNG or JPEG image file, etc. PDF files and other non-webpages may not have any Supplementary Content. For these types of pages, the absence of Supplementary Content is expected and is not a sign of low quality. Please rate these types of pages using your best judgment.

Here is an example of a high quality PDF file, which we display here as an image (just like all other examples in this guideline). There is no Supplementary Content, but none would be expected for this type of file: Example 2.5.11.

Of course, not all PDF files are high quality. Here is an example of a gibberish PDF which should be rated overall **Lowest** quality. Example 2.5.12.

2.6   Rating the Layout of the Page/Use of Space on the Page

Use of space refers to the position of and the amount of space on the page dedicated to Main Content, Supplementary Content, and Ads. You will rate on this scale:

- misleading or deceptive
- poor
- so-so
- good
- excellent

The page should be designed and organized to accomplish its purpose. While every page is different, pages that use space effectively should have these characteristics:

- The Main Content should be prominently displayed and "front and center". It should be immediately visible when a user opens the page.
- It should be very clear what the Main Content actually is. The page layout, organization and use of space, as well as the choice of font, font size, background, etc., of the page should make this clear.
- The Main Content and Supplementary Content together should take up most of the space on the page.
- Ads and Supplementary Content should be arranged so as not to distract from the Main Content.
- It should be clear what parts of the page are Ads, either by explicit labeling or simply by page layout.

Some pages are "prettier" or more "professional" looking than others, but you should not rate based on how "nice" the page looks. Pages can have lovely images, a pretty background, or a professional "look", but fail to achieve their purpose. Example 2.6.7 has a nice image at the top, but otherwise the use of space is poor: the Main Content is placed below many ads (and note that the Main Content is also low quality). Here is an example of a very low quality page that looks professional: Example 3.4.11 (see section 3.4 for more information on this example).

On the other hand, a page can be very functional, use space well, and achieve its purpose without being "pretty": Example 2.6.4

Like everything else, good layout and use of space depends on the purpose of the page. Let's look at a few examples. Please note that we are only considering the use of space in these examples. Several of these are low quality pages.

Use of Space Rating	Example	Explanation
misleading or deceptive	Example 2.6.6	Advertising should never disguise itself as the Main Content of the page. Pages with Ads that are designed to look like Main Content should be considered deceptive.
poor	Example 2.6.3	Advertising should never make it very difficult or impossible to find or use the Main content. In this example, some users might not even notice the Main Content because it is under a long list of Ads. A lot of scrolling is required to see the Main Content.
poor	Example 2.6.8	This example has mildly deceptive layout. This is a download page, so the download text and links are the Main Content. But the large green "Download now" button at the top is an Advertisement. Users could easily click on that button without realizing that it is an Ad, and not part of the Main Content.

Here are some examples for Q&A pages. Q&A pages with good use of space should devote much or most of the space to the questions and answers, and the layout/use of space should make the question and the answers clear. Good layout might also help users understand at a glance how many answers there are, which are the best answers, and clearly highlight the author and the date of the answer. The Supplementary Content should play a supporting role on the sides of the page, and Ads should be clearly distinguished from Supplementary Content and the answers.

Use of Space Rating	Example	Explanation
misleading or deceptive	Example 2.6.2	The use of space in this example is poor or even deceptive. The Ads and Supplementary Content take up most of the space. This page also uses the same "look" (color, font, and layout of the text), both for answers to the question and for the Ads/links, making it very difficult to distinguish which are answers to the question and which are Ads. (There are other aspects of this page which make it overall **Low** or **Lowest** quality: there are no features for users to rate the answers, and the page is overwhelmed by Ads and links.)
poor	Example 2.6.5	This is another example of a Q&A page that has poor use of space. This question does not have an answer, but it might take you a while to realize this. The page has a lot of Supplementary Content and Ads, much of which is shown where you would expect answers to be displayed. There is a section called "Relevant answers", but in fact these are pages from the same website with different questions, many of which are actually unrelated to this question. The "Relevant answers" section should be considered Supplementary Content. Compare this page to this example used earlier: Example 2.4.3. The use of space in Example 2.4.3 makes it immediately obvious that the question is unanswered. Of course, with no answer, we will consider the page overall **Low** quality, even though the use of space and layout is good for Example 2.4.3, but poor for Example 2.6.5.
good	Example 2.6.1	This is an example of good layout and good use of the space on a Q&A page, though the overall page quality is **Medium**. It's very clear what each part of the page is. The question is immediately visible, the authors are next to the questions and answers, all responses are dated, and the answers are sorted according to user feedback so that the best answer is at the top, immediately following the question.

Note: Invasive pop-ups or large flashing/animated/distracting Ads that cannot easily be closed are an example of poor use of space because they take attention away from the Main Content.

3.0   Answering Homepage and Website Questions

For Page Quality rating, we ask you to explore the website and answer some questions about the website.

Website checks are very important for distinguishing between overall **High** and **Highest** quality, as well as between overall **Low** and **Lowest** quality.

These website checks are especially important for stores or other websites which users entrust with credit card or other personal or financial information.

3.1   Finding the Homepage of the Website

To answer the website questions, you must visit the homepage associated with the URL in the task.

How do you find the homepage of the task URL? Click on the URL and examine the landing page. If the landing page is not the homepage of a website, you will usually see either a link labeled "home" or a logo to click on. If all else fails, examine and modify the URL by removing everything to the right of ".com" or ".org" in the URL. For example, to get to the Apple homepage from this URL: http://www.apple.com/support/iphone/, you would remove "support/iphone/" from the URL.

In the following examples, we have included the URL of the task page, as well as the URL of its associated homepage. We have also included an image that shows where to click on the task landing page to navigate to the homepage. You will see a red box around the link or the logo you would click to navigate to the homepage.

URL of the task page	Homepage of the website	Image that shows where to click
http://www.williams-sonoma.com/products/shun-classic-7-piece-knife-block-set/?pkey=cknife-sets|cutsetblk	http://www.williams-sonoma.com/	Example 3.1.1
http://www.mattcutts.com/blog/spam-reports-in-five-languages/	http://www.mattcutts.com/blog/ If you are curious as to why http://www.mattcutts.com/ is not considered the homepage, go to http://www.mattcutts.com/ and see what that page looks like. It's pretty clear that the blog homepage is actually http://www.mattcutts.com/blog/, not http://www.mattcutts.com/.	Example 3.1.3
http://answers.yahoo.com/question/index;_ylt=AnAYEU1fED6ncg1jRCFy30kk5XNG;_ylv=3?qid=20091214193523AAQqHQS	http://answers.yahoo.com In this case, we will consider http://answers.yahoo.com the homepage, rather than http://yahoo.com. Why? Because clicking on the logo takes the user to http://answers.yahoo.com. In addition, http://answers.yahoo.com has information about the Yahoo! Answers website. It is very difficult to find specific information about answers.yahoo.com on the yahoo.com homepage.	Example 3.1.5
http://hms.harvard.edu/hms/facts.asp	http://hms.harvard.edu/hms/home.asp In this case, we will consider the Harvard Medical School page at http://hms.harvard.edu/hms/home.asp to be the homepage, rather than http://www.harvard.edu/ (which is the homepage of Harvard University). Clicking the logo at the top of http://hms.harvard.edu/hms/facts.asp takes users to http://hms.harvard.edu/hms/home.asp, not to http://www.harvard.edu/.	Example 3.1.7

URL of the task page	Homepage of the website	Image that shows where to click
http://basketbawful.blogspot.com/2011/10/nba-lockout-lock-in.html	http://basketbawful.blogspot.com/ In this case, we will consider http://basketbawful.blogspot.com/ the homepage, rather than http://blogspot.com. Clicking the "Basketbawful" logo at the top of http://basketbawful.blogspot.com/2011/10/nba-lockout-lock-in.html takes users to http://basketbawful.blogspot.com/, not http://blogspot.com	Example 3.1.2
http://www.facebook.com/sarahpalin	http://www.facebook.com/ In this case, we will consider http://www.facebook.com/ the homepage. Clicking the "facebook" logo at the top of http://www.facebook.com/sarahpalin takes users to http://www.facebook.com/	Example 3.1.4
http://twitter.com/#!/BarackObama	http://twitter.com/ In this case, we will consider http://twitter.com/ the homepage. Clicking the "twitter" logo at the top of http://twitter.com/#!/BarackObama takes users to http://twitter.com/	Example 3.1.6

Whenever possible, look for a "home" link or try clicking on the logo. Sometimes, the navigation links and structure also have a clear hierarchy, with the homepage featured prominently. Webmasters often make it easy to get to the homepage of the website, and homepages usually have links to a lot of helpful information you need to evaluate page quality.

If there is no clear home link (or logo or navigation structure on the page you are evaluating), try looking at some other pages on the site. If finding the homepage is hard, usually either the website design is amateurish or the website is low quality. Use the questions in the next section to decide which.

Occasionally, your rating task will include a URL for which there are two or more justifiable "homepage" candidates. For example, you may find a landing page that has a "home" link, as well as a website logo with a link. Usually these links go to the same page, but sometimes they have different landing pages.

Or you may find a complicated relationship between subdomains and the top level domain associated with a URL. For example, you may not be sure whether the homepage of the URL http://finance.yahoo.com/news/category-stocks/ is http://finance.yahoo.com/ or http://www.yahoo.com/). Many websites have a complicated directory structure, and you may not be sure which page is reasonably the "homepage" associated with the landing page of a given URL.

As always, please use your judgment. You may consider information from any reasonable homepage candidate. In general, please prefer the homepage candidate which has the most information relevant to the landing page.

Please note that we include finding the homepage as part of the Page Quality guidelines because we are trying to find information and answer questions about a specific landing page and the website it is associated with. Frequently, some of the information we need to know about the landing page (such as authorship, contact information, etc.) is not on the landing page itself because it is the same for every page on the site or subdomain, or it is contained within a certain branchy of a directory on the site. Finding the homepage (or relevant subdomain or appropriate page in a directory) is a good technique for locating this kind of information. Please use any reasonable "homepage" candidate that allows you to find the information you need to rate effectively.

Note: Amateurish (i.e., non-professional) website design may not be a page quality issue, at least for certain types of pages (for example, websites of family photos). Amateurish website design is less acceptable for topics or purposes needing a high degree of professionalism or trust (for example, legal, financial, and medical websites or shopping websites that ask for your credit card information).

3.2   Is the Purpose of the Page Consistent with the Website?

You have already looked at the page and figured out its purpose. Now do the same thing for the website.

Sometimes websites explain their purpose right on the homepage. Some websites put this information on a subpage. Examine the homepage for this information. Also look for "about" or "about us" links or "contact us" links or even "FAQ" links.

You want to make sure the purpose of the page you are evaluating and the purpose of the website are consistent. Here are the website purpose questions in the rating task:

   a)   Describe the purpose of the website.
   b)   Is the page consistent with the purpose of the website?

Here are some examples:

Consistency Check	Example	Information about the website	Purpose of the page	Purpose of the website	Explanation
consistent	Example 3.2.1	Example 3.2.2	Share a recipe	Provide recipes and other cooking resources	The stated purpose of the website and the page are consistent.
inconsistent	Example 3.2.3	Example 3.2.4	Share a large collection of clearly non-recipe videos	Provide recipes	The stated purpose of the website (to provide recipes) is inconsistent with the presumed purpose of the page (to display videos, all of which happen to be unrelated to recipes).

Note: Example 3.2.3 above has a collection of random and unrelated content and videos. Some people might justifiably say the page has no clear purpose. It also appears that the content may be copied or scraped from some other page on the Web. These are some of the many indications that this page is overall **Lowest** quality. In many cases, landing page and website questions give a consistent picture. We ask you to do page and website checks because sometimes a **Lowest** quality page only clearly fails on one question. Example 3.2.3 fails on many.

If the purpose of the page is very clearly inconsistent with the purpose of the website, please examine closely. Usually, **Lowest** will be the most appropriate overall Page Quality rating.

If the purpose of the *website* is deceptive or harmful, or if the website has no purpose (for example, the website is made up of gibberish pages), all pages on the website should be considered overall **Lowest** quality.

3.3   Who is Responsible for the Content of the Website and the Content of the Page?

Every page belongs to a website, and it should be clear:

   1.   Who (what company, business, foundation, individual, etc.) is responsible for the content on the website?
   2.   Who is responsible for the content on the page you are evaluating?

The website check questions you will answer in the rating task are:

   a)   Is it clear who is responsible for the content on the website?
   b)   Is it clear who is responsible for the content on the page?

Websites are usually very clear about who is responsible for the content on the website and the content on the page. There are many reasons for this. Commercial websites have copyright material they want to protect. Businesses want users to know who they are. Artists, authors, musicians, and other original content creators usually want to be known and appreciated. Foundations often want support and even volunteers. High quality stores want users to feel comfortable buying online.

The homepage of the website should indicate who is responsible for the content of the website. Most websites have "contact us" or "about us" or "about" pages that give you information about who owns the site. Many companies have an entire website or blog devoted to who they are and what they are doing, what jobs are available, etc. Please look carefully and explore the website. Blogs or even social networking pages may be the way companies communicate who they are.

An example of a website where it is clear who is responsible for the content of the website: Example 3.3.1 This website has a nice page about who they are. Note that this page is titled "WHO IS GuS?", not "about" or "about us". You will need to look at the information provided on the website, not necessarily the title of the page on which the information appears.

Website Owner and Content Creator are Different

In some cases, the website is owned by one company, but the content on the page is provided by a different company or individual.

For example, social networking sites allow individuals to create pages. The social networking website is responsible for many aspects of the site, but individuals are responsible for what is on their pages (subject to the terms and restrictions of the social networking site).

Here are some examples:

Who is responsible?	Example	Explanation
clear	Example 3.3.2	This is Oprah's Facebook page. She or the Oprah Winfrey show is responsible for the content on this page. Facebook (the company) is responsible for the website.
unclear	Example 3.3.3	There is no "about us" information on who is responsible for the page content or for the website.
unclear	Example 3.3.4	The "About Us" page does not give the name of a company or physical address. No other page on this site has information either. Clicking the "contact us" link will open the default email program on your computer with an email addressed to the website. Clicking the link will not provide contact information for the website.

3.4 Does the Website Have an Appropriate Amount of Contact Information?

Most websites are interested in communicating with their users. Usually, this means that websites offer paths of communication that at the very least include phone numbers and email addresses. High or highest quality websites will usually offer many ways by which users can get in touch, such as email addresses, phone numbers, and physical addresses. Sometimes, this contact information is even organized by department and provides the names of individuals to contact.

The types and amount of contact information needed depend on the type of page and the type of website. In other words, the amount differs based on the purpose of the page and the website. Contact information is extremely important for online stores. Be extra critical of shopping websites. Most stores have contact information or contact processes prominently featured. Sometimes this information is listed as "customer service". Users often need to contact stores for questions and returns. Stores usually work very hard to win users' trust by offering many ways to contact them. Some retailers have special sections of their website devoted to customer service.

Contact information is also very important for any website requiring users to input personal information or financial information. In general, contact information is very important for websites that require a high level of user trust, such as banks, medical websites, etc.

Some kinds of websites need less detailed and a smaller amount of contact information for their purpose. For example, humor websites may not need the kind of detailed contact information we would expect an online banking website to have. Websites that require a lower level of trust should still have the name of the company, a phone number and/or an email address, at a minimum.

Find the contact information associated with the website. Usually, homepages have a "contact us" link. There should be clear contact information somewhere on the website. Sometimes, the contact information is located on the "about us" page and sometimes it is displayed directly on the homepage. Please explore the website if you cannot find a "contact us" page. Sometimes you will find the contact information on a "corporate site" link. Be a detective.

The website check question you will answer in the rating task is: Does the website have sufficient contact information for its purpose?

Here are some examples:

Sufficient contact information?	Example	Explanation
sufficient	Example 3.4.1	This page provides the street address and phone number of the company, as well as email addresses and phone numbers (and even names) of various people/departments in the company.
sufficient	Example 3.4.7	In this example, there is a "Company Site" link in the upper right hand corner (we added the blue box to help you find it). Clicking on that link gives the corporate site homepage: 3.4.8. From the corporate homepage, there is a "Global Contacts" page under the "Company" tab: 3.4.9. Clicking on this link gives the "Global Contacts" page: 3.4.10. This rating task required raters to be detectives. (Note: this is an example of a highly reputable financial site—we definitely do not want this type of page to get low quality ratings!)
sufficient	Example 3.4.2 Example 3.4.3 Example 3.4.4	These three examples are customer service pages on popular and well-regarded stores/shopping sites that help customers contact the store. Notice that these pages use a variety of approaches, from giving users a list of phone numbers to providing interactive features that navigate users through the customer service process.
insufficient	Example 3.4.5	This is a blank email form with no indication to whom it would go. There is absolutely no other contact information anywhere on the website
insufficient	Example 3.4.6	There is no contact information at all anywhere on this website.
insufficient	Example 3.4.11	There is no contact information at all anywhere on this website. (Note: there is also absolutely no information about who is responsible for the content of this website.)

3.5 What Kind of Reputation Does the Website Have?

A website's reputation is based on the experience of real users, as well as the opinion of people who are experts in the topic of the website.

Stores frequently have user ratings, which can help you understand a store's reputation based on the reports of people who actually shop there. Many other kinds of websites have reputations as well. You might find that a newspaper website has won journalistic awards. You might find that a medical information site is endorsed by physician groups.

The reputation of a website is especially important for monetary transactions or when private data is involved.

The reputation of a website is also very important when the information on the website demands a high level of authoritativeness or expertise, such as medical information websites. When a high level of authoritativeness or expertise is needed, the reputation of a website should be judged by what expert opinions have to say. For example, high quality medical websites are endorsed by prominent physician groups.

Reputation research in Page Quality rating is very important. A positive reputation from a consensus of experts is often what distinguishes an overall **Highest** quality page from a **High** quality page. A negative reputation should not be ignored and is a reason to give an overall Page Quality rating of **Low** or **Lowest**.

The website check question you will answer in the rating task is:

What kind of reputation does the website have?

- negative or malicious reputation
- mixed reputation
- positive or OK reputation
- little or no information found

Reviews and ratings of websites can be very helpful for determining reputation. Any store or website can get a few negative reviews, and that is normal. Large stores and companies have thousands of reviews and most receive a few negative ones. However, it is **not** normal for businesses or websites to have all or predominantly negative reviews.

Use the "negative or malicious reputation" rating if there are many believable, detailed negative reviews (and very few positive reviews), or if there are informative/believable complaints from multiple sources detailing malicious behavior.

Note: Frequently, you will find little or no information about the reputation of a small website. This is not indicative of positive or negative reputation.

Sometimes, you will find information about a website, but it is not related to its reputation. For example, pages with information about Internet traffic to the website are not about the reputation of the website. Please select "little or no information found" if the information you find is unrelated to the reputation of the website.

How to Search for Reputation Information - The usual way to find out about reputation is by searching for comments, reviews, and articles and references about the website. Usually, one quick search is all you need. Here is how to research the reputation of the website:

1. Identify the "homepage" of the website.
2. Try one or more of the following searches on Google:
   [homepage]
   [homepage.com]
   [homepage reviews]
   [homepage complaints]
   [homepage -site:homepage.com]
   ["homepage.com" -site:homepage.com]
   ["homepage.com" -site:homepage.com reviews]
   [link:homepage.com]
3. Browse through the results to see what others have to say about the website.
4. For businesses, there are many sources of reputation information and reviews. Here are some examples: RepExample1, RepExample2, RepExample3, RepExample4, RepExample5
5. For many other types of websites, you may find newspaper articles, encyclopedia articles, and other informational pages which can be helpful.
6. Look for other websites that reference the website you are researching. On Google search, you can use the "link:" operator.

Here is an example that uses some of these techniques.
URL of the landing page: http://www.decormyeyes.com/pd.asp?prod_id=1940

1. Identify the homepage of the website, which is "decormyeyes.com".
2. Issue the query [decormyeyes reviews] to see what others have to say about the website. Here is what this search looks like: Example 3.5.3. You will notice many extremely negative reviews and complaints about this malicious site. The New York Times has an article extensively detailing the malicious behavior of this website.
3. Also try this query: [decormyeyes -site:decormyeyes.com]
4. Try the search [décor my eyes site:bbb.org]. You will see that the Better Business Bureau (BBB) gives this business a very low rating: Example 3.5.5

Please note the following

- The Better Business Bureau does not have ratings for all businesses. You will sometimes find high ratings on BBB because there is very little data on the business, not because the business has a positive reputation. However, low ratings on BBB are usually the result of multiple unresolved complaints. Please consider very low ratings on the BBB site to be evidence for negative reputation.

- Including "site:homepage.com" in your query restricts the search results to include only pages from the site http://www.homepage.com/. Another helpful trick is to try adding "-site:homepage.com" to the query, which restricts your search results to pages *about* homepage.com, but not *on* the website homepage.com.

Sometimes it is helpful just to search using the name of the homepage (i.e., "homepage"), and sometimes it is helpful to search for the URL (i.e., "homepage.com"). If you are having trouble finding any information, try both.

Here are some other examples:

Example	Reputation Research	Reputation	Explanation
atgstores.com	Google Search Results page for [atgstores.com complaints] Google Search Results page for [atgstores.com reviews]	Negative	You will notice many detailed extremely negative reviews about this site. Please see these pages: Example3.5.6.1, Example3.5.6.2, Example3.5.6.3
hhv.org/	Google Search Results page for [help hospitalized veterans scam]	Negative	You will notice many detailed negative articles about this organization on news sites and charity watchdog sites: Example3.5.10.1, Example3.5.10.2 , Example3.5.10.3, Example3.5.10.4,
denisonyachtsales.com	Google Search Results page for [denisonyachtsales.com reviews – site:denisonyachts.com]	Positive to Mixed	You will find two negative reviews. This one has lots of detail and is believable: Example3.5.7.1 This review at the top of this page is questionable: Example 3.5.7.2. It offers no details and feels spammy. Notice that there is a credible response by the owner further down the page. Although we could find pages with positive reviews about this business, we can infer a positive reputation by looking at the following: Example3.5.7.3, Example3.5.7.4, Example3.5.7.5 In this case, we should consider the very credible sounding negative review, but it should not outweigh the overall generally positive reputation of the business. We would consider denisonyachtsales.com to have an OK or positive reputation.
csmonitor.com (Christian Science Monitor)	Google Search Results page for [csmonitor.com - site:csmonitor.com]	Positive	Notice the highlighted section in this article about the Christian Science Monitor newspaper, which tells us that the newspaper has won seven Pulitzer prizes. Example 3.5.8.1 From this information, we can infer that the csmonitor.com website has a positive reputation.
llbean.com	Search results for [llbean.com] Search results for [llbean.com reviews –site:llbean.com]	Positive	From the numerous positive user reviews on these sites, we can infer that llbean.com has a positive reputation: Example3.5.9.1, Example3.5.9.2, Example 3.5.9.3

Final note: Reputation is different than popularity. Popular websites are used by many people and are frequently (but not always) high quality. Reputation is based on both experiences of users as well as the opinion of experts in the topic of the website. So medical websites with positive reputations are those recommended by prominent medical groups. Stores with good reputations are ones that serve their customers well (not just the ones that sell the most products). Reputation research is necessary. Do not just assume websites you personally use have a good reputation. Please do research! You might be surprised at what you find.

High or highest quality websites stay current. Most continue to add high quality content on a frequent basis. For example, most news websites add content continuously throughout the day and archive content that is relevant.

Users can trust high quality websites because they maintain their content. For example, medical advice changes all the time. A high or highest quality medical information site will remove information that has become outdated and will quickly correct errors.

The frequency of maintenance should depend on the purpose of the website. We would expect the homepage of a news website to be updated many times a day. Other websites might have a much slower cycle. Most homepages should be updated at least every few years.

The website check question you will answer in the rating task is: Is the homepage of the website maintained or cared for?

Homepage Checks - You cannot check every page on a website to make sure the content is maintained, but please do inspect the homepage. Make sure that at least the homepage looks reasonably maintained and updated. Check the following:

1. Check copyright dates, if available. These should be recent- or at least not more than 2 years old. Note: some amateur websites may not be vigilant in updating the site's copyright date. If you see an old date but strongly feel the page is OK, you may consider the website maintained.   For example, it's probably OK and even expected if a website of family photos does not have a copyright date.   It's also OK if a small volunteer organization website has not updated its copyright date if the group is obviously still updating other parts of the homepage and website. But we would expect business and professional websites to have recent copyright dates.
2. Look for a "last updated" date or other indication that the website is maintained. Most homepages should have evidence of recent updates.
3. Check to see if links on the homepage work and that basic functionality of the homepage works. (One broken link or so is OK- every website has those now and then.) Do a few random clicks to check. The images and formatting should also look OK. The page should feel cared for.

Medical Information Examples - Here are two examples (both from medical information sites) that do not appear to be updated/maintained. Note that medical websites require up to date information and a high degree of user trust.

1. Example 3.6.1: The information at the bottom of the page indicates it was last updated in 2005. It does not look like anyone is taking care of this website/maintaining the accuracy of this information.
2. Example 3.6.2: There is a note at the top that the site is for sale. There are many other indications that this page is overall **Low** quality, but the "for sale" notice should make you suspicious about updates/maintenance of the information.

4.0   Assigning an Overall Page Quality Rating

Each Page Quality task begins with a click on the task URL. Page Quality rating is a series of steps, including examining both the landing page and the website it belongs to.
Which questions are most important? How do you combine your responses into an overall Page Quality rating?

As with the rest of this guideline, the answer depends on the purpose of the page and type of website. You will have to use your judgment to combine all that you learned about the page and the website into one Overall Page Quality rating. In this section, we'll give you some guidance and some examples.

Remember: In order to give an overall rating above **Medium**, every aspect of page and website quality you look at must be medium or high. If any aspect of quality seems low, whether the aspect pertains to the page or the website, choose a lower overall rating.

4.1   Highest Quality Pages

Highest quality pages are highly satisfying to users. **Highest** quality pages have a large amount of very high quality content, very helpful supplementary content, and use very good webpage layout. The author(s) of the content on **Highest** quality pages should have a very high level of expertise in the subject.

In addition, **Highest** quality pages are often found on websites that have a very good reputation from experts in the topic (even if average users or raters are unaware of the site or its reputation).   Reputation checks are an important part of identifying H**ighest** quality pages.

- **Highest** quality pages have an obvious purpose and they achieve that purpose very well.
- The Main Content of **Highest** quality pages is created by people with a high level of expertise in the topic.
- **Highest** quality pages have a very satisfying amount of Main Content.
- The page layout on **Highest** quality pages makes the Main Content immediately visible ("front and center").
- The space on **Highest** quality pages is used well.
- The Supplementary Content on **Highest** quality pages is helpful and contributes to a very satisfying user experience.
- **Highest** quality pages usually have near professional quality content, even though ordinary individuals may create the content.
- **Highest** quality pages frequently appear on high quality websites with very positive reputations for their purpose or topic, such as:

  - Award winning newspaper sites for news
  - Authoritative sites for medical information
  - Well-known "go-to" recipe sites for recipes
  - Highly regarded and trusted shopping sites

4.2 High Quality Pages

High quality pages are satisfying to users.   **High** quality pages have a large amount of high quality content, good use of space, and are trustworthy and authoritative.

- **High** quality pages have an obvious purpose and they achieve that purpose well.
- **High** quality content is created by people with appropriate expertise.
- **High** quality pages have high or highest quality Main Content.
- **High** quality pages have a satisfying amount of Main Content.
- The page layout of **High** quality pages makes the Main Content clearly visible.
- The space on **High** quality pages is used reasonably well.
- The Supplementary Content on **High** quality pages is helpful.
- **High** quality pages appear on all sorts of websites, large and small, but the website of the page you are evaluating should "pass" all website checks, including the reputation check. **High** quality pages may be found on websites with a positive reputation or no reputation. They may even be found on mixed reputation websites, if they have enough positive evidence to support an overall **High** quality rating. **High** quality pages will not be found on websites with a negative reputation.

4.3 Medium Quality Pages

Many pages are overall **Medium** quality. There is nothing "wrong" with **Medium** pages. However, it is easy to identify dimensions along which the page could be improved: they could have higher quality content, more content, more helpful and specific supplementary content, better layout, a better reputation, etc.

- **Medium** quality pages have a purpose and they achieve that purpose.
- **Medium** quality pages may have medium quality content, or even high quality content
- The amount of Main Content on **Medium** quality pages is OK, though it may not be extensive.
- The page layout on **Medium** quality pages makes the Main Content visible.
- The space on **Medium** quality pages is used reasonably well.
- The Supplementary Content on **Medium** quality pages is helpful or OK.
- **Medium** quality pages appear on all sorts of websites (and, in fact, many pages on the web are **Medium** quality). The website of the page you are evaluating should still "pass" all website checks.
- **Medium** pages may appear on websites with positive, mixed reputation, or no reputation. Having many negative reviews is a possible reason to give a **Medium** rating, even to pages on a popular website. You will need to use your judgment, taking into consideration the mix of positive and negative reviews, the reasons for the negative reviews, and the overall reputation and popularity of the website.

4.4 Low Quality Pages

Low quality pages come in many flavors. Usually, **Low** quality pages are lacking in one or a few of the aspects we consider for the page or the website.

Low quality pages may only be acceptable to users if there are no other higher quality pages. Many **Low** quality pages feel unsatisfying, especially to more discriminating users. Some **Low** quality pages may feel "ok", but the website itself lacks enough information to feel credible or trustworthy.

Please note: if a page has low quality main content, please rate the page **Low** or **Lowest** quality. If the website that the page is on "fails" one of the website checks, please rate the page **Low** or **Lowest** quality. In other words, if any of your checks find an area of concern, do not hesitate to use **Low** or **Lowest** ratings.

- **Low** quality pages usually have a purpose, though the purpose may be somewhat unclear or the page may not achieve that purpose well.
- **Low** quality pages may have low quality Main Content. The Main Content may be higher quality but copied from another source (perhaps with minimal alteration).
- The amount of Main Content on **Low** quality pages may be lacking.
- The page layout on **Low** quality pages may be poor.
- The space on **Low** quality pages may not be used well.
- The Supplementary Content on **Low** quality pages may be unhelpful or distracting or lacking.
- **Low** quality pages may have an obvious problem with functionality or may have errors in displaying content. For example, the Main Content of the page may fail to load. (Please note that if many pages on the website have problems, you should consider the website unmaintained and use the **Lowest** rating).
- **Low** quality pages exist on all sorts of websites. In fact, many high quality websites have a few low quality pages. For example, most websites have a few pages with non-loading content or very little Main Content.
- Negative reputation alone can be the reason for a **Low** rating, but to assign this rating there must be evidence of an overwhelmingly negative reputation found on multiple sources.
- There are many "flavors" of **Low** quality pages. If a page does not live up to the standards established in this guideline for any reason, please use the **Low** (or **Lowest**) rating.

4.5 Lowest Quality Pages

Lowest quality pages may be severely lacking in one or more of the aspects we consider for the page or the website. The **lowest** overall Page Quality rating should be used for pages on websites that obviously fail one or more website checks, even if all page level ratings are OK.

Lowest quality pages contribute little to the internet.   Many or most users would find **lowest** pages unsatisfying. In many cases, users would be better off without **lowest** quality pages.

- **Lowest** quality pages may not have a purpose, or the page may have an unclear/deceptive or malicious purpose.
- **Lowest** quality pages may have a purpose but fail to achieve that purpose
- **Lowest** quality pages may have low or lowest quality Main Content. The Main Content may be medium or higher quality but completely copied from another source.
- The amount of Main Content on **Lowest** quality pages may be lacking. There may not be any Main Content.
- The page layout on **lowest** quality pages may be poor.
- The Supplementary Content on **lowest** quality pages may be unhelpful or distracting or nonexistent.
- The page may appear on a website which significantly "fails" the website level checks; for example, a website with absolutely no information about who created the website or how to contact the owner of the site.
- Extremely negative reputation can be a justification for the **Lowest** rating in cases of deceptive or malicious behavior.

The page may appear on a website that obviously or significantly violates the "Quality Guidelines" section of Google's "Webmaster Guidelines".

5.0   Additional Page Quality Rating Guidance

This section is included to give additional rating guidance, answer rater questions, and address common rating mistakes.

5.1   Assigning a Page Quality Rating to Pages with no Main Content/Error Messages

Pages may lack Main Content for various reasons.   Sometimes, the page is "broken" and the content does not load properly or does not load at all. Sometimes, the content is no longer available and the page displays an error message with this information. Sometimes the page is deliberately designed without Main Content.

Every website probably has some "broken" non-functioning pages. This is normal, and those *individual* non-functioning or broken pages on an otherwise maintained site should be considered overall **Low** quality. This is true even if other pages on the website are overall **High** or **Highest** quality.

Sometimes website checks for a broken or error message page reveal that the individual page is not an isolated example, but rather a symptom of an unmaintained site (or possibly a deceptive or malicious site). If that is the case, the page should be rated **Lowest** quality.

If the page has no Main Content by design, the rating should be Lowest.

Not all pages with error messages are **Low** or **Lowest** quality pages, however. If the purpose of the page is to communicate that content has been moved or is no longer available and the page does a good job of communicating this message, the overall **Page Quality** rating may be higher; it may be **Medium** or even **High**. The Page Quality rating will depend on the website level checks and the content of the page.

Reminder: Page Quality ratings are query-independent.   The full range of the Page Quality rating scale (from **Lowest** to **Highest**) can be used when rating pages, even "custom 404" pages. The rating will depend on how well the page achieves its purpose and how well it passes page and website checks.

Here are some examples of "broken" or error message pages:

Example	Rating	Discussion
Example 3.6.3	Lowest	This is an example of a "broken" looking page which appears to be missing Main Content. You might think that this page is just "missing" the main content due to a problem with this particular page. In fact, this website has hundreds of pages that look the same way—no Main Content, just Ads. This website is either unmaintained or deceptive. Either way, this page should be rated **Lowest** quality.
Example 3.6.4	Low	This is another example of a "broken" looking page which appears to be missing Main Content. Even though the website is a legitimate news website and passes all other website checks, we will consider this an overall **Low** quality page.
Example 3.6.6	Medium	This is an example of a "custom 404 page". "Custom 404 message" pages are designed to alert users that the URL they are trying to visit no longer exists. Some websites do a nice job of not only alerting users about a problem, but also giving them help. This particular page displays the bare minimum of content needed to explain the problem to users, and the only help offered is a link to the homepage. The website is high quality, passes all website checks, and has a good reputation. But there obviously is not a lot of time and effort put into the particular page. The page accomplishes its purpose and there is nothing "wrong" with the page, but it is not a high quality page. The overall Page Quality rating here should probably be **Medium.**
Example 3.6.5	High	This is an example of a very high quality "custom 404" page. The Main Content of this page is the cartoon, the caption, and the search functionality (which is specific to the content of the website). Clearly, time, effort, and talent were involved in the creation of the Main Content. This page achieves its purpose well, and the website hosting the page passes all website checks and has a good reputation. An overall Page Quality rating here should probably be on the high end of the range.

5.2   Balancing Page Level and Website Level Questions to Assign an Overall Page Quality Rating

For some Page Quality rating tasks, the **page level** questions (i.e., the questions about the landing page) will be the most important questions to consider. For other tasks, the **website level** questions (i.e., the questions about the website) will play an equally large or even a larger role. As always, use your best judgment.

Important: If either page level or website level questions reveal a significant quality concern, the page should get an overall **Low** or **Lowest** rating. Please give a high rating only if both the page level and the website level questions can justify a high rating.

Website level questions play the biggest role (and you should use an overall **Low** or **Lowest** Page Quality rating) when:
*   The website is harmful/deceptive/malicious. In this case, all pages on the website should be rated overall **Lowest** quality.
*   The website has an extremely negative or malicious reputation. If so, all pages on the website should be rated overall **Lowest** quality.
*   The website obviously or significantly violates the "Quality Guidelines" section of Google's "Webmaster Guidelines". If so, all pages on the website should be rated overall **Low** or **Lowest** quality.
*   The website clearly *completely* fails one or more of the website level checks in this document. If so, all pages on the website should be rated **Low** or **Lowest** quality. The overall rating depends on the purpose of the site and the check which revealed the quality concern. For example, all pages should be considered **Lowest** if the website is a store and there is absolutely no contact information, customer service information, or ownership information. For a store, this is an extremely important quality check!

Website level questions play a *medium to large role* in determining the overall Page Quality rating when:
*   The content on the site is very uniform, i.e., when all content is produced by the same organization and there is little variation in the content or quality of pages. An example is a medical website produced by a reputable physician group where each page on various diseases has a similar amount of high quality content.
*   The content of the website is produced by different authors or organizations, but the website has very active editorial standards and guarantees a level of accuracy/quality. An example is a science journal with very high standards for publication.
*   The website has an extremely positive reputation from experts in the topic of the website, i.e., the website is acknowledged to be one of the most authoritative sources on the topic.

Page level questions are most important when:
*   The website clearly "passes" all website level checks. If the website checks are "cleared", the overall Page Quality rating will usually depend almost entirely on the content of the page.
*   The website has many different authors/contributors, or there is a large variation in the quality of the pages. Even if a website is popular and has a good reputation, you should rely on page level checks when the content is produced by different people. Page level questions are *particularly* important when the website has little or no active editorial standards, i.e., anyone can publish any content on the website.

So how do you determine whether the content is copied or copied with minimal alteration? How do you identify the original source of the content? This can be difficult, but please follow these steps.

1. Copy a sentence or a series of several words in the text. It may be necessary to try a few sentences or phrases just to be sure. When deciding what sentence or phrase to copy, try to find a sentence or series of several words without punctuation, unusual characters, or suspicious words that may have replaced the original text.

2. Search using Google by putting the entire sentence or phrase within quotation marks inside the search box. For example, try searching for the sentence [“Many details are omitted or altered while many of the perils that Dorothy encountered in the novel are not at all mentioned in the feature film”] or the phrase [“timid Munchkins come out of hiding to celebrate the demise”]. Sometimes, it is helpful to try the same search without the quotes, e.g. [timid Munchkins come out of hiding to celebrate the demise].

3. Compare the pages you find that match the sentence or phrase. Is most of their Main Content the same? If so, does one clearly come from a highly authoritative source which is known for original content creation (newspaper, magazine, medical foundation, etc)? Does one source appear to have the earliest publication date? Does one source seem to reasonably be the original?

Use your best judgment. Sometimes it is clear that the content is copied from somewhere, but you cannot tell what the original source is. Or sometimes the content found on the original source has changed enough that searches for sentences or phrases may no longer match the original source. For example, Wikipedia articles can change dramatically over time. Old copies may not match the current content. If you strongly suspect the page you are evaluating is not the original source, go ahead and use the Low quality Main Content rating.

Sometimes content is intentionally revised to make it difficult to determine that the content has been copied. Content may even be put through a translator to revise it. For example, if the original content is in English, it may be put through a translator twice: first to change it to a foreign language and second to translate it back to English. Text that has been changed in this was will often sound nonsensical.

Any time you find copied content or suspect the page has copied content, please explain in the comment box. Please include the original source (URL or description) if you are able to find it.

We will now walk you through two examples to determine if the content is copied.

Example 1 – No clear original source

1. [Example 2.3.3](#).   There is a paragraph at the top, followed by a line. Then there is an article below. Notice something a bit funny? Instead of the word "flower" or "flowers", this article uses "f" or "fs". This page looks suspicious. Let's try to figure out if the content is copied from elsewhere. Let's use the sentence: "Flowers which last only one day, like day lilies, do not dry well." It does not have the odd "fs" abbreviation.

2. Do a search on Google with that sentence in quotes: ["Flowers which last only one day, like day lilies, do not dry well."].

You will see that there are many webpages with this sentence, though most use the word "flowers" rather than "fs". In fact, if you go to the last page of Google results, you'll find this:

"In order to show you the most relevant results, we have omitted some entries very similar to the 25 already displayed. If you like, you can repeat the search with the omitted results included."

Clicking on the blue link will give you over a hundred results for this sentence, many of which contain this article text or links to pages containing this article text. But none of these results seems to be very authoritative, and no single web result looks like it is the original source of the article.

In some cases, even though the original source of content may be difficult or impossible to identify, you can still be fairly certain you have a copy. In the example above, it's highly unlikely that this is the original source of the content. It has the odd "fs" and "f" abbreviations (probably done so that search engines will not detect that this is copied content). There are also copying errors and other alterations on this page. Look at the bottom and you'll see "For Ber Colors: Rapid drying in a very warm, dry and bly-lit place will produce b blossoms; s drying in a more humid spot will produce more muted colors."

While we cannot be sure of the original source, this is clearly a copy. It's actually less helpful than an unaltered copy. The abbreviations make it difficult to understand the text in places. If you see similar issues when rating, please make a note in your comments.

Example 2 – Clear original source

1. Example5.3.1. This example is from an actual Page Quality rating task. Some raters indicated that this page has copied content. Does it? Let's continue through the steps to see.

2. Search for this phrase: "A master of creating the illusion of three-dimensional forms and figures on flag walls." You will find many results. We should try to determine if there is one plausible original source for this content.

3. Let's start with our URL. On the landing page, click the "about us" link to see who is responsible for the content of the website: Example5.3.2. You'll find this information which shows that this is a very authoritative source for the content:

> Since the laying of the Capitol cornerstone by George Washington in 1793, the
> Architect of the Capitol (AOC) has served the United States as builder and steward
> of many of the nation's most iconic and indelible landmark buildings. These include
> the U.S. Capitol, Capitol Visitor Center, Senate Office Buildings, House Office Buildings,
> Supreme Court, Library of Congress, U.S. Botanic Garden and Capitol Grounds.

Now let's look at this result which came up when we searched for the phrase on Google: Example5.3.3.This is a far less authoritative source. In addition, the text at the bottom of the article cites "the Architect of the Capitol" at the bottom in the Credits section: "Images and descriptions online, courtesy Architect of the Capitol."

There are other copies of this article on the Web, but we can see that the original URL is a page on a highly authoritative website for this content and that other sources cite this page. At this point, we can conclude that Example5.3.1 is the original source of the article, though there are many copies on other websites.

Raters have requested more guidance on how to assign URL ratings for high and low quality pages. Here are some basic principles. Please use your best judgment.

- **URL ratings** are based on both the landing page AND the query in a task. You must think carefully about the user intent and then assign a rating based on the helpfulness of the page. **Page Quality rating**, on the other hand, is query-independent. The rating you assign for Page Quality does not depend a query. In fact, Page Quality rating tasks do not have a query.

- In **URL rating**, if the landing page is useless for the query for any reason, it should receive the **Off-Topic or Useless** rating. High quality pages may be useless for a query because they are off topic. Extreme low quality can also make a page useless. In other words, pages that are on-topic but extremely low quality can be given **Off-Topic or Useless** ratings in URL rating.

- High quality pages should not always receive high URL ratings. Just because a page is high quality does not mean it will be helpful for a query. A high quality page can be rated from **AV** to **OT**, depending on the fit to the query. Here are some examples of high quality pages that should receive low URL ratings: A high quality page about zebras when the query is [horses]; a high quality page about the 2004 US presidential elections when the query is [2008 US presidential elections]; a high quality page selling pencils from a New Zealand website when the query is [buy pencils], English (US).

- Usually, on-topic low quality pages should be rated lower on the URL rating scale than on-topic higher quality pages because low quality pages are usually not very helpful for users. For example, consider a medical query such as [food poisoning symptoms].   An authoritative high quality page from a reliable medical website about the topic should probably be rated **Useful**, whereas a low quality, poorly written article on a non-medical website by an anonymous author should probably be rated **Slightly Relevant** or possibly even **Off-Topic or Useless**.

- The **Useful** rating should be given to helpful, high quality pages which are also a good fit for the query. The **Useful** rating may also be used for results that are medium quality but are a good fit for the query and have other very desirable characteristics, such as very recent information.

- The **Appropriate Vital** rating is special and does not depend on the quality of the page. If the query requests a specific page, then that page should get the **Appropriate Vital** rating, even if it is low or even lowest quality.

7.0   Page Quality Rating FAQs

Question	Answer
Why do we have to do all these steps? This takes forever!	With practice, the amount of time needed for accurate ratings will decrease. The steps are important and are designed to help you assess many different aspects of Page Quality. You may be surprised by what you find. Pages which initially look low quality may turn out to be medium or high quality with careful inspection. The reverse may happen as well. We want your most informed, thoughtful opinion.
Are we just giving high quality ratings to pages that "look" good?	No! The goal is to do the exact opposite. These steps are designed to help you analyze the page without using a superficial "does-it-look-good?" approach.
You talked about expertise when rating Main Content. Does expertise matter for all topics? Aren't there some topics for which there are no experts?	Remember - we are not just talking about expertise. High quality pages involve time, effort, expertise and/or talent. But since you ask, pretty much any topic has some form of experts, though there are some topics or types of pages where expertise is less important than other aspects for Main Content quality rating.

For most page purposes and most topics, you can find experts, even when the field itself is niche or non-mainstream. For example, there are expert alternative medicine websites with leading practitioners of acupuncture, herbal therapies, etc. There are also pages about alternative medicine written by people with no expertise or experience. The Main Content quality ratings should distinguish between these two scenarios. |
| Aren't there some types of pages that always have low quality content? | For almost any type of page, there is a range of content quality. Remember - high quality content is defined as content that is very satisfying, useful, or helpful for its purpose.

For example, there are high quality celebrity gossip pages and low quality celebrity gossip pages. Often, the purpose of a gossip page is to share scandalous plausibly true personal information about well known people, so the Main Content of a gossip page is high quality when it is juicy and from a somewhat plausible source. Gossip pages are not judged by their accuracy. And, yes, there are high quality gossip pages! |
I've never seen a high quality page of type X. If there are no high quality pages of this type, why are we giving existing pages a low quality rating?	For some topics or page purposes, there may not be many (or any!) high quality pages **now**, but in the future there may be. We need a uniform set of standards that apply to all pages, even for pages that have not yet been created.
Some of these criteria seem unfair. For example, some art pages do not have a purpose. Are these pages low quality?	Art pages have a purpose: artistic expression. Certainly, pages created for artistic expression do not deserve the low quality rating simply because they have no other purpose. Artistic expression, humor, and entertainment are valid page purposes.
Are forum pages always low quality?	No. Forum pages vary. We need to evaluate forum pages using the same criteria as all other pages. There are some forum pages with detailed information on specific issues written by people who are experts in the topic being discussed. There are also shallow discussion threads with very little content. No type of page (news, forum, shopping, encyclopedia, etc.) is automatically high or low quality.
Are Q&A pages necessarily low quality?	No. Q&A pages vary. We need to evaluate these pages with the same criteria as all other pages. Sometimes, it can be difficult to assess the accuracy of the information or the expertise/knowledge of the person answering the question. In these cases, you may need to do a bit of research. If the page is asking for medical advice, be skeptical about the expertise of the participants in the discussion. If the question is about daily life, then it is far more likely that the participants in the discussion have the necessary experience/expertise.

Some Q&A pages are detailed and have accurate and reliable information. Many others have little participation or inaccurate/incomplete information. On some Q&A pages, the question itself is incomplete or fragmented. We must evaluate these from the perspective of web users, not the participants in the discussion on the page.

Remember - no type of page (such as news, Q&A, store, etc.) is automatically high or low quality. |

Part 4: Rating Examples

In this section, you will see examples of some of the types of queries and landing pages you will evaluate, along with suggested ratings. Most queries can be categorized as action, information, or navigation (**do-know-go**), but many queries fall into more than one category. As you work on URL rating tasks, remember that you must always consider user intent and how helpful the landing page would be for users who issue the query.

1.0 Named Entity Queries

Some queries are for named entities. Different types of named entities include:

- People (celebrities, public figures, ordinary people, etc.)
- Geographic locations (a country, a region, a state, a province, a county, a city, etc.)
- Famous locations (monuments, tourist attractions, natural wonders, etc.)
- Companies, products, and brand names (IBM, Apple iPod, Nintendo, Toyota Camry, etc.)
- Organizations and other institutions (United Nations, The World Bank, Harvard University, etc.)
- Books, shows, movies, musical pieces ("War and Peace", "Mission Impossible", Handel's "Messiah", etc.)
- Events (the Olympics, a marathon, a lottery drawing, a sweepstakes, etc.)

	[Elton John], English (US)
Query Description	▪ *Elton John is a famous English singer and songwriter*
Likely User Intent	▪ ***Do*** *– Purchase music performed by Elton John* ▪ ***Know*** *– Users want information or news about Elton John* ▪ ***Go*** *– Users want to go to the homepage of Elton John's official website*
Vital	▪ The homepage of Elton John's official and well maintained website: http://web.eltonjohn.com/index.jsp
Useful – helpful for most users	▪ Quality pages with biographical or good general information, such as this Wikipedia page about Elton John: http://en.wikipedia.org/wiki/Elton_John and http://www.imdb.com/name/nm0005056/ ▪ Elton John's official YouTube channel page which has links to many videos performed by him: http://www.youtube.com/eltonjohn ▪ Amazon page with Elton John MP3 downloads and CDs for sale: http://www.amazon.com/s/ref=nb_sb_noss_1?url=search-alias%3Daps&field-keywords=elton+john ▪ Elton John's official Facebook page which has a lot of information that would be interesting to most of his fans: http://www.facebook.com/EltonJohn
Relevant – helpful for many or some users	▪ A timely article about Elton John ▪ Amazon page with Elton John's "The Greatest Hits 1970-2002" album, or individual songs from the album", to download: http://www.amazon.com/The-Greatest-Hits-1970-2002/dp/B000WTXD30/ref=sr_1_1?ie=UTF8&qid=1340050305&sr=8-1&keywords=elton+john. The page also offers over 200 reviews of the album. **Useful** is also acceptable for this album with 34 of his greatest hits. ▪ Lyrics website page with links to lyrics of many Elton John songs, organized by album: http://www.azlyrics.com/j/john.html. Most lyrics websites are low or medium quality. **Relevant** is an appropriate for most lyrics pages.
Slightly Relevant – helpful for few users	▪ An outdated article about Elton John ▪ Lyrics website page with the lyrics of one Elton John song: http://www.azlyrics.com/lyrics/eltonjohn/candleinthewind.html ▪ Low quality page with an article about Elton John: http://www.eltonjohngreatesthits.net/
Off-Topic or Useless – helpful for very few or no users	▪ Very low quality page with keyword stuffing including multiple displays of "elton john" and "elton john tickets": http://cheapest-elton-john-tickets.download-418-91217.datapicks.com/

	[Nicole Kidman], English (US)
Query Description	*Nicole Kidman is a well-known, award winning movie star. She is in the news frequently because of her acting career, and also because of her previous marriage to Tom Cruise and her current marriage to singer Keith Urban.*
Likely User Intent	▪ **Know** – *Users want information, news, video clips, pictures, etc. related to Nicole Kidman* ▪ **Go** – *Users want to go to the homepage of Nicole Kidman's official website*
Vital	▪ The homepage of Nicole Kidman's official and well maintained website: http://nicolekidmanofficial.com/
Useful – helpful for most users	▪ Quality pages with biographical or good general information about Nicole Kidman, such as http://www.imdb.com/name/nm0000173/. Such pages might include a biography, filmography, pictures, etc. ▪ A very high quality personal fan page ▪ A page with many images of Nicole Kidman, such as http://images.search.yahoo.com/search/images;_ylt=A0geup.yzVBMzyIAIftXNyoA?ei=UTF-8&p=nicole+kidman
Relevant – helpful for many or some users	▪ A short article with timely information about Nicole Kidman ▪ A video of Nicole Kidman in an ad for Chanel: http://www.youtube.com/watch?v=yTO4FHf8MBs
Slightly Relevant – helpful for few users	▪ An outdated, unimportant article about Nicole Kidman, such as http://www.smh.com.au/news/people/nicole-kidman-cup-cancelled/2007/05/15/1178995148978.html
Off-Topic or Useless – helpful for very few or no users	Note: The names of well-known actresses and personalities are often used to draw users to spam and porn pages. The following page is **Off-Topic or Useless** and should be assigned a **Spam** flag: http://www.nicolekidman.org.

	[A O Smith], English (US)
Query Description	*A.O. Smith is a company that makes electric motors, water heaters & storage tanks.*
Likely User Intent	▪ **Go** – *Users want to go to the company's official homepage* ▪ **Do** – *Users want to purchase products manufactured by the company* ▪ **Know** – *Users want information about the company*
Vital	▪ Corporate homepage for A.O. Smith http://www.aosmith.com/
Useful – helpful for most users	▪ A.O. Smith division webpages at http://www.aosmithmotors.com/ and http://www.hotwater.com/ ▪ Pages that sell, distribute, or review multiple A.O. Smith products. **Relevant** may also be acceptable, depending on how helpful the page is. ▪ A page with current news articles about A.O. Smith, such as http://www.google.com/news/search?aq=f&pz=1&cf=all&ned=us&hl=en&q=a+o+smith
Relevant – helpful for many or some users	▪ Helpful subpages on the A.O. Smith website, such as the webpage for investors at http://investor.shareholder.com/aosmith/ ▪ A current news article about A.O. Smith ▪ A.O. Smith's Facebook page: http://www.facebook.com/pages/A-O-Smith/220554620563
Slightly Relevant – helpful for few users	▪ Outdated article about the A.O. Smith company ▪ Subpages on the A.O. Smith website, which would not be helpful to most users, such as: http://www.aosmith.com/Governance/Detail.aspx?id=328&ekmensel=c580fa7b_14_0_328_3 ▪ Amazon product review written by someone named A.O. Smith, http://www.amazon.com/gp/cdp/member-reviews/A3CWREGQNQJAQD?ie=UTF8&sort_by=MostRecentReview. Since it is very unlikely that this page would be helpful to the user who typed the query, **Off-Topic or Useless** is also an acceptable rating.
Off-Topic or Useless – helpful for very few or no users	▪ "About us" page for David Smith, a pharmacist associated with A&O Pharmacy in Salinas, California. http://www.aopharmacy.com/about_us.htm

[For Other Living Things in Sunnyvale], English (US)	
Query Description	*For Other Living Things is a pet supply store in Sunnyvale, California.*
Likely User Intent	▪ ***Go*** *– Users want to go to the official homepage of the company* ▪ ***Do*** *– Users want to make a purchase* ▪ ***Know*** *– Users want information about the store*
Vital	▪ Official homepage at http://www.forotherlivingthings.com/
Useful – helpful for most users	▪ Directory pages with contact information, a map, and reviews about the store, such as: http://www.yelp.com/biz/for-other-living-things-sunnyvale or http://local.yahoo.com/info-21336044-for-other-living-things-sunnyvale
Relevant – helpful for many or some users	▪ Helpful pages on the website, such as: http://www.forotherlivingthings.com/contact_us.php, http://www.forotherlivingthings.com/about_us.php, and http://www.forotherlivingthings.com/all-products-c-142.html ▪ A directory page with contact information: http://www.zvents.com/sunnyvale-ca/venues/show/125217-for-other-living-things ▪ The company's Facebook page: http://www.facebook.com/pages/Sunnyvale-CA/For-Other-Living-Things/96204195772? **Useful** is also acceptable.
Slightly Relevant – helpful for few users	▪ Subpage that would not be helpful to most users: http://www.forotherlivingthings.com/privacy.php ▪ A page about guinea pigs that mentions the store and has a link to the company's website: http://community.babycenter.com/journal/wheekergal/685/are_guinea_pigs_the_right_pet_for_your_kids
Off-Topic or Useless – helpful for very few or no users	▪ Page with a 2006 article about cat behavior written by Marilyn Krieger, who teaches cat behavior classes at For Other Living Things. **Slightly Relevant** is also an acceptable rating for this page.

[Perkins], English (US)	
Query Description	*There are many companies and people with the name Perkins.*
Likely User Intent	▪ ***Go*** *– Users want to go to the official homepage of the Perkins Restaurant & Bakery chain, the dominant interpretation, or to the official homepage of another entity with the Perkins name* ▪ ***Know*** *– Users want information about Perkins Restaurant & Bakery, other companies with the Perkins name, or people with the Perkins name*
Vital	▪ Official homepage of Perkins Restaurant & Bakery at http://www.perkinsrestaurants.com/, the dominant interpretation of the query
Useful – helpful for most users	▪ Official homepages of common interpretations for this query, such as: http://perkins.com, homepage of Perkins Engines, and http://www.perkins.org/, homepage of Perkins School for the Blind ▪ Subpages on the Perkins Restaurant website which would be helpful to many or some people, such as the locations subpage, and http://www.perkinsrestaurants.com/menu, the menu subpage. **Relevant** is also acceptable for thèse two subpages.
Relevant – helpful for many or some users	▪ Official homepages of less common or minor interpretations, such as: http://www.perkinsmedicalsupply.com/, homepage of Perkins Medical Supply, a small company, and http://www.ed.gov/programs/fpl/index.html, homepage of the Federal Perkins Loan Program ▪ Wikipedia article about Perkins restaurant ▪ Timely articles about Perkins restaurant
Slightly Relevant – helpful for few users	▪ Subpages on the Perkins Restaurant website, which would not be helpful to most users, such as http://www.perkinsrestaurants.com/privacy ▪ Outdated news articles about the Perkins restaurant ▪ The homepage of someone whose last name is Perkins. Since no first name is specified in the query, a higher rating is not appropriate.
Off-Topic or Useless – helpful for very few or no users	▪ Video of a private birthday party at a Perkins Restaurant: http://www.youtube.com/watch?v=TZuvYSOsHug

	[iphone], English (US)
Query Description	*The iPhone is a popular mobile smartphone made by Apple.*
Likely User Intent	▪ ***Do*** *– Users want to purchase an iPhone* ▪ ***Know*** *– Users want information (reviews, specifications, features, etc.) about the iPhone* ▪ ***Go*** *– Users want to go to the official product page on the Apple website*
Vital	▪ The iPhone page on the Apple website: http://www.apple.com/iphone/
Useful – helpful for most users	▪ The Apple website homepage: http://www.apple.com/ ▪ The Apple Store page on the Apple website: http://store.apple.com/us ▪ The iPhone page of the Apple Store: http://store.apple.com/us/browse/home/shop_iphone/family/iphone?mco=OTY2ODA2OQ ▪ High quality sites that review or provide comprehensive information on the iPhone, such as http://www.cnet.com/apple-iphone.html ▪ The AT&T page where users can purchase the iPhone: http://www.att.com/wireless/iphone/ ▪ The Apple iPhone discussion board: http://discussions.apple.com/category.jspa?categoryID=201
Relevant – helpful for many or some users	▪ Page with many iPhone accessories for sale ▪ A timely article about the iPhone ▪ A helpful video about the iPhone, such as http://www.youtube.com/watch?v=IpQ9RESJnWM ▪ A Wikipedia article about the iPhone, http://en.wikipedia.org/wiki/Iphone
Slightly Relevant – helpful for few users	▪ Review about the HTC Touch phone that mentions the iPhone ▪ Outdated article on the iPhone ▪ The MacPro page on the Apple website: http://www.apple.com/macpro/. There is a link on the page for the iPhone, but the page is not about the iPhone. Acceptable ratings are **Slightly Relevant** and **Off-Topic or Useless.**
Off-Topic or Useless – helpful for very few or no users	▪ Page about a different type of smartphone, such as: http://www.sonyericsson.com/cws/products/mobilephones/overview/p990i

	[Honda Pilot], English (US)
Query Description	*The Pilot is a popular Honda SUV.*
Likely User Intent	▪ ***Do*** *- Users want to purchase a Honda Pilot* ▪ ***Know*** *– Users want information (reviews, specifications, features, etc.) about the Honda Pilot* ▪ ***Go*** *– Users want to go to the official Pilot page on the Honda site*
Vital	▪ The official Pilot page on the Honda site
Useful – helpful for most users	▪ The automobiles page on the Honda website: http://automobiles.honda.com/ ▪ High quality pages that review or provide comprehensive information about the current model of the Honda Pilot, such as http://www.edmunds.com/honda/pilot/review.html ▪ The Insurance Institute for Highway Safety (IIHS) page about the Honda Pilot: http://www.iihs.org/ratings/ratingsbyseries.aspx?id=391. **Relevant** would also be acceptable.
Relevant – helpful for many or some users	▪ High quality pages with comprehensive information about previous year models of the Honda Pilot, such as: http://autos.aol.com/honda-pilot-2007:8689-overview. If the information is more than a year or two old, **Slightly Relevant** is also appropriate. ▪ A relatively short article about the current year's Honda Pilot ▪ A Wikipedia article on the Honda Pilot, http://en.wikipedia.org/wiki/Honda_Pilot
Slightly Relevant – helpful for few users	▪ Shopping page for Pilot headlights and fog lights: http://shopping.yahoo.com/s:Headlights:4168-Brand=Pilot ▪ Amazon page with Honda Pilot repair manual for sale: http://www.amazon.com/Honda-Pilot-Acura-MDX-Haynes/dp/1563926903
Off-Topic or Useless – helpful for very few or no users	▪ High quality page about the Honda Civic: http://www.edmunds.com/honda/civic/review.html, a different Honda vehicle

	[Nevada], English (US)
Query Description	*Nevada is one of the 50 states in the United States. Many people visit Nevada, especially the city of Las Vegas.*
Likely User Intent	▪ **Do** – Users want to make travel plans and reservations ▪ **Know** - Users want general information about Nevada or travel and tourism information ▪ **Go** - Users want to navigate to the official Nevada government website
Vital	▪ The official homepage for the state of Nevada: http://www.nv.gov/
Useful – helpful for most users	▪ The state of Nevada's official travel and tourism website: http://travelnevada.com/ ▪ High quality, comprehensive pages about Nevada: http://en.wikipedia.org/wiki/Nevada ▪ High quality travel and tourism pages for Nevada, such as http://travelnevada.com/ and http://travel.yahoo.com/p-travelguide-191501966-nevada_vacations-i
Relevant – helpful for many or some users	▪ Homepages of Nevada's flagship universities: University of Nevada, Las Vegas and University of Nevada, Reno: http://www.unlv.edu/ and http://www.unr.edu/home/ ▪ Page with facts about Nevada: http://leg.state.nv.us/General/NVFacts/index.cfm ▪ Wikipedia page with links to other pages about specific Nevada cities: http://en.wikipedia.org/wiki/List_of_cities_in_Nevada
Slightly Relevant – helpful for few users	▪ IMDb page for a movie titled "Nevada Smith": http://www.imdb.com/title/tt0060748/. **Off-Topic or Useless** is also acceptable. ▪ Homepage of the Nevada Republican Party: http://www.nevadagop.org/ ▪ Outdated article about an election in Nevada.
Off-Topic or Useless – helpful for very few or no users	▪ Homepage for the UCMT Family of Schools, which has massage therapy schools in Utah, Nevada, Arizona, and Colorado: http://www.ucmt.com/

	[Chicago], English (US)
Query Description	*Chicago is a big city in the United States.*
Likely User Intent	▪ ***Do*** – Users want to make travel plans and reservations for visiting Chicago ▪ ***Know*** – Users want travel and tourism information or general information about Chicago ▪ ***Go*** – Users want to navigate to the official Chicago city government website *When a city (or state, country, etc.) is a major travel destination, it is likely that the users want to make travel plans or find information. However, if the city (or state, country, etc.) has an official page, that page should get a **Vital** rating.*
Vital	▪ The official homepage for the city of Chicago: http://www.cityofchicago.org/city/en.html
Useful – helpful for most users	▪ High quality pages with helpful travel & tourism information, such as http://www.choosechicago.com/Pages/default.aspx ▪ High quality pages about Chicago: its history, climate, travel, culture, public transportation, etc., http://www.lonelyplanet.com/worldguide/usa/chicago and http://en.wikipedia.org/wiki/Chicago ▪ An excellent blog or collection of personal information, which would be helpful to someone visiting the city, such as http://www.gochicagocard.com/blog/ ▪ A comprehensive collection of high quality images of the city of Chicago, http://images.google.com/images?q=chicago&sourceid=navclient-ff&ie=UTF-8&rls=GGGL,GGGL:2006-33,GGGL:en&um=1&sa=N&tab=wi ▪ A high quality map of the city, such as http://travel.yahoo.com/p-map-191501928-map_of_chicago_il-i ▪ Official homepage of Chicago, the band, http://www.chicagotheband.com/
Relevant – helpful for many or some users	▪ Homepage for the main regional newspaper, Chicago Tribune, at http://www.chicagotribune.com/. ▪ Homepages of large, prominent entities that most users would associate with the city of Chicago, such as The University of Chicago at http://www.uchicago.edu/, The Chicago Bulls at http://www.nba.com/bulls/, the Chicago Cubs at http://chicago.cubs.mlb.com/, etc. ▪ YouTube Channel page of Chicago's official tourism site: http://www.youtube.com/user/explorechicago ▪ Videos of the band "Chicago" performing in concert, such as http://www.youtube.com/watch?v=QECAViP4U1Y&feature=PlayList&p=59E9DEA4BBF87639&index=2
Slightly Relevant – helpful for few users	▪ Local weather forecasts for Chicago, http://www.wunderground.com/US/IL/Chicago.html ▪ Homepages of universities or businesses in the Chicago area that are not as closely associated with the city, such as Northwestern University, http://www.northwestern.edu/ ▪ Homepages of other newspapers that cover the Chicago area, but are not the "main" newspaper of the city, such as http://www.chicagoweeklynews.com/
Off-Topic or Useless – helpful for very few or no users	▪ Webpage of the summer music program at Northwestern University (a university located just outside Chicago), http://www.music.northwestern.edu/summer/ ▪ Video of the Blue Brothers performing the song, "Sweet Home Chicago", http://www.youtube.com/watch?v=Tlou_2lMLAc

	[white house], English (US)
Query Description	*The residence and workplace of the President of the United States is called the White House.*
Likely User Intent	▪ **Go** *– Users want to go to the official White House page* ▪ **Know** *– Users want information about the White House*
Vital	▪ The official page of the White House on the US government website: http://www.whitehouse.gov
Useful – helpful for most users	▪ The President's page on the official White House site: http://www.whitehouse.gov/administration/president-obama/ ▪ Pages on the official White House website that would be helpful to many users, such as the Briefing Room subpage (http://www.whitehouse.gov/briefing-room) and the White House Blog subpage: (http://www.whitehouse.gov/blog) ▪ Wikipedia page about the White House: http://en.wikipedia.org/wiki/White_House ▪ White House Twitter page: http://twitter.com/whitehouse **Relevant** is also acceptable.
Relevant – helpful for many or some users	▪ Pages on the official White House website that would be helpful to some users, such as: http://www.whitehouse.gov/about/white-house-101/ and http://www.whitehouse.gov/about/ ▪ Homepages of common or somewhat minor interpretations, such as the homepage of this city in the state of Tennessee: http://www.cityofwhitehouse.com/ . **Slightly Relevant** is also acceptable.
Slightly Relevant – helpful for few users	▪ Pages on the official White House website which would be helpful to few users, such as this page with a 2003 memo about privacy and cookies at http://www.whitehouse.gov/omb/memoranda_m03-22/#20 ▪ Homepages of minor interpretations, such as the homepage of The White House Federal Credit Union: (http://www.whcu.org/home.aspx) and the homepage of White House Florist (http://www.whitehouseflower.com/)
Off-Topic or Useless – helpful for very few or no users	▪ A page about removing white house paint from brown boots: http://www.answerbag.com/q_view/507910

	[whitehouse.gov], English (US)
Query Description	*This is a special type of query, which we refer to as a URL query. The query is the URL of the official White House webpage.*
Likely User Intent	▪ **Go** *– Users want to go to http://www.whitehouse.gov*
Vital	▪ The official page of the White House on the US government website: http://www.whitehouse.gov
Useful – helpful for most users	▪ The President's page on the official White House site: http://www.whitehouse.gov/administration/president-obama/, which is very similar to the White House page, and possibly matches user intent
Relevant – helpful for many or some users	▪ Pages on the official White House site that would be helpful to some users
Slightly Relevant – helpful for few users	▪ Wikipedia page about the White House, which has a link to the official website: http://en.wikipedia.org/wiki/White_House ▪ Pages on the official White House website which would be helpful to few users.
Off-Topic or Useless – helpful for very few or no users	▪ The homepage of the White House Restaurant in Laguna Beach, California at http://www.whitehouserestaurant.com/

2.0 Action Queries

When typing an action query, users are trying to accomplish a goal or engage in an activity, such as to download software, play a game online, send flowers, find entertaining videos, etc. These are "**do**" queries: users want to do something. Here are some examples of action queries:

- Download software for free or for money
- Purchase a product
- Pay a bill online
- Play a game online
- Take an online survey
- Print a calendar
- Send flowers
- Organize photos or order prints online
- Find a video clip
- Copy an image or piece of clipart
- Take an online personality test

	[adobe reader download], English (US)
Query Description	*Adobe Reader software allows the user to view and print PDF files.*
Likely User Intent	• *__Do__ – Users want to download Adobe Reader* • *__Know__ – Users want information about Adobe Reader* • *__Go__ – Users want to go to the download page on the Adobe website*
Vital	• Adobe Reader download page on official Adobe website: http://get.adobe.com/reader/
Useful – helpful for most users	• The Adobe homepage: http://www.adobe.com/. Reader is one of Adobe's most well-known products. **Relevant** is also acceptable.
Relevant – helpful for many or some users	• A page on a reputable website with information and reviews on Adobe Reader and a link to the download page on the Adobe website, such as http://www.download.com/Adobe-Acrobat-Reader/3000-2378_4-10000062.html. **Useful** is also acceptable.
Slightly Relevant – helpful for few users	• A Yahoo! Answers page with a user's explanation about what Adobe Reader does, and which has a link to Adobe: http://answers.yahoo.com/question/index?qid=1005111000036
Off-Topic or Useless – helpful for very few or no users	• A page about the Omea Reader, a free RSS reader: http://www.jetbrains.com/omea/reader/

	[text twist], English (US)
Query Description	*TextTwist is a popular computer game that can be played online or downloaded.*
Likely User Intent	▪ **Do** – *Users want to play the game online or download it (for free or for a fee)*
Vital	▪ None possible
Useful – helpful for most users	▪ Pages where users can play or download the game, such as http://get.games.yahoo.com/proddesc?gamekey=texttwist
Relevant – helpful for many or some users	▪ An article which contains tips for playing the game, such as http://videogames.lovetoknow.com/wiki/Text_Twist_Tips_and_Strategies
Off-Topic or Useless – helpful for very few or no users	▪ A page on which to download Tetris, a different computer game.

	[take an online personality test], English (US)
Query Description	*Personality tests help people to understand their behavior and can help them learn what type of career they might be suited for*
Likely User Intent	▪ **Do** – *Users want to take an online personality test for free or for money*
Vital	▪ None possible
Useful – helpful for most users	▪ Online personality tests based on the famous Myers-Briggs Type Indicator which identifies 16 distinct personality types, such as http://www.humanmetrics.com/cgi-win/Jtypes2.asp and http://kisa.ca/personality/
Relevant – helpful for many or some users	▪ A very short online personality test, based on the famous Myers-Briggs personality test, at http://www.personalitytype.com/quiz.html ▪ The website of a company that offers the Myers-Briggs Type Indicator online for a fee, and offers clients many kinds of reports based on test results. The company's clients include many well-known US corporations. http://www.knowyourtype.com/
Slightly Relevant – helpful for few users	▪ An online personality test that helps identify personality disorders. There is no way to tell anything about the quality of the test. http://www.4degreez.com/misc/personality_disorder_test.mv
Off-Topic or Useless – helpful for very few or no users	▪ A page that offers "The Original Internet Love Test", a test that predicts compatibility between two people. http://www.lovetest.com/

	[skateboarding dog video], English (US)
Query Description	*There are videos on the Web of dogs using skateboards*
Likely User Intent	▪ **Do** – *Users want to watch a video of a skateboarding dog*
Vital	▪ None possible
Useful – helpful for most users	▪ Pages on video websites with highly entertaining skateboarding dog videos that would be interesting to many users, such as http://www.youtube.com/watch?v=ziDeUbifKIM, http://www.youtube.com/watch?v=i3T3sYZ9eBk and http://www.metacafe.com/watch/914414/skateboarding_dog_amazing_funny/
Relevant – helpful for many or some users	▪ Pages on video websites with somewhat entertaining skateboarding dog videos that would be interesting to some users, such as http://www.metacafe.com/watch/925757/barney_the_skateboarding_dog/ , http://uk.youtube.com/watch?v=nhE9Y1tEwQw&NR=1, andhttp://uk.youtube.com/watch?v=tlx-AdIR7ew
Slightly Relevant – helpful for few users	▪ A video of a skateboarding dog made out of clay: http://www.youtube.com/watch?v=WVUoTigp7qo, which would be interesting to few users.
Off-Topic or Useless – helpful for very few or no users	▪ A video of a person skateboarding, such as: http://www.youtube.com/watch?v=UMg44qXLaNw

3.0 Information Queries

When typing an information query, users are trying to find information. These are "**know**" queries: users want to know something. For many information queries, it would be difficult to imagine user intents other than looking for information. Below are some examples of information queries.

Please note that in the last two information query examples, a page exists that warrants a rating of **Vital**. User intent is to find information, and these pages provide exactly what users are looking for on the official, authoritative page associated with the query. Even when user intent is to find information that can be found on many pages on the Web, a **Vital** rating is sometimes possible.

	[retina and laser surgery], English (US)
Query Description	*Laser surgery can be performed on the retina to treat a variety of retinal problems.*
Likely User Intent	▪ ***Know*** *– Users want information about laser surgery for the retina*
Vital	▪ None possible
Useful – helpful for most users	▪ Pages from high quality sources providing information on laser surgery for the retina, http://www.kellogg.umich.edu/patientcare/conditions/detached.retina.html ▪ Newsgroups or message boards which are focused on the subject and would be very helpful to users, such as http://www.afb.org/message_board_replies2.asp?TopicID=3067&FolderID=14
Relevant – helpful for many or some users	▪ Individual retinal laser surgery practitioner pages that provide information on the topic, such as http://www.socalretina.com/html/procedures.html ▪ Wikipedia page on eye surgery that discusses many types of eye surgery, including laser retina surgery: http://en.wikipedia.org/wiki/Eye_surgery ▪ Yahoo! Answers page on the topic of the query: http://au.answers.yahoo.com/answers2/frontend.php?question?qid=20070724160757AAHmLJy ▪ Article on diabetic retinopathy that discusses laser treatment: http://www.solomoneyeassociates.com/procedures/diabetic_eye_treatment.htm
Slightly Relevant – helpful for few users	▪ Site that describes a retinal fellowship program: http://www.maculasurgery.com/Fellowship%20Goals.htm
Off-Topic or Useless – helpful for very few or no users	▪ Sites about laser surgery and acne: http://www.lasersurgery.com/acne/ ▪ Sites about a type of eye surgery that does not involve the use of lasers, such as http://en.wikipedia.org/wiki/Strabismus_surgery

	[what can I do with coffee grounds], English (US)
Query Description	*Used coffee grounds do not need to be thrown away; there are many uses for them.*
Likely User Intent	▪ ***Know*** *– Users want information about uses for coffee grounds*
Vital	▪ None possible
Useful – helpful for most users	▪ Pages (including FAQs and message board pages) with advice on many ways to use coffee grounds (deodorizer, fertilizer, dye, etc.), such as http://www.gomestic.com/Homemaking/10-Uses-for-Used-Coffee-Grounds.75800
Relevant – helpful for many or some users	▪ Pages that provide one or just a few tips for using coffee grounds, http://www.goodhousekeeping.com/home/heloise/kitchen/recycle-coffee-grounds-sep06
Slightly Relevant – helpful for few users	▪ A page that discusses whether coffee grounds can be put down a garbage disposal, which includes a suggestion that coffee grounds can be composted, http://wiki.answers.com/Q/Can_you_put_coffee_grounds_in_a_garbage_disposal
Off-Topic or Useless – helpful for very few or no users	▪ Online directory listing for a restaurant called "The Coffee Grounds" in St. Paul, Minnesota: http://phoenix.citysearch.com/profile/1701833/tempe_az/coffee_grounds.html

[HTML lessons], English (US)	
Query Description	HTML stands for HyperText Markup Language, the markup language for the creation of most webpages.
Likely User Intent	▪ **Do** – Users want to take on online tutorial on HTML ▪ **Know** - Users want pages that provide information about using HTML
Vital	▪ None possible
Useful – helpful for most users	▪ Pages that offer lessons, step-by-step instructions, or tutorials for learning HTML, such as http://www.utexas.edu/learn/html/ and http://www.w3schools.com/html/default.asp
Relevant – helpful for many or some users	▪ Pages that offer short tutorials on using HTML
Slightly Relevant – helpful for few users	▪ A Wikipedia page with good information about HTML and links to tutorial pages: http://en.wikipedia.org/wiki/HTML
Off-Topic or Useless – helpful for very few or no users	▪ Pages that offer lessons or tutorials for learning XML, not HTML, such as http://www.w3schools.com/xml/default.asp ▪ An article that discusses HTML 5, a major upgrade to HTML, but does not provide lessons, http://www.news.com/World-Wide-Web-Consortium-releases-draft-of-HTML-5/2100-1007_3-6227721.html

[map collins ave south beach], English (US)	
Query Description	South Beach is a section of Miami Beach, Florida. Collins Avenue is a major street in Miami Beach.
Likely User Intent	▪ **Know** – Users want a map of South Beach that displays Collins Avenue.
Vital	▪ None possible
Useful – helpful for most users	▪ Map that shows the South Beach area of Miami Beach, and identifies Collins Avenue, such as http://www.miamibeach411.com/maps_south_beach.html
Slightly Relevant – helpful for few users	▪ Map that shows the South Beach area of Miami Beach, but does not identify Collins Avenue without zooming in, http://miami.citysearch.com/profile/map/11344117/miami_beach_fl/south_beach.html ▪ Wikipedia page about South Beach that does not display a map, but which discusses north-south and east-west roads, including Collins Avenue, http://en.wikipedia.org/wiki/South_Beach
Off-Topic or Useless – helpful for very few or no users	▪ Map finder page in which users can type "Collins ave, south beach, fl" in the search box and get a map of the area, such as http://maps.yahoo.com/ .

[international telephone codes], English (US)	
Query Description	Every country has a country calling code (dialing prefix) that is dialed before the telephone number when calling that country.
Likely User Intent	▪ **Know** – Users want a list of country calling codes
Vital	▪ None possible
Useful – helpful for most users	▪ Pages that provide a comprehensive set of international calling codes, such as http://en.wikipedia.org/wiki/List_of_country_calling_codes ▪ A page that describes how to dial an international call and provides a link to a page with a list of country calling codes, http://www.wiktel.com/standards/howdial.htm
Relevant – helpful for many or some users	▪ Pages with international telephone codes, but for Europe only, http://www.europe.org/dialingcodes.html
Slightly Relevant – helpful for few users	▪ A page that describes how to call to and from just one country, such as http://www.japan-guide.com/e/e2223_how.html
Off-Topic or Useless – helpful for very few or no users	▪ A page with a United States National Area Code Map: http://www.whitepages.com/maps. Area codes in the US are not the same as country calling codes.

	[enable javascript ie], English (US)
Query Description	"ie" is an abbreviation for Internet Explorer, which is Microsoft's web browser. The most current version is Internet Explorer 8.
Likely User Intent	▪ **Do** – Users want to enable JavaScript in Internet Explorer ▪ **Know** – Users want to learn how to enable JavaScript in Internet Explorer ▪ **Go** – Users want to go the a page in the Microsoft website to find this information
Vital	▪ Page on Microsoft's website that tells how to enable JavaScript in Internet Explorer: http://support.microsoft.com/gp/howtoscript
Useful – helpful for most users	▪ Pages on other reputable websites that provide detailed instructions on enabling JavaScript in Internet Explorer, such as http://kb.iu.edu/data/ahqx.html and http://gsaauctions.gov/brow_details/IE6instr.htm
Relevant – helpful for many or some users	▪ Page with detailed instructions for enabling JavaScript in Internet Explorer versions 5, 6, and 7, but not 8: http://www.tranexp.com/win/JavaScript-enabling.htm. This page would be helpful for some or few users. **Slightly Relevant** is also acceptable.
Slightly Relevant – helpful for few users	▪ Page on low quality site with basic instructions for enabling JavaScript in Internet Explorer versions 3 through 6, but not 7 or 8.
Off-Topic or Useless – helpful for very few or no users	▪ Pages that tell users how to enable JavaScript in browsers other than Internet Explorer, such as http://kb.iu.edu/data/aeet.html

	[Louvre visiting hours], English (US)
Query Description	The Louvre is a famous museum in Paris.
Likely User Intent	▪ **Know** – Users want to find the museum's visiting hours ▪ **Go** – Users want to find this information on the official Louvre website
Vital	▪ Visiting hours page on the site of the Louvre at http://www.louvre.fr/llv/pratique/horaires.jsp?bmLocale=en
Useful – helpful for most users	▪ A page from a reputable travel website that provides visiting hours and other useful information http://www.frommers.com/destinations/paris/A25285.html
Relevant – helpful for many or some users	▪ Official homepage of the Louvre. The page does not display the visiting hours, but there is a link to the "Visit" section of the website. http://www.louvre.fr/llv/commun/home.jsp?bmLocale=en
Slightly Relevant – helpful for few users	▪ A page from a museum guidebook that displays the Louvre's hours, but in 24-hours time (which US users are less familiar with). **Relevant** is also acceptable for this page. http://www.holidaycheck.com/things_to_do-travel-information+The+Louvre-zid_7700.html
Off-Topic or Useless – helpful for very few or no users	▪ General travel information about Paris with a brief mention of the Louvre, but no reference to visiting hours, http://www.tripadvisor.com/Tourism-g187147-Paris_Ile_de_France-Vacations.html ▪ Wikipedia page on the Louvre, which does not provide visiting hours or even have a link to a page with visiting hours. . http://en.wikipedia.org/wiki/Louvre

4.0 Queries that Ask for a List

After typing a query, the search engine user sees a result page. You can think of the results on the result page as a list. Sometimes, the best results for "queries that ask for a list" are the best individual examples from that list. The page of search results itself is a nice list for users.

A landing page that provides links to many good individual results can also be very helpful to users.

"Queries that ask for a list" may be typed in singular or plural form. For example, the query may be [bank], English (US) or [banks], English (US).

Here are some examples of queries that ask for a list:

	[credit cards], English (US)
Query Description	*In the United States, most credit cards are issued by financial institutions or organizations, and most of these are affiliated with one of the major credit card associations: Visa, MasterCard, etc.*
Likely User Intent	▪ ***Do*** *– Users want to sign up for a credit card online* ▪ ***Know*** *– Users want to research credit cards before signing up*
Vital	None possible
Useful – helpful for most users	▪ Since the user has not specified a particular credit card association or financial institution, homepages of well-known credit card companies or issuers of credit cards in the US are **Useful**. **Relevant** is also acceptable. http://www.americanexpress.com/ http://www.usa.visa.com/personal/ http://www.mastercard.com/us/gateway.html http://www.citicards.com/cards/wv/home.do http://www.discovercard.com/ ▪ Pages on reputable sites that offer credit card comparisons, such as: http://moneycentral.msn.com/banking/services/CreditCard.asp
Relevant – helpful for many or some users	▪ Pages with information about how credit cards work, such as http://www.howstuffworks.com/credit-card.htm ▪ Pages on reputable sites with information about credit cards, such as http://www.ftc.gov/bcp/menus/consumer/credit/loans.shtm
Slightly Relevant – helpful for few users	▪ The credit card application page for a credit card that requires union membership, such as http://www.unionplus.org/benefits/money/card.cfm ▪ The credit card application page for a company that issues cards to permanent Australian residents only, http://virginmoney.com.au/credit_card/. **Off-Topic or Useless** is also acceptable.
Off-Topic or Useless – helpful for very few or no users	▪ College webpage that tells students that a convenience fee is charged when tuition payments are made with a credit card: https://tuitionpay.salliemae.com/tuitionpay/tpphome.aspx?csusm

	[banks], English (US)
Query Description	*Banks are financial institutions that offer services to individuals and businesses. There are many well-known national banks, as well as many smaller regional/local banks in the United States.*
Likely User Intent	**Do** – Users want to open a bank account **Know** – Users want to research banks before opening a bank account
Vital	None possible
Useful – helpful for most users	▪ Since the user has not specified a particular bank, homepages of well-known banks in the US are **Useful**. **Relevant** is also acceptable. Here are some examples (there are many others): http://www.citibank.com/ https://www.bankofamerica.com/ http://www.chase.com/ ▪ Website with links to banks in the United States, organized by state: http://www.thecommunitybanker.com/bank_links/
Relevant – helpful for many or some users	▪ Official government webpage that displays contact information for US Federal Reserve Banks, http://www.federalreserve.gov/fraddress.htm ▪ The homepage of a small regional bank, which serves communities in that region, http://www.albanybank.com/ . **Slightly Relevant** is also acceptable.
Slightly Relevant – helpful for few users	▪ The homepage of a bank in another country, such as http://www.barclays.co.uk/. **Off-Topic or Useless** is also acceptable. ▪ Outdated article on bank interest rates, http://money.cnn.com/magazines/moneymag/moneymag_archive/2004/12/01/8192192/index.htm
Off-Topic or Useless – helpful for very few or no users	▪ An article about someone who was injured while washing the windows of a bank, http://www.wect.com/Global/story.asp?S=5841672

	[bikes], English (US)
Query Description	*Bikes, also known as bicycles, are two-wheel, human-powered vehicles that people use. There are different types of bikes, such as mountain, road, hybrid, comfort, recumbent, etc.*
Likely User Intent	▪ **Do** – Users want to purchase a bike ▪ **Know** – Users want to research bikes before making a purchase
Vital	None possible
Useful – helpful for most users	▪ Since the user has not specified a particular bike manufacturer, homepages of well-known bike manufacturers would be **Useful**. **Relevant** is also acceptable. Here are some examples (there are many others): http://www.schwinnbike.com/usa/eng/ http://www.trekbikes.com/us/en/ http://www.specialized.com/us/en/bc/home.jsp ▪ Pages on reputable sites with a wide range of bikes for sale, such as http://www.amazon.com/s/ref=nb_sb_noss?url=search-alias%3Daps&field-keywords=bikes and http://www.rei.com/category/4500003_Bicycles ▪ Pages on reputable sites with a comprehensive list of bike reviews or information about many bikes
Relevant – helpful for many or some users	▪ Pages with information about how bikes work , such as http://www.howstuffworks.com/bicycle.htm
Slightly Relevant	▪ The "privacy policy" subpage on the Trek website, http://www.trekbikes.com/us/en/general/privacy_policy/ ▪ Homepage of ConferenceBike, manufacturer of a bike that can be ridden by seven riders, http://www.conferencebike.com/
Off-Topic or Useless – helpful for very few or no users	▪ Article that talks about children putting playing cards in the spokes of their bicycle wheels in the 1930s and 1940s, http://www.otal.umd.edu/~vg/amst205.F97/vj14/cards/children.html

	[airlines], English (US)
Query Description	*There are many airline companies that operate in the United States and throughout the world.*
Likely User Intent	▪ **Do** – *Users want to purchase airline tickets* ▪ **Know** – *Users want to find information (such as prices and schedules) before purchasing tickets*
Vital	▪ None possible
Useful – helpful for most users	▪ Homepages of online travel companies that offer flights on numerous airlines. Here are some examples (there are many others): http://www.orbitz.com/ http://www.expedia.com/ http://www.travelocity.com/ ▪ Since the user has not specified a particular airline, homepages of well-known US airline companies would be **Useful** or **Relevant**. Here are some examples (there are many others): http://www.united.com/ http://www.aa.com/ http://www.usairways.com/ https://www.southwest.com/ ▪ The Federal Aviation Administration's page of links to US airline companies: http://www.fly.faa.gov/FAQ/Airline_Links/airline_links.jsp ▪ Wikipedia page with links to airlines that operate in the United States: http://en.wikipedia.org/wiki/List_of_airlines_of_the_United_States
Relevant – helpful for many or some users	▪ Homepages of major airlines not based in the US. **Slightly Relevant** is also acceptable. http://www.alitalia.com/us_en/?no http://www.jal.co.jp/en/ ▪ Wikipedia page that contains a list of airlines, organized by continent and country: http://en.wikipedia.org/wiki/List_of_airlines
Slightly Relevant – helpful for few users	▪ A two-year old article that discusses rumors about mergers between US airline companies.
Off-Topic or Useless – helpful for very few or no users	▪ The homepage of a company that gives airplane tours of the Grand Canyon, http://www.airgrandcanyon.com/

	[hotels], English (US)
Query Description	There are many hotel companies that operate in the United States and throughout the world.
Likely User Intent	▪ **Do** – Users want to make a hotel reservation ▪ **Know** – Users want to find information about hotels before making a reservation
Vital	▪ None possible
Useful – helpful for most users	▪ Since the user has not specified a particular hotel, homepages of well-known hotel chains would be **Useful**. **Relevant** is also acceptable. Here are some examples (there are many others): http://www.radisson.com/ http://www.hilton.com/ http://www.marriott.com/ ▪ Homepages of online hotel and travel companies that allow users to make reservations with many different hotel chains: http://www.hotels.com/ http://www.orbitz.com/ http://www.expedia.com/ http://www.travelocity.com/
Relevant – helpful for many or some users	▪ Websites that allow users to make reservations with many different bed and breakfast inns, which are a specific type of hotel. **Slightly Relevant** is also acceptable. http://www.bedandbreakfast.com/ http://www.bbonline.com/ ▪ Wikipedia page with general information about hotels: http://en.wikipedia.org/wiki/Hotels. **Slightly Relevant** is also acceptable.
Slightly Relevant – helpful for few users	▪ Page about hotel chains in India: http://www.indfy.com/hotel-chains-of-india/
Off-Topic or Useless – helpful for very few or no users	▪ Wikipedia page about the song "Hotel California": http://en.wikipedia.org/wiki/Hotel_California_(song)

	[London Boutiques], English (US)
Query Description	Boutiques are small specialty shops.
Likely User Intent	▪ **Do** – Users want to shop at a boutique in London ▪ **Know** – Users want information about boutiques in London
Vital	▪ None possible
Useful – helpful for most users	▪ Pages with good information about many London boutiques, such as http://www.timeout.com/london/shopping/features/2067/London-s_50_best_boutiques.html. Such pages might include pictures, addresses, descriptive information, reviews, price ranges, store hours, maps, etc. ▪ Map result page displaying information about many London boutiques, such as http://maps.google.com/maps?f=l&view=text&q=boutique&near=London%2C+United+Kingdom&btnG=Search+Businesses
Relevant – helpful for many or some users	▪ A review of an individual London boutique, with address and contact information, such as http://www.frommers.com/destinations/london/S27883.html . **Slightly Relevant** is also acceptable.
Slightly Relevant – helpful for few users	▪ Outdated article (February 1999) titled: "London's Top 15 Boutiques" - http://www.travelandleisure.com/articles/cheaper-and-chicer/1
Off-Topic or Useless – helpful for very few or no users	▪ A travel page about boutiques in Paris, not London: http://www.francetoday.com/travel/paris/listings/boutiques.html

5.0 Rating Examples for Task Locations other than English (US)

	[IBM], English (IN)
Query Description	*IBM (International Business Machines) is a multinational computer technology company with offices around the world.*
Likely User Intent	▪ ***Go*** *– Users want to go the IBM India website.*
Appropriate Vital	▪ IBM India webpage: http://www.ibm.com/in/
International Vital	▪ "Choose your country/region and language" IBM webpage: http://www.ibm.com/planetwide/select/selector.html
Other Vital	▪ IBM Australia webpage: http://www.ibm.com/au/en/ ▪ IBM Spain webpage: http://www.ibm.com/es/es/ ▪ IBM China webpage: http://www.ibm.com/cn/zh/
Useful – helpful for most users	▪ IBM India "profile" page, which has contact information and information about the various groups and facilities in India: http://www.ibm.com/ibm/in/en/
Relevant – helpful for many or some users	▪ India IBM contact information page: http://www.ibm.com/contact/in/ ▪ Wikipedia article about IBM India: http://en.wikipedia.org/wiki/IBM_India ▪ 2011 news article about IBM India's revenues: http://articles.economictimes.indiatimes.com/2011-06-01/news/29608432_1_ibm-india-ibm-japan-revenues-cross
Slightly Relevant – helpful for few users	▪ 2007 news article about an increase in IBM's India headcount: http://news.zdnet.co.uk/itmanagement/0,1000000308,39285764,00.htm
Off-Topic or Useless – helpful for very few or no users	▪ Homepage of HP India: http://welcome.hp.com/country/in/en/welcome.html

	[Match], English (UK)
Query Description	*There are two equally likely interpretations for this query for U.K. users: Match, the online dating company and Match, the British football magazine*
Likely User Intent	▪ ***Go*** *– Users want to go either* http://uk.match.com/ *or* http://www.matchmag.co.uk/
Vital	▪ *Since neither interpretation is clearly dominant, no **Vital** rating is possible.*
Useful – helpful for most users	▪ U.K. Match dating company webpage: http://uk.match.com/ ▪ Homepage of Match, the football magazine: http://www.matchmag.co.uk/
Relevant – helpful for many or some users	▪ Homepage of Match, research collaboration between five leading UK universities: http://www.match.ac.uk/ . **Useful** is also acceptable. ▪ Wikipedia article about the football magazine: http://en.wikipedia.org/wiki/Match_magazine ▪ Wikipedia article about the dating company: http://en.wikipedia.org/wiki/Match.com ▪ Wikipedia article about matches that people use to light a fire: http://en.wikipedia.org/wiki/Match ▪ "Match of the Day" football page on the BBC website: http://news.bbc.co.uk/sport1/hi/football/match_of_the_day/default.stm
Slightly Relevant – helpful for few users	▪ Careers webpage for the dating company which shows jobs in the US: http://uk.match.com/careers/index.aspx
Off-Topic or Useless – helpful for very few or no users	▪ Wikipedia page about the musical, "Fiddler on the Roof". One of the characters in the musical is a matchmaker: http://en.wikipedia.org/wiki/Fiddler_on_the_Roof.

	[Sephora], English (CA)
Query Description	*Sephora is a beauty supply company that sells products online and in stores around the world.*
Likely User Intent	▪ **Go** – Users want to go the Sephora website
Appropriate Vital	▪ Canada Sephora webpage: www.sephora.com/canada
International Vital	▪ "Choose your country" Sephora webpage: http://www.sephora.com/international.jhtml
Other Vital	▪ US Sephora homepage: http://www.sephora.com/ ▪ France Sephora homepage: http://www.sephora.fr/ ▪ Italy Sephora homepage: http://www.sephora.it/
Useful – helpful for most users	▪ Canada Sephora Store Locator webpage: http://www.sephora.com/help/stores/allStores.jhtml?country=canada. **Relevant** is also acceptable.
Relevant – helpful for many or some users	▪ Yelp map/review page with information about the Toronto Sephora store: http://www.yelp.ca/biz/sephora-beauty-canada-toronto ▪ Amazon.ca page with Sephora beauty guide book for sale: http://www.amazon.ca/Sephora-Ultimate-Makeup-Beauty-Authority/dp/0061466409 **Slightly Relevant** is also acceptable. ▪ Wikipedia article about Sephora: http://en.wikipedia.org/wiki/Sephora **Slightly Relevant** is also acceptable.
Slightly Relevant – helpful for few users	▪ Checkout page on Canada Sephora website: https://www.sephora.com/secure/arc20/richCheckout.jhtml;jsessionid=ZXBKWD2KQ0NBICV0KRTQQAQ
Off-Topic or Useless – helpful for very few or no users	▪ Homepage for FabaoCanada, a different Canadian beauty supply company: http://www.fabaocanada.com/

	[Orange], French (FR)
Query Description	*Orange is a French telecommunications company*
Likely User Intent	▪ **Go** – Users want to go the Orange website
Appropriate Vital	▪ Orange homepage for consumers: http://www.orange.fr
International Vital	▪ Top level page in English: http://www.orange.com/
Other Vital	▪ Austria Orange homepage: http://www.orange.at/Content.Node/
Useful – helpful for most users	▪ Mobile subpage: http://mobile-shop.orange.fr/ ▪ Internet subpage: http://abonnez-vous.orange.fr/residentiel/accueil/accueil.aspx
Relevant – helpful for many or some users	▪ Orange corporate homepage: http://www.orange.com/fr_FR/index.jsp. Most users would be more interested in the consumer homepage, so this page should not get a **Vital** rating. **Useful** is also acceptable. ▪ Women's page: http://femmes.orange.fr/ ▪ News page: http://actu.orange.fr/ ▪ Wikipedia article about Orange: http://actu.orange.fr/
Slightly Relevant – helpful for few users	▪ 2009 press release about high-definition voice service for mobile phones in Moldova: http://www.orange.com/en_EN/press/press_releases/cp090910en.jsp
Off-Topic or Useless – helpful for very few or no users	▪ Article about jobs in Orange County in California: http://www.ocregister.com/articles/economy-259910-improve-flexible.html

Part 5: Webspam Guidelines

1.0   What is Webspam ?

Webspam is the term for webpages that are designed by webmasters to trick search engines and draw users to their websites. In these guidelines, we sometimes refer to webspam as "spam", and webmasters who use deceptive techniques as "spammers".

In the coming pages, you will learn how to identify some of these deceptive techniques. When you see them being used, you will assign a **Spam** flag. Please note that pages that are merely annoying, junky, or low quality, such as pages with lots of pop-ups or ads, are not necessarily spam.

1.1   The Relationship between Ratings and Spam

In the "Rating Guidelines", you learned that landing pages are rated according to their utility to users for a particular query. You would not be able to assign a rating to a page without knowing the query.

Spam flags do **not** depend on a relationship between the query and the landing page. A page should get a **Spam** flag if it is created using deceptive techniques - no matter what the query is or how helpful the page might be.

Some spam pages are very low quality and have little or no content which would be helpful for users.   These pages will usually be assigned a low rating, either **Slightly Relevant** or **Off-Topic or Useless**, in addition to the **Spam** flag.

Other spam pages, which are not as low quality and have some helpful content, may be assigned a rating of **Slightly Relevant** or **Relevant**.

In some specific cases, it is also possible for a page to receive a **Vital** rating, and also be assigned a **Spam** flag.   For example, if there is a sneaky redirect and the landing page is the target of the query, the page will get a **Vital** rating and a **Spam** flag. You will learn about "sneaky redirect" spam in Section 3.3.

1.2   Why do Spammers Create Spam Pages?

Spammers create spam pages to make money. Sometimes, they make money directly, by placing moneymaking links on the spam page. Here are two types of moneymaking links:

- **Pay-Per-Click (PPC)** ads: Spammers get paid each time ads are clicked on their webpages. Another term for PPC ads is "sponsored links".
- **Thin Affiliates**: Spammers make money when a transaction is completed after the user has clicked through to the merchant's site from their webpages. A thin affiliate typically doesn't add much value compared to many other sources of information on the Web, and often sends the user to another website to complete the actual purchase.

PPC ads appear on many, many webpages. Some pages with PPC ads are spam, but many pages with PPC ads are not. Pages should **not** be assigned a **Spam** flag if they are created to provide information or help to users. Pages are spam if they exist primarily to make money and not to help users.

Sometimes, spam pages do not have moneymaking links. These spam pages are created to change search engine rankings or even to do harm to users' computers with sneaky downloads. They are spam because they use deceptive techniques, even though you are unable to see how they are making money.

1.3   When to Check for Spam

There are some pages, such as the main page of a well-known website (e.g. http://www.apple.com), that you may feel do not need to be evaluated for spam. However, even webmasters for highly reputable websites occasionally use deceptive techniques. Therefore, we ask that you use the following two quick and easy spam detection techniques on **all** webpages that you evaluate.

- Apply "Ctrl-A" (or apply "⌘" and "A" for Apple computer users) to the landing page to look for hidden text. You will learn about using "Ctrl-A" in Section 3.1.1.
- Scroll all the way down and to the right on the page to look for hidden text on areas of the page outside the normal viewing area. You will learn more about hidden text outside the normal viewing area in Section 3.1.5.

You should use the other spam detection techniques described in these guidelines when you feel the page needs further investigation.

Throughout the Webspam Guidelines, you will be given links to spam URLs that you can use to practice spam detection techniques. Please be aware that spam pages can change very quickly. Sometimes, they change from one type of spam to another type. Sometimes, the pages just stop loading. Because spam pages change so quickly, you will also be given links to screenshot examples. You can "walk through" the spam examples using the live links (if they work) and/or by clicking the "Screenshot Example" links. You may notice that some examples fall into more than one spam category.

2.0   Browser Requirement

Unless told otherwise in the project-specific instructions, from now on you must do ALL of your rating work in Firefox. You must not use any other browser for your rating work.

By rating work, we mean doing query research, viewing tasks in EWOQ, submitting tasks in EWOQ, etc. You must not use any other browser for any aspect of your rating work.

Here are some of the benefits of using Mozilla Firefox:

- Mozilla offers a Firefox Add-on called "Web Developer", which provides you with a special toolbar containing tools helpful in spam detection. The two buttons on the toolbar that will probably be the most helpful are the "Disable" button, which allows you to quickly disable JavaScript, and the "CSS" button, which allows you to quickly disable CSS (Cascading Style Sheets). You will learn how these tools will help you to detect spam in a later section of these guidelines. Here is a link to download the Web Developer toolbar, if you would like to do so: https://addons.mozilla.org/en-US/firefox/addon/60

- Firefox allows you to add tabs for webpages, which can be helpful in web browsing and spam detection. Here is a description of this Firefox feature: http://www.mozilla.com/en-US/firefox/tabs.html. Customizing your browser in this way will allow you to quickly navigate to pages that you visit frequently and save you time. Using tabs will also allow you to open different versions of the same page, which can be helpful in spam detection. Specifically, you will be able to load versions of a page before and after disabling JavaScript and CSS, and then toggle between them to see the differences.

3.0   Looking for Technical Signals

When evaluating a page for spam, you should start by looking for the following "technical signals":

- Hidden text and hidden links
- Keyword stuffing
- Sneaky redirects
- Cloaking with JavaScript redirects and 100% frame

This section describes these technical signals and provides tips and tools on how to identify them.

3.1   Hidden Text and Hidden Links

Webmasters add hidden text and/or hidden links to lure search engines and users to their pages. Hidden text is visible to the search engine, but not to the user, who might find it distracting or annoying. Here are some things you should know about hidden text:

- It may be completely invisible to the human eye.
- It may be in the same color as the background color on the page, or in a color that is so close to the background color that it almost invisible and will not be noticed.
- It may be formatted in a very, very small font size (e.g., 1-point) so that it will not be noticed.
- It may be placed outside the normal viewing area. For example, there may be a large blank space between the normal viewing area and a "hidden" area of text all the way at the bottom of the page or far to the right.
- Sometimes there is just a line or two of hidden text, but you may even see a whole page of it.
- Most hidden text is there to trick the search engine, but occasionally you will find hidden text that is not spam. For example, if the webmaster merely hides the date of an update, it is not spam.

Hidden text may be revealed by:

- Applying Ctrl-A (or "⌘" and "A" for Apple computer users)
- Disabling CSS
- Disabling JavaScript
- Viewing the source code
- Looking outside the normal viewing area

3.1.1   Apply Ctrl-A to the Landing Page

After you have clicked on the URL, simultaneously press the **"Ctrl" and "A"** keys (the keyboard shortcut for "Select All" for PC users), or **"⌘"** and **"A"** or "**Command**" and "**A**" (the keyboard shortcuts for Apple computer users) and then scroll down the whole page. This technique sometimes reveals text that has been hidden.

Using Ctrl-A to reveal hidden text
Screenshot Example

Tiny text is not always exposed using Ctrl-A. You should be suspicious of horizontal lines or bars on the page because sometimes they contain hidden text. A simple technique for revealing this type of hidden text is to select and copy the suspicious line or bar, paste it in your word processor, and increase the font size. You may also try using the techniques described below.

3.1.2   Disable CSS

Disabling CSS sometimes reveals hidden text. Here are instructions for disabling CSS using the Web Developer toolbar:

1. Click on "CSS".
2. On the dropdown menu, click on "Disable Styles".
3. Click on "All Styles".

You do not need to check every page for hidden text in CSS, but please do check if the page is suspicious. If you download the Web Developer toolbar, you will find it is simple to use.

Disabling CSS to reveal hidden text
Screenshot Example

3.1.3 Disable JavaScript

Spammers sometimes use JavaScript to hide text. Here are instructions for disabling JavaScript using the Web Developer toolbar:

1. Click on "Disable".
2. On the dropdown menu, click on "Disable JavaScript".
3. Click on "All JavaScript".
4. Refresh the page.

You can also disable JavaScript using your browser menu in Firefox; however, it takes more steps and more time than using the Web Developer toolbar:

Disabling JavaScript using your browser window in Firefox:
1. Go to "**Tools**". 2. Click on "**Options**". 3. Click on "**Content**" or "**Web Features**". 4. To disable JavaScript, make sure the "**Enable**" box is not unchecked. 5. Click "**OK**".

Disabling JavaScript to reveal hidden text
Screenshot Example

Important: When you are done looking for spam on a particular page, please remember to go back and enable JavaScript. If you do not do this, certain features on pages you open will not work.

3.1.4 View the Source Code

Viewing the source code sometimes reveals hidden text.

Viewing Source Code in Firefox:
1. Go to "**View**". 2. Click on "**Page Source**".   or 1. Right click on the page. 2. Click on "**View Page Source**".

Here is an example of hidden text that is revealed by viewing the source code. Look for large areas of keyword stuffing in the source code. Keyword stuffing is discussed in Section 3.2.

Viewing Source Code to find hidden text
Screenshot Example

Please note that a **Spam** flag should *not* be assigned when the keyword stuffing appears in the meta tags only. Meta tags are easy to identify because they start with the words "meta name". Here is an example:

Not Hidden Text: Keyword stuffing in the meta tags only
Screenshot Example

3.1.5 Look Outside the Normal Viewing Area

Be suspicious of large blank areas on the bottom and far right portions of the page. Use the vertical and horizontal scroll bars to see if it appears there is text on the portion(s) of the page outside the main viewing area.

3.2 Keyword Stuffing

Keyword Stuffing: Webmasters sometimes load pages with keywords that are related to the query. Here are descriptions of what you might see:

- Keywords repeated many times on the page
- Words that are related to keywords repeated many times on the page
- Multiple misspellings of keywords on the page

Webmasters also sometimes load pages with *irrelevant* keywords on topics that are unrelated to the query, such as mortgages, cell phones, ringtones, gambling, weather, etc.

Whether the keywords are related or unrelated to the query, the intent is to draw search engines and users to the page.

It is sometimes difficult to decide when the keywords on a page should be considered keyword stuffing. We ask you to assign a **Spam** flag if you think the number of keywords on the page is excessive and would be annoying and distracting to the real user. If you do not feel the number of keywords would bother the user, please do not assign a **Spam** flag.

Please note: Hidden text and keyword stuffing often go together. Hidden text frequently contains keyword stuffing.

Recognizing keyword stuffing

Some keyword stuffing is visible to the human eye and you will not have to use any special techniques to see it. In other cases, it is hidden. You will discover *hidden* keyword stuffing by using the techniques in Section 3.1.1. **Important:** *hidden* keyword stuffing will **always** be considered spam (unless it is only in the source code meta tags).

Here are some examples that most users would consider excessive and annoying, even though in some cases the keywords are in the portion of the page "below the fold", which users would have to scroll down to see:

Keyword Stuffing Examples	
Fake Feed Example	Screenshot Example
Fake Blog Example	Screenshot Example
Computer-Generated Text Example	Screenshot Example

3.2.1 Keyword Stuffing in the URL

URLs may also contain keyword stuffing. These URLs are computer-generated based on the words in the query and are often formatted with many hyphens (dashes) in them. They are a strong spam signal.

Keyword Stuffing in the URL Examples
Screenshot Examples

Here are some additional examples of keyword stuffing in the URL. We have removed the hyperlinks from these examples because some of them have stopped working and others have become malicious. You do not need to click through to the landing page in order to see that there is keyword stuffing in the URL and that they are spam.

- http://frat-boy-blog-gay.grandbrooklynlodge.cn/boy-brief-frat-in-their-wet.html
- http://brazilian-model-alexandra.wantloweryour.cn/brazilian-model-adriana-lima.html
- http://where-do-hot-girls-hang-in-philadelphia.heartlandvalleymiles.cn/hang-it-all.html

3.3 Sneaky Redirects

Sneaky Redirects: We call it a sneaky redirect when a page redirects the user from a URL on one domain to a different URL on a different domain, *with spam intent*. Search engines "see" the first page, while the user is sent to a different page and sees different content. Here are some other things you should know about sneaky redirects:

- While being redirected, you may notice that the page redirects through several URLs before ending up on the landing page.
- Sneaky redirects may take the user to one of several rotating domains; so clicking on the same URL several times may send you to different landing pages each time.
- Some sneaky redirects take users to well-known merchant websites, such as Amazon, eBay, Zappos, etc.

Recognizing sneaky redirects

- **Compare the two URLS:** Compare the URL in the rating task to the URL of the landing page to see if it makes sense that one would redirect to the other. A redirect from a company's old homepage to its new homepage on a different domain is not sneaky. Redirects from one page on a domain to another page on the same domain are also not sneaky.
- **Look at the domain registrants:** If you suspect that a sneaky redirect has taken place, you should check to see "who is" the registrant (or owner) of the two domains. If the registrant is the same, the redirect is not sneaky. Please see Section 3.3.1 for instructions on checking "who is".

3.3.1 Using "Whois"

Here are instructions for checking "who is" the domain registrant:

1. Go to the site of a "whois" provider. Here are two you can use: http://www.domaintools.com/ and http://whois.mtgsy.net/default.php
2. Enter the URL of one domain in the search box on the "whois" page. Sometimes, you will need to delete some leading or following characters. For example, if the URL is http://supportapj.dell.com/support/, you will enter just "dell.com" in the search box of the whois provider.
3. Open another "whois" page.
4. Enter the URL of the other domain in the search box on the second "whois" page.
5. Compare the domain registrants for the two URLs. If you find that they have the same domain registrant, you will conclude that the page is not spam. If they are different and do not seem related, it is probably spam.

Sneaky Redirect Example	
http://www.kqzyfj.com/go65biroiq57A8E7A6577BDAA6 redirects to http://www.jcwhitney.com/Auto-Parts/10101.jcw	Screenshot Example

Example of a Non-Sneaky Redirect
Screenshot Example

Please be aware that domains with the same domain registrant can look very different. For example, Barnes and Noble, the bookseller, owns the following domains: www.barnesandnoble.com, www.bn.com, and www.books.com.

3.4 Cloaking

It is called "cloaking" when the webmaster shows different pages to the search engine and the user. Two cloaking techniques used by spammers are:

- JavaScript redirects
- 100% frame

3.4.1 JavaScript Redirects

Spammers use JavaScript redirects to create two different pages. Looking at the page first with JavaScript enabled and then with JavaScript disabled reveals the differences.

3.4.2 100% Frame

Webmasters sometimes cloak what users see by using frames. Two frames (pages) exist, but one frame takes up 100% of the screen. The user sees one frame (page), but the search engine sees both frames. Here are instructions for looking at the different frames in Firefox:

Viewing Frame Information in Firefox
1. Right-click on the page. 2. Click "**This Frame**". 3. Click "**View Frame Info**". 4. Compare the URL of the frame with the URL of the page. If they are different, the page is probably 100% framed, and should be flagged as spam.

100% Frame Example
Screenshot Example

4.0 Helpful Webpages vs. Spam Webpages

Search engines want to display webpages that are helpful to users. In this section, you will learn how to determine if pages with ads on them are spam, or if they have utility to the user. We will talk about:

- Pages with PPC ads and other content, which *are designed to help users in some way*
- Pages with PPC ads and other content, which *only exist to make money*

Some pages contain PPC ads only, or have very, very little on them besides the PPC ads. We refer to these pages as "pure PPC" pages. You will learn more about pure PPC pages in Section 4.2. When the page containing PPC ads is created to be helpful to users, it is not spam. Here are examples of content that is helpful to users:

- Price comparison functionality: Some webpages offer price comparisons for shoppers looking to make a purchase. The shopper then has ability to take price into consideration. Even if the user has to click an affiliate link to go to another site to place the order, it is helpful to have price comparisons on the page.
- Product reviews: Some pages provide *original* product reviews that are helpful to the user in deciding whether to make a purchase. Items that are commonly reviewed are books, electronics, and hotels.
- Recipes: Some pages provide recipes. If the recipes on the page are helpful, for example, if the recipes are original or the page includes reviews of original or non-original recipes, the page is not spam.
- Lyrics, quotes, proverbs, poems, etc.: Some pages display this type of content. If the page is designed to help users find song lyrics or poems, etc., it is not spam.

- Contact information: Some pages provide contact information for companies. If the contact information includes physical addresses, phone numbers, maps, etc., the page is helpful and not spam.
- Coupon, discount, and promotion codes: Some affiliate pages provide coupon, promotion, or discount codes for the consumer, in addition to a link to the merchant. Since these types of codes are helpful to the user, they provide added value.

Please note that recipes, lyrics, quotes, poems, etc. do not usually have authoritative pages. Anyone can obtain and put this content on webpages.

4.1 Pages with Copied Content and PPC Ads

Copied content refers to content that has been copied from other sources. Webmasters sometimes use special "scraper" software to search the Web for content to put on their websites that is related to specific keywords. Content can also be taken from another website using the simple "copy and paste" method.

4.1.1 Copied Text and PPC Ads

Content that has been copied from sources such as Wikipedia (http://www.wikipedia.org/) and the Open Directory Project (http://www.dmoz.org/), sites that allow the distribution of their content and may even encourage it, is still considered to be copied content.

Copying content from such sources is not necessarily illegal, nor is it plagiarism. Webmasters who copy content usually do not claim to be original content creators and may, in fact, assign credit to the originator of the content. However, even if they do give credit to others, it is considered to be copied content.

These copies are often old, not updated, and may not be trustworthy. Users want information they can trust. A copy of a Wikipedia article on an unknown website accompanied by ads offers little utility to users. We will call a page spam if it is created to make money from ads on the page.

Copied Text Examples		
Wikipedia Example	Wikipedia URL: http://en.wikipedia.org/wiki/Magnetite Spam URL: http://www.nationmaster.com/encyclopedia/magnetite	Screenshot Example
DMOZ Example	DMOZ URL: http://www.dmoz.org/Computers/Security/ Spam URL: http://contentguarder.com	Screenshot Example

4.1.2 Feeds and PPC Ads

Web publishers (such as the BBC, CNN, Usenet, CNet, NYTimes, and others) publish information online that is readily available to users through RSS (Really Simple Syndication) and XML (Extensible Markup Language) feeds. Companies, such as Searchfeed.com, provide feeds of PPC ads and links to most qualifying webmasters.

A page that just contains freely available feeds and PPC ads, and was created just to make money, is spam.

4.1.3 Doorway Pages

Doorway pages are sets of pages that have been created for search engines to deliver the user to a common destination page. The pages all look very much the same and do not provide meaningful content for users. Please see the examples in the table below.
The top level URL http://www.hair-removal-hair-laser.com/ contains links for all of the states in the US. Clicking on a link makes you think that you are getting a customized page for that state, but if you click on another link, you will find that every page is really the same. These pages are spam. They are created to send users to a moneymaking page.

Doorway Pages Example		
Top level URL	http://www.hair-removal-hair-laser.com/	
California page URL	http://www.hair-removal-hair-laser.com/ca.html	
Florida page URL	http://www.hair-removal-hair-laser.com/fl.html	Screenshot Example
San Francisco page URL	http://www.hair-removal-hair-laser.com/City/California/Hair-removal-SanFrancisco.html	
Miami page URL	http://www.hair-removal-hair-laser.com/City/Florida/Hair_Removal_Miami_FL.html	

4.1.4 Templates and Other Computer-Generated Pages

Some websites use templates to mass-reproduce webpages automatically. The content is usually copied from sources that provide such content. You will learn to recognize templates, which usually follow a generic format or pattern. Look for slight keyword variations that suggest automated use of a keyword suggestion tool. If the keyword is "mortgage", you may see words such as "mortgages", "mortgage loan", "mortgages loans", etc. in the title, snippets, and/or URL

These spam pages contain links to other pages that usually contain some combination of copied content, PPC ads, and other spam links. Clicking on links on these pages will land you on other pages on the same domain with similar content and links.

Template Examples		
Computer-generated text	http://iponsel.com/ebook/hp-pavilion-dv2500-maintenance-and-service-manual/2008/05/01/	Screenshot Example
Computer-generated pages	Screenshot Example	

4.1.5 Copied Message Boards

Sometimes you will see copied message boards (user forums) and ads. When the page contains only the copied message board and PPC ads, the page is spam.

4.1.6 Recognizing Copied Content

Here are some things you can do to help you recognize copied content:

- **Search for an exact sentence from the text on the page:** Copy and paste a distinctive sentence in the search box of a search engine. When you paste the sentence in the search box, put quotation marks around it so that the search engine will search for the exact string of words. From the search results displayed, you may find where the content originated. If the content is original and has not been copied from another source, it probably was written to be helpful to users.
- **Look for PPC ads surrounding the content**. Wikipedia and DMOZ do not display ads. If you see Wikipedia or DMOZ content and PPC ads with no original content on the page, it is spam.
- **Become familiar with the format of Wikipedia and DMOZ pages**: The section headings and links on Wikipedia pages usually follow the same format. DMOZ pages use a directory pathway that is easy to recognize. In addition, DMOZ pages have these links: "submit a site" and "become an editor", which also appear on copied pages.

- **Look for suspicious, computer-generated grammar:** Look at the text on the page. When it is computer-generated, it often looks like "gibberish", which means that it does not make sense. You may also see hyperlinked keywords inside the text.
- **Look at URL formatting:** Look for URL formatting that suggests that a template or other automation was used to create it. Often, you will see keywords contained in the URL, separated by hyphens. Here is an example: http://nzealand.co.nz/blog/thelawmail/2007/12/29/com-search-extreme-belladonna-users-search-expired-domain-names-search-expired-domains/.
- **Look to see if the page appears to have been created to help users:** Look for features, such as lyrics, recipes, quotes, contact information, phone numbers, physical addresses, original reviews, a working comment box, etc.
- **Think about whether it seems as if the page was created by a human or by a machine:** Pages created by machines are usually not designed to be helpful for users and are usually spam.

4.2 Fake Search Pages with PPC Ads

A fake search page is a page with a list of links that looks like a page of search results. You will see a "search box" on the page, but if you submit a new query in the search box, you just get a different page of links. If you click on a few of the links, you will see that the page is just a collection of PPC links disguised as search engine results.

Fake Search Page Examples
Screenshot Examples

4.3 Fake Blogs with PPC Ads

A fake blog contains fake blog entries that are either nonsensical or copied from another source. Fake blogs often contain keyword stuffing, which is described in Section 3.2. The page exists so that the PPC links on the page will be clicked. PPC links may appear within the text of the fake blog entry, or on other parts of the page. Fake blogs may appear to allow the user to post a comment, but the feature does not work. Fake blogs are spam.

Spammed Blogs: Spammed blogs are different from fake blogs. A spammed blog is a real working blog with real blog entries, but has been spammed with entries that contain PPC ads and/or porn links. We do not want to penalize a blog because someone else has put spam on it. If you believe that the blog is a good, legitimate blog that has been spammed by someone else, please do **not** assign a **Spam** flag.

4.4 Fake Message Boards with PPC Ads

A fake message board is similar to a fake blog. It contains what appear to be "messages", but are not. The text in the message may be nonsensical or it may contain PPC links. Fake message boards may appear to have comment, registration, and login sections, but either these features do not work at all, or you are redirected back to the same page. On real message boards, you will see responses to posts. On fake message boards, either there are no responses, or the responses themselves are spam.

Fake Message Board Examples	
▪ http://www.cosmicscripts.com/boards/message/mainboard.html ▪ http://www.priyablue.com/msg/	Screenshot Examples

Copied Message Boards with PPC Ads: You may also find entire message boards that have been copied. If you suspect this has happened, copy and search for a snippet of text. Copied message boards are spam.

Spammed Message Boards: Spammed message boards are different from fake message boards. A spammed message board is a real message board with real posts and real responses, but which posts with PPC ads and/or porn links have spammed. We do not want to penalize a message board because someone has put spam posts up on it. If you believe the message board is a good, legitimate message board that has been spammed, please do **not** assign a **Spam** flag.

4.5 Copied Content that is NOT Spam

Some copied content is not spam. Here are some examples: lyrics, poems, proverbs, quotes, etc. This type of content has no unique or central authority.

If the page you are evaluating appears to be from a legitimate lyrics, poetry, etc. website, do not assign a **Spam** flag. If you think the page exists primarily to make money, you should assign a **Spam** flag.

5.0 Commercial Intent

In this section, we will talk about how spammers make money and how to look for commercial intent.

Most spam pages have commercial intent. Spammers create spam pages to make money and earn commissions when users make a purchase on an affiliate merchant site or when they click on a PPC ad.

If a page exists primarily to make money without sufficient added value for users, the page is spam.

Please remember: Some spam pages do not have obvious moneymaking intent. If a page is created to change search engine rankings or even to do harm to users' computers with sneaky downloads, it is spam even though you are unable to see how the page is making money.

5.1 Thin Affiliates

A thin affiliate is a website that earns money from affiliate commissions. It exists primarily to make money. The spammer shows content from other "real" merchant sites, such as Amazon or eBay, or a good hotel or travel website. When users click on links to buy products or make reservations, they are redirected to the "real" merchant page.

The thin affiliate offers little additional information and does not offer substantial value to users. This is a moneymaking spam technique.

5.1.1 Recognizing Thin Affiliates

To help determine if a page is a thin affiliate, you can do the following:

- **Click buttons on the page.** Click on a "More Information" or "Make a Purchase" button. If you are taken to a merchant on a different domain, it is probably a thin affiliate. You will not be able to make the purchase on the affiliate webpage.
- **Check properties of images on the page**. Right-click on an image on the page with your mouse and look at "Properties" to see where the image originates. Check to see if the address of the image is the same as the address of the page or if it is the address of a "real" merchant.
- **Look for original content on the page.** The quality of an affiliate page or site depends on how much added value, usefulness, or original/additional information is available on the page that is not easily available elsewhere on the Web. If the page has the same "cookie-cutter" text or functionality found on dozens or hundreds of other sites, it is more likely to be spam.
- **Look at the domain registrants.** If clicking a button takes you to another page, check to see "who is" the registrant (or owner) of the two domains. If the registrant is the same, the page is not a thin affiliate. Please follow the instructions for checking "who is" in Section 3.3.1.

5.1.2 Recognizing True Merchants

Features that will help you determine if a website is a true merchant include:

- a "view your shopping cart" link that stays on the same site
- a shopping cart that updates when you add items to it
- a return policy with a physical address
- a shipping charge calculator that works
- a "wish list" link, or a link to postpone the purchase of an item until later
- a way to track FedEx orders
- a user forum that works
- the ability to register or login
- a gift registry that works

Please note the following:

- A page does not need to have all of these features to be considered a true merchant.
- Yahoo! Stores are true merchants – they are not thin affiliates.
- Some true smaller merchants take users to another site to complete the transaction because they use a third party to process the transaction. These merchants are not thin affiliates.

Many large web retailers offer affiliate programs. Some of the most common examples are Amazon.com, eBay.com, Zappos.com, Allposters.com, Hotels.com, Orbitz.com, and Overstock.com. Here are some thin affiliate examples:

Thin Affiliate Examples		
ShoeMall Example	Thin affiliate URL: http://www.shoes.jalfrezi.com	Screenshot Example
Travel Site Example	Thin affiliate URL: http://www.travelnotes.org	Screenshot Example
Thin Affiliate on an Expired Domain Example	Screenshot Example	

5.2 Pure PPC Pages

We refer to pages with PPC ads only (or with PPC ads and very little other content on them) as pure PPC pages. The spammer makes money when a link is clicked. No purchase is necessary. Pure PPC pages may have links to other spam pages that also contain PPC ads. Pure PPC pages are spam.   Fake directory pages also can be considered pure PPC pages.

Pure PPC Example
Screenshot Example

5.3 Parked (Expired) Domains

Definitions of "Domain": The word "domain" can have two different meanings for raters:

- It can refer to one of the elements in the DNS (Domain Name System), such as .com, .org, .edu, .net, .gov, .it, .uk, .cn, .es, etc., that organize Internet addresses.

- It can refer to the set of words (URL) that identifies the web address of a specific entity, such as "microsoft.com", "harvard.edu", "baidu.cn", etc.

In this section, when we use the word "domain", we are referring to the second meaning.

When companies go out of business, are acquired by another company, change their name, or fail to pay their domain registration fee, the domain name "expires" and may be purchased by someone else.

Parked Domains: Spammers sometimes buy expired or expiring domains and put their own content on the page. Such sites are referred to as "parked domains" or "expired domains". Their value to spammers is in their pre-existing links. Pages that previously linked to the expired domain will now link to the spammer's page.

Spammers also purchase the following kinds of domains, which we will also refer to as parked domains, since they are similar in appearance:

- Domains which are close in spelling to real domains, hoping that users will mistype the domain name or URL and land on their websites, which contain PPC ads.
- Domains that users might type when looking for a website to use.

A typical parked/expired domain contains some or all of the following:

- A list of sponsored links
- A list of popular categories
- A list of categories that contains the keywords

Recognizing Parked/Expired Domains

- **Look at the links.** All of the links on a parked domain are paid links. There is no original content on the page.
- **Look at the domain name (URL).** On a parked domain, the domain name (URL) often has little or nothing to do with the content on the webpage. You may see the keywords, but the links are usually generic and the linked pages are not really associated with the query.
- **Look at the page on the Internet Archive**. Go to http://www.archive.org/index.php to enter the URL and view the page as it appeared previously, when its original owner maintained it. If the original site was different, it is probably a parked domain.

You will soon become familiar with the format of parked / expired domains.

Parked Domain Examples
Screenshot Examples

5.4 Pages with Unhelpful Content and PPC Ads

Some webpages with content are created just for the purpose of putting ads on them; writers are paid by spammers to create articles on a wide range of topics. Often the articles are very generic and do not provide a lot of good information, but they are original. You will not find the articles on another website. Although you may be convinced that the intent is to deceive, if the content makes sense and appears to be original, you will not be able to assign a **Spam** flag to such pages. You will have to use your judgment.

- Decide if you think the content is helpful to users or if it is too general, too poorly written, or gibberish.
- Try to determine if the page was made by a human or by a computer.
- Try to determine why the page was created.

Unhelpful Content Examples	
■ http://super-choice.blogspot.com/2005/06/super-calculator.html ■ http://www.impotence-erectile-dysfunction.com/viagra_drug_the_little_blue_pill.htm	Screenshot Examples

6.0 Phishing Websites

Phishing is an attempt by unscrupulous people to obtain sensitive information from Internet users. Some of you may have received emails in your own email accounts that look as if they're from legitimate companies, but upon closer inspection are not. Often these emails ask for sensitive information.

The landing page in the following task also asks for sensitive information and is another type of phishing.

Query [runescape gold], English (US)
URL http://www.gprunescape.com/

This landing page should make users (and raters) very suspicious and cautious. The spelling and grammar are bad and unprofessional, and the page feels "spammy". What is most worrisome is that the page asks for the user's bank password and pin number!

Even though we would not want to interact with the page, this type of phishing does not go against the Webspam Guidelines and the page should not be flagged as spam or malicious.

Please remember to only flag pages that fall in one of the spam categories described in the guidelines. Some phishing pages may be spam, but this one is not.

7.0 Spam and the Resolving Stage

It is not uncommon for tasks to go into the "resolving" stage because raters disagree on whether a page should be assigned **Unratable: Didn't Load** or a rating from the rating scale and a **Spam** flag. The disagreement occurs because raters see different pages when they click on the link in the task. These differences may be due to timing, or they may be due to Firefox browser version and/ or setting differences.

When a task goes into the resolving stage for this reason and the page you see matches the criteria for **Unratable: Didn't Load**, please take another look. Since other raters see a spam page, it is obvious that they are looking at something different from what you see. Here are some things you can try:

1. Update to the most current version of Firefox.
2. Look at the source code or disable JavaScript.

If you still do not detect spam, do not assign a **Spam** flag.

Please be aware that spam pages frequently stop loading after a period of time. If you detect spam one day, but the page does not load for you the next day, please do not change your rating, (i.e. do not remove the **Spam** flag).

You will learn more about the "resolving" stage in Part 6: Using EWOQ.

Spam recognition is a skill that is developed through practice and exposure. Open discussion of difficult cases in the resolving stage in EWOQ will help you develop your skills.

Remember to look at the page as a whole. Spam pages usually have some of these characteristics:

- PPC ads are usually very prominent on the page, and it is obvious that the page was created for them.
- If you do a text search, you will find that the content has been copied.
- If you visually remove all of the spam elements from the page (PPC ads and copied content), there is nothing of any value remaining.

Good pages usually have these characteristics:

- The page is well-organized. There may be ads on the page, but they are well identified and not distracting.
- If you do a text search, the original page is usually the first result displayed.
- The page will have value to the user. A good search engine would want the page in a set of search results.

Here are the spam flags that you will use:

- **Not Spam:** If you do not believe that a page is spam, you should assign a **Not Spam** flag.
- **Maybe Spam**: If you find a page to be "spammy", but you do not feel comfortable saying that the page is definitely spam, you should assign a **Maybe Spam** flag.
- **Spam**: If you believe that a page has been designed using the deceptive web design techniques described in these guidelines, you should assign a **Spam** flag.

When unsure which flag to use, remember to ask yourself these questions:

- Does the page provide the user with a good search experience?
- Does the page contain original content that would be helpful to users?
- Do you think the page should be included in a set of search results?
- Is the page designed for users? Is there a human element to the page?
- If you removed the PPC ads and copied text from the page, is there anything helpful left?

If you answer "yes" to these questions, the page is probably **not** spam.

Part 6: Using EWOQ

1.0   Introduction

Welcome to EWOQ !

EWOQ is the evaluation system you will use as a rater. You will acquire tasks and rate them based on the guidelines given to you.

For URL rating, a task consists of a pair: a **query** and a **URL**. As you work in the EWOQ interface, you will acquire tasks as you need them and submit your ratings as you complete them.

2.0   Accessing the EWOQ Rating Interface

Go this link to access the EWOQ URL rating interface: https://www.google.com/evaluation/search/rating/home

You will supply your Gmail user ID and password for authentication.

3.0   Rating

In general, rating a task involves the following steps:

1.   Acquiring tasks   (See the "Rating Home Before and After Task Acquisition" screenshots)
2.   Starting to rate   (See the "Rating Task Home" screenshot)
3.   Submitting your initial rating (See the "Rating Task Home" screenshot)
4.   Re-rating unresolved tasks   (See Section 5)
5.   Commenting   (See Section 6)
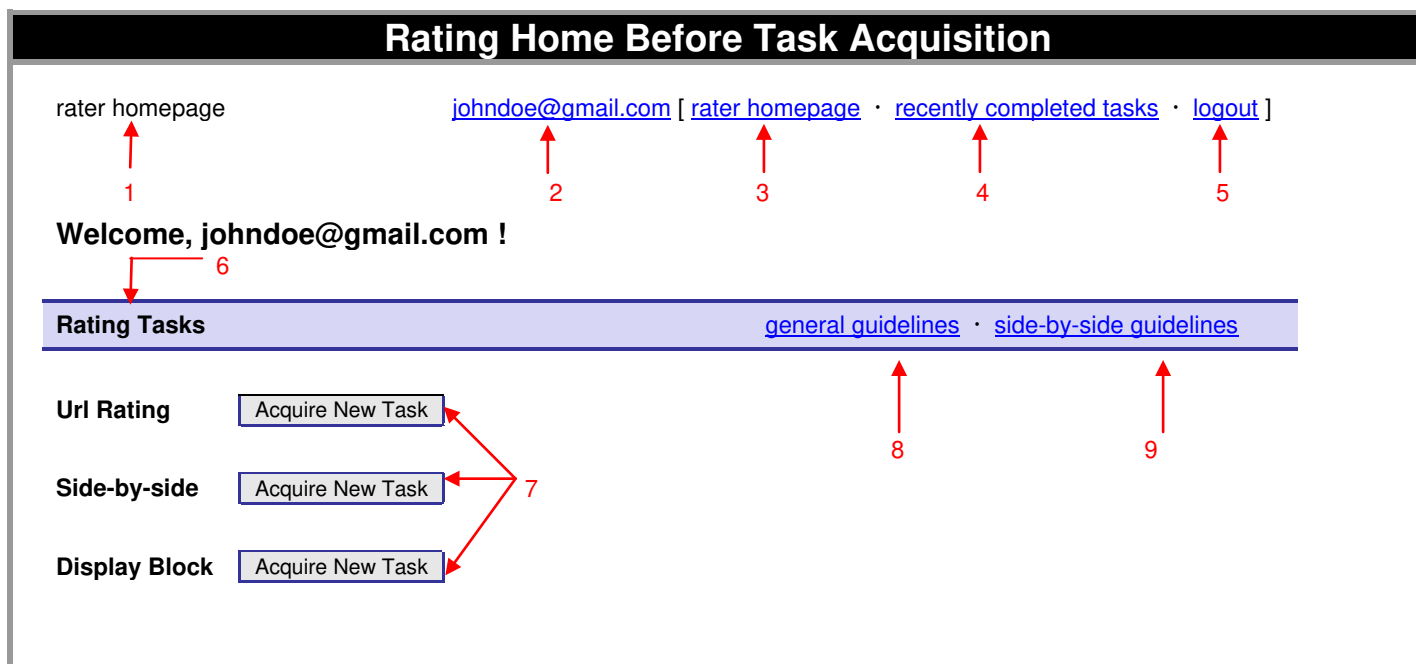
Rating Home Before Task Acquisition

rater homepage   johndoe@gmail.com [rater homepage · recently completed tasks · logout]

↑   ↑   ↑   ↑   ↑
1   2   3   4   5

Welcome, johndoe@gmail.com !

6

Rating Tasks   general guidelines · side-by-side guidelines

↑   ↑
8   9

Url Rating   [Acquire New Task]

Side-by-side   [Acquire New Task] ← 7

Display Block   [Acquire New Task]

The red numbers represent the following:

1. **rater homepage**
   This text shows that you are at the **Rater Homepage**.

2. **johndoe@gmail.com**
   Your Gmail account.

3. **rater homepage**
   Click on this link to go back to the **Rater Homepage**.

4. **recently completed tasks**
   Click on this link to change ratings on tasks completed in the last several minutes. Currently, the option to change ratings on recently completed tasks only applies to **Side-by-Side** and **URL Rating** tasks.

5. **logout**
   Click on this link to end your EWOQ session. Please logout to end your EWOQ session.

6. **Rating Task**
   This section lists available project types. The screenshot shows that tasks from "**Url Rating**", "**Side-by-Side**", and "**Display Block**" projects are currently available.

7. **Acquire New Task**
   Click this button to acquire a new task. The new Rater Homepage will allow you to acquire only one task from one of the project types displayed on your Rater Homepage. When tasks are available, you will see buttons for up to three different project types displayed. Please click on the button next to the project type you wish to work on. If there are no available tasks, you will see a "**No rating tasks**" message instead of the "**Acquire New Task**" button.

8. **general guidelines**
   Click on this link to read the "General Guidelines".

9. **side-by-side guidelines**
   Click on this link to read the "Side-by-Side Rating Guidelines".

Rating Home After Task Acquisition

rater homepage   johndoe@gmail.com [rater homepage · recently completed tasks · logout]

Welcome, johndoe@gmail.com !

Rating Tasks   general guidelines · side-by-side guidelines

You have a **URL Rating** task in your queue, please | continue | .

                12   11

Resolving Tasks

Resolving tasks in your queue:

Task ID	Status	Language	Query	URL	Last Modified	Expires	Rating
1234567	Unresolved	English (US)	hawaii	http://www.hawaii.gov	2/20/2008	2/20/2008	Off-Topic or Useless
7654321	**Unresolved**	**English (US)**	**sea turtle**	**http://www.turtle.com**	**2/21/2008**	**2/21/2008**	**Vital**

The red numbers represent the following:

10. **You have a "project type" task in your queue, please continue**
    The **continue** button indicates that you have an acquired but unrated task in your queue. In this example, the **"project type"** is **URL Rating**. Please click on the **continue** button to go to the **URL Rating Task Home** and rate the task.


11. **Resolving Tasks**
    Every task will be acquired and rated by a group of raters, each working independently. If raters disagree with one another by a wide margin, the task will be returned to the raters involved for re-rating in the "resolving stage". This resolving section will appear on your Rater Homepage only if there are task(s) that need to be resolved.   Please participate in the resolving process as soon as possible.

Rating Task Home

---

1   2   3   4   5   6

Rating Task - icq

7 → [search results: google]   [release task] ← 8   general guidelines   9

10 →	**Query**	Icq
11 →	**Query Description**	*This field is present only if there is a description for the query.*
12 →	**URL**	http://www.mobicq.info/
13 →	**Task Location**	Ukraine (UA)
14 →	**Task Language**	Ukrainian
15 →	**Other Acceptable Languages**	Russian

URL RATING

16 →	**Rating**
	Choose one
		○ **Vital** (choose one geographical location) ← 17
		  ○ **Appropriate Vital**
		  ○ **International Vital**
		  ○ **Other Vital**
		○ **Useful**
		○ **Relevant**
		○ **Slightly Relevant**
		○ **Off-Topic or Useless**
		○ **Unratable** ← 18
		  ☐ Didn't Load
		  ☐ Foreign Language

19 →	**Landing Page Language** Choose one	○ Ukrainian ○ Russian ○ English ○ Foreign Language ○ None of the above
20 →	**Spam** Choose one	○ Not Spam ○ Maybe Spam ○ Spam
21 →	**Other Flags** Choose all that apply	☐ Pornography ☐ Malicious
22 →	**Comment**	

[Cancel]   [Submit]

23   24

The red numbers represent the following:

1. **rater homepage**
   This text shows that you are at the **Rater Homepage**.

2. **rater homepage → rating task**
   This shows your location in the EWOQ system; in our screenshot, the display shows the path from the **rater homepage** to the current **Rating Task** page.

3. **johndoe@gmail.com**
   Your Gmail account.

4. **rater homepage**
   Click on this link to go to the **Rater Homepage**.

5. **recently completed tasks**
   Click on this link to change ratings on tasks completed in the last several minutes. Currently, the option to change ratings on recently completed tasks only applies to **Side-by-Side** and **URL Rating** tasks.

6. **logout**
   Click on this link to end your EWOQ session. Please logout to end your EWOQ session.

7. **search results**
   Clicking these links automatically displays search results for the query.

8. **release task**
   Clicking on this link allows you to remove the task from your task list. To ensure you indeed mean to give up a task, a dialogue box will appear before the task is released. This is what releasing the task accomplishes:

   a. The released task will not be considered part of your workflow.
   b. The task will return to the pool of tasks, to be reassigned to other raters via a randomized process based on availability and priority. The task will not come back to you.

Option	Use this option when:	Can the task (same query and URL pair) come back ?
"release task" button	You personally cannot rate the query, but you think other raters will be able to rate it. For example the query is technical or scientific, and you believe that other raters may do a better job than you evaluating landing pages for the query.	No

9. **general guidelines**
   Click on this link to view the "General Guidelines".

10. **Query**
   Make sure you understand the query. Please research the query to learn about its meaning and the user intent behind it.

11. **Query Description**
   <mark>This field is present when there is a description for the query. Raters are expected to look for query descriptions and pay attention to them when assigning ratings.</mark>

12. **URL**
   This is the URL that you will click to view the landing page.

13. **Task Location**
   The location associated with the task.

14. **Task Language**
   The language associated with the task.

15. **Other Acceptable Languages**
   Please refer to the "Rating Guidelines" for information on acceptable languages.

16. **Rating**
   Please refer to the "Rating Guidelines" for information on each rating category.

17. **Vital**
   If the page is **Vital**, please choose one of the three geographical location **Vital** ratings. Please note that clicking on one of the three buttons will simultaneously select the **Vital** button.

18. **Unratable**
   If the page is **Unratable**, please choose any checkboxes that represent your reason(s) for selecting **Unratable**. Please note that:
   - Clicking on one of the two checkboxes will simultaneously select the **Unratable** button.
   - Clicking on the **Foreign Language** checkbox will simultaneously select the **Foreign Language** button in the **Landing Page Language** section.

19. **Landing Page Language**
   Please refer to the "Rating Guidelines" for information on selecting the landing page language.

20. **Spam**
   Assign one of the three spam flags to pages that load and can be rated. Spam flags are optional when you select either of the **Unratable** options. If you notice that an **Unratable: Didn't Load** or **Unratable: Foreign Language** page is spam, please assign a **Spam** flag. Please note that you are required to leave a comment if you choose **Spam** or **Maybe Spam**.

21. **Other Flags**
   Please choose **Pornography** and/or flags when appropriate.

22. Comment

New raters are REQUIRED to comment on every URL task in the initial rating stage for the first three weeks. After that, commenting is required only when you assign **Spam**, **Maybe Spam**, and/or **Malicious** flags. Please note that you will not be notified when the three week mandatory commenting period is over, and that you will not need to comment on every task after the first three weeks.

Exam takers: Please note that the commenting requirement applies to the first three weeks of employment after raters are hired. It does not apply to exam takers. While taking the exam, you do not need to leave any comments. Your exam will be graded only on the answers you select.

23. Cancel

You may select "Cancel" to retain a task without saving any information. Choosing this option will take you back to the Rater Homepage with a message "You have a **url rating** task in your queue, please continue ."

24. Submit

You will submit your rating to finalize your work on a task.

5.0   Resolving Tasks (Re-rating Unresolved Tasks) / Moderators

Every task will be acquired and rated by a group of raters, each working independently. If the raters disagree with one another by a wide margin, the task will be returned to the raters involved for re-rating in the "resolving" stage. It will reappear in your task list on the Rater Homepage with the status "Unresolved" and will be highlighted in yellow to catch your attention.

In addition, each time an action has been taken on the "Unresolved" task by someone other than you, the task will remain highlighted, but will also be shown in **bold** text. The actions that will cause this to happen are rating changes made by other raters and/or commenting by raters, administrators, or moderators. This is analogous to how unviewed messages appear in bold text in an e-mail inbox.

When you see that a task has entered the "Unresolved" state, **or that a previously resolved task appears again in bold text**, you are required to revisit the task to participate in the resolving process. In other words, even though you and the other raters have come to agreement on a task, the resolving process may not be over. A rater, moderator, or administrator might have something important to communicate and may have added a comment even though the task is in the "Resolved" state. **Anytime a task appears in bold text, please revisit the task.**

Moderators

For some unresolved tasks, you may see comments written by a moderator. Please pay attention to these comments just as you would comments from an administrator. The moderator helps resolve tasks and contributes to discussions by:
- monitoring tasks
- highlighting rater comments
- leaving comments and helpful tips

Rating Task Home

Rating Task - icq

[search results: google] ·   release task   · general guidelines

Query	icq
URL	http://www.b-mobil-pho-cheap-get-free-great-deals.com /
Task Location	Ukraine (UA)
Task Language	Ukrainian
Other Acceptable Languages	Russian

Related Ratings

1 →

Rater	Last Modified	Rating	Spam	Flags
Rater 2	3/14/08 10:36 AM	Slightly Relevant	Maybe Spam	
Rater 3	3/12/08 9:02 AM	Off-Topic or Useless	Spam	Pornography, Malicious
Rater 4	3/14/08 7:55 AM	Unratable: Didn't Load	*None*	
me (Rater 1)	3/15/08 10:38 AM	Off-Topic or Useless	Spam	Pornography
Rater 5	3/14/08 6:36 PM	Relevant	Not Spam	

2 → (points to "me (Rater 1)" row)

Comments on this Rating

3 →

Comment	Rater	Timestamp
Article not found message, therefore DL.	Rater 4	3/14/08 7:55 AM
There is pornographic hidden text and links. Attempted to download spyware.	Rater 3	3/12/08 9:02 AM
Confirming that there are hidden text and links to pornographic sites.	Rater 1	3/15/08 10:38AM

The red numbers represent the following:

1. **Related Ratings**
   This section shows the ratings submitted by other raters with a "Last Modified" timestamp. Everyone participating in a task will stay anonymous. In fact, all raters are identified by "Rater" plus a number. Administrators will be shown as Administrator instead of Rater. Moderators will be shown as Moderator plus a number.

2. **Me (Rater 1)**
   You will be able to see your initial rating with its timestamp. In this example, the rater is identified as Rater 1.

3. **Comments on this Rating**
   This section displays all comments left in the task, including your initial comments, *if any*. As you and other participants enter more comments in the future, the comments will be posted in this box. The most recent comments will appear on the bottom of the page.

153

Example 1: User / Moderator

Comment	Rater	Timestamp
Appropriate Vital – www.wine.com	Rater 3	3/14/08 7:55 AM
Can generic subjects have Vital results ?	Moderator	3/14/08 8:03 AM

Example 2: Users / Administrator

Comment	Rater	Timestamp
There is hidden text on this page	Rater 1	3/14/08 7:06 AM
Indeed hidden text down the bottom	Administrator	3/14/08 1:02 PM
Landing page DL --- User 2 8/20/06 1:07 PM	Rater 2	3/15/08 6:28 PM

Example 3: Users / Moderator / Administrator

Comment		
Sneaky redirect to www.sdasdfasde-asdf-zzzz.com	Rater 3	3/15/08 6:38 AM
Landing page DL --- User 3 at 8/20/06 7:00 PM	Rater 2	3/15/08 8:08 AM
Please refer to guidelines for more information on spam and resolve disagreements as soon as possible.	Moderator	3/15/08 1:35 PM
Also check to see if there is any hidden text	Administrator	3/15/08 8:30 PM
Sneaky redirect, keyword stuffing and hidden text. Changing from DL to OT/Spam	Rater 1	3/16/08 1:26 AM

6.0 Commenting Etiquette

The following are guidelines for effective communication during the resolving process in EWOQ.

1. It is important to share relevant background information (reasons, explanations, etc.) when stating your opinion. Indicate your source of information whenever possible. If you come across an important website in your research, please give its full URL.

2. Please do not use abbreviations.
   Exception: To save space and time, the following abbreviations for ratings and flags should be used:

V	(Vital)	OT	(Off-Topic or Useless)
AV	(Appropriate Vital)	DL	(Unratable: Didn't Load)
IV	(International Vital)	FL	(Unratable: Foreign Language)
OV	(Other Vital)	Mal	(Malicious)
Usf	(Useful)	PPC	(pay-per-click)
Rel	(Relevant)	LP	(landing page)
SR	(Slightly Relevant)		

Please refrain from using message board lingo (IMO, FWIW, AFAIK, etc.).

3. Please write concisely. Do not make unnecessary comments such as "Oh, I see your point" or "Sorry, I missed that". But do write enough to explain yourself clearly to other raters who might not have your background or expertise.

4. Please do not type your comments in all capital letters. The use of all capitals is generally considered shouting and may bother other raters.

5. Sometimes the most efficient way to make your point is to quote guidelines. Please be very specific about how the information you quote relates to the situation at hand. When quoting from the "General Guidelines", please include the version number and page number.

6. When commenting on a query, describe your interpretation of user intent. This is very important for ambiguous or poorly phrased queries. You may include whether you believe the query is a navigation, information, or action query. If you disagree with the Query Description you see on the EWOQ interface, please be explicit about that as well.

7. State your reason for assigning "Spam", "Maybe Spam", and "Malicious" flags.

   Spam and **Maybe Spam** flag comment examples:
   - Hidden text
   - Keyword stuffing
   - Sneaky redirect to eBay
   - Sneaky redirect to << enter URL of page redirected to >>
   - JavaScript redirect
   - 100% frame
   - Copied text from Wikipedia plus ads
   - DMOZ content plus ads
   - News feed plus ads
   - Templated spam page
   - Computer-generated gibberish
   - Copied message board
   - Fake search page
   - Fake blog
   - Fake message board
   - Amazon thin affiliate
   - PPC only
   - Parked domain

   Malicious flag comment examples:
   - Pop-ups would not go away
   - Page forced me to close Firefox to continue working
   - Page downloaded Trojan on my computer
   - My anti-virus software detected a virus

8. Brief comments to confirm your rating in the resolving stage are always appreciated:
   - "Still DL for me."
   - "Confirming Usf: it's the best result I could find."

Part 7: Quick Guide to URL Rating

Welcome to URL Rating

The "Quick Guide to URL Rating" is an abbreviated version of the "Rating Guidelines".

IMPORTANT DEFINITIONS:

Search Engine: A website that lets users search the Web by typing words, numbers, and/or symbols into a search box.
Query: The words, numbers, and/or symbols user types in the search box of a search engine.
Task Language and Task Location: Every query has a task language and task location associated with it using this format: [digital cameras], Spanish (MX), which indicates that a Spanish reading user in Mexico typed "digital cameras" in the search box. As a rater, you will represent users in your task location who read the task language.
Homepage: The main page of a website, for example: http://www.apple.com.
Subpage: A page on a website that is not the homepage.
Webpage: Any page on a website: a homepage or subpage.
URL: The web address of the page you will evaluate.
Page or Landing Page: The page you will evaluate. It is the page you see after you click on the URL. You must visit the landing page on every URL rating task.
User Intent: What the user is trying to accomplish by typing the query.
Topic: What the query is about.
Utility: A measure of how helpful the page is for the user intent. Pages with good utility are helpful for users.

Internet Safety Information: We strongly recommend that you have anti-virus and anti-spyware protection on your computer that you update regularly. We suggest that you only open files with which you are comfortable. File formats are generally considered safe: .txt, .ppt, .doc, .xls, and .pdf.

Understanding the Query: Before evaluating a task, you must understand the query. Use an online encyclopedia (such as http://www.wikipedia.org) and/or do web research. Keep in mind, however, that pages helpful to you may not be helpful to users (who already understand the query). All web research must be done using the Firefox browser.

Understanding User Intent: You also need to understand user intent to evaluate a page. When a user types [tetris], English (US), the likely user intent is to play the game online. A page that allows users to play the game fits the user intent. A page about the history of the game does not.

Issues to Consider

Task Language and Task Location: Users in different parts of the world have different expectations for the same query. English (US) and English (UK) users will have different interpretations for the query [football].

Queries with Multiple Meanings: Many queries have more than one meaning. The query [apple], English (US) could refer to the computer brand or the fruit. We call these possible meanings "query interpretations".

Dominant Interpretation: The one query interpretation that most users have in mind. The Microsoft operating system is the dominant interpretation for [windows], English (US).

Common Interpretations: Sometimes, there is no dominant interpretation. The car, the planet, and the chemical are common interpretations for [mercury], English (US).

Minor Interpretations: Sometimes you will find less common interpretations. Mercury Marine Insurance Company is a minor interpretation for [mercury], English (US).

Timeliness: A query can be interpreted differently at different points in time. In 1994, the user who typed [President Bush], English (US) was looking for information on President George H.W. Bush. In 2010, his son George W. Bush is the more likely interpretation.

Classification of User Intent: Do-Know-Go: It is helpful to classify the query according to user intent. Note: Many queries have more than one type of user intent.

Action Intent (Do): The user wants to accomplish a goal or engage in an activity, such as make a purchase, download software, play a game, print a calendar, send flowers, watch a video, copy an image, etc.

Information Intent (Know): The user wants to find information.

Navigation Intent (Go): The user wants go to a specific website or webpage, such as the IBM homepage or the Camry page on the Toyota website.

The Language of the Landing Page: You will look at the landing page and determine which of the following best describes the language on it:

Task Language: The page is in the task language.
Acceptable Languages: The page is in another language that is commonly used in the task location.
English: The page is in English.
Foreign Language: The page is in a language other than the task language, an acceptable language, or English.
None of the above: The page has no language or does not load in a way that the language can be evaluated.

Please use your judgment when there is more than one language on the landing page.

The Rating Scale

The Rating Scale rating options are: Vital, Useful, Relevant, Slightly Relevant, Off-Topic or Useless, and Unratable.

Vital (V) is used for these very special situations:
- The dominant interpretation of the query is navigation and the page is the target of the navigation query, e.g. [yahoo], English (US) and http://www.yahoo.com.

- The dominant interpretation of the query is an entity (such as a person, place, business, restaurant, product, company, organization, etc.) and the page is the official page associated with that entity, e.g. [ipod nano], English (US) and http://www.apple.com/ipodnano/.

ENTITY QUERIES WITH VITAL PAGES

Some entity queries are **Go** queries, while others are **Know** queries. For entity queries, the official page of the entity is **Vital**, even if you think the user wants information. Examples of entity types: celebrities, restaurants, movies, companies, books, specific products, famous locations, special events, government officials, blogs, universities, etc.

VITAL PAGES FOR PEOPLE QUERIES:

For a query which is the name of a real (non-fictional) living person, the **Vital** rating should be used when:
- The query has a clear dominant interpretation, i.e. most people issuing the query are looking for information about one particular individual.
- The result is the homepage of the person's official website, if such a website exists.

VITAL PAGES AND GEOGRAPHIC LOCATION: We have 3 different **Vital** ratings because some official sites or pages have multiple versions for different languages or countries.

Appropriate Vital (AV): Use **AV** if (1) there is only one version of the page, (2) there is more than one version, and the page seems right for the task location, or (3) if the page is the one "asked for" in the query.

International Vital (IV): Use **IV** if (1) the page is a "choose your language" or "choose your location" page, or (2) for an English version which is designed to be an international page, helpful to many users.

Other Vital (OV): Use **OV** if the language or location of the official page does not match the task location, and a better version exists. (If a better version for the task location does not exist, then use **Appropriate Vital**).

Important Vital Concepts:
- The query must have a dominant interpretation. If there is no dominant interpretation, no **Vital** rating is possible.
- Most **Vital** pages have very high or the highest possible utility, but some **Vital** pages do not.
- Information queries usually do not have **Vital** pages.
- Some URLs that "look" **Vital** are not. www.diabetes.com cannot be **Vital** for [diabetes], English (US) because this is an information query and no one can own it.
- A query can have more than one **Vital** page. For the query [barnes and noble], English (US), www.books.com www.bn.com, and www.barnesandnoble.com all have the same landing page and are all **Vital** for the query.

Useful (Usf) pages are very helpful for most users. They should be (1) high quality, and (2) a good "fit" for the query. They often have some or all of these characteristics: comprehensive, highly satisfying, authoritative, well-organized, entertaining and/or recent (such as breaking news

on a topic. Spammy pages should not be rated **Useful.** Note that more than one page can be rated **Useful** for a query.

Relevant (Rel) pages are helpful for many or some users. They should still "fit" the query, but might have fewer valuable attributes than were listed for **Useful** pages. **Relevant** pages may be less comprehensive, less satisfying, come from a less authoritative source, etc. They should not be low quality.

Slightly Relevant (SR) pages are generally not helpful, but are still marginally on-topic. They may be low quality, outdated, too narrowly regional, too specific, too broad, or service a minor interpretation, etc. They may have less information and come from a less authoritative source. **Slightly Relevant** is also appropriate for superficially relevant or shallow pages.

Off-Topic or Useless (OT) pages are not helpful for most users. They are unrelated to the query and/or have no utility.

Unratable: Pages that you are unable to evaluate are **Unratable**. There are two **Unratable** categories: **Didn't Load** and **Foreign Language**.

Unratable: Didn't Load (DL): This is a special rating category for pages that truly do not load or have any content at all. Assign this rating to:
- Pages with *error messages and no other content.*
- Pages with non-working redirects and *no other content.*
- Completely blank pages.
- Pages with malware warnings, such as "Warning-visiting this web site may harm your computer."

Unratable: Foreign Language (FL): Assign this rating when the landing page is not the task language, an acceptable language, or English:
- And the landing page is not clearly **Vital** for the query, based on the appearance of the URL of the landing page.
- Even if you can tell that the page is off-topic.

From User Intent to Assigning a Rating

Location is Important – Sometimes you will need to lower the rating if the page content is from another country.

Language is Important – Landing pages in the task language are clearly good. Landing pages in English or an acceptable language may not be a good "fit" for users in the task location.

Multiple Interpretations – Pages associated with minor interpretations and unlikely user intents should be rated lower. Pages for common interpretations and reasonable user intents should not be rated lower. Only queries with a dominant interpretation can have **Vital** pages.

Specificity of Queries and Landing Pages – Some queries are general, some are specific, and some are in between. Good landing pages need to "fit" the specificity of the query to be helpful to users. When there is a mismatch between the query and the landing page, think about how helpful the page would be for users.

Common Rating Problems

There are some situations in which it is difficult for raters to assign good ratings. This is often because the experience of the rater is very different from the experience of the user. You do not write the queries you rate, and you cannot be sure what the user really wants. Also, you rate one result at a time without the context of a search engine result page, whereas the user is able to see the full page of search results. Here are some hard rating situations:

Dictionary or Encyclopedia Results - These types of pages are often helpful to raters who are trying to understand the query. They can also sometimes be helpful for the user, but not when the user already understands the words in the query, and is looking for something different.

Queries That Ask for a List - When the query seems to ask for a list that includes many, many possibilities, individual examples usually are not as helpful as a list. When the list of possibilities is short, then individual examples are helpful. Sometimes, there are very famous or popular examples on the list. In these cases, the individual famous or popular examples are helpful, even if the list of possibilities is long.

Misspelled and Mistyped Queries – For **obviously** misspelled or mistyped queries, you should base your rating on user intent, not necessarily on exactly how the query has been spelled. For queries that are **not** obviously misspelled, you should assume users are looking for results for the query as it is spelled. [federal expres] is obviously misspelled. [micheal Jordon] is not obviously misspelled.

URL QUERIES - These are "go" queries that are URLs or look like parts of URLs.
Working URL queries -[www.ebay.ca], [mail.yahoo.com], [http://www.amazon.com], [rei.com].
Non-working or "Imperfect" URL Queries - [ebay.cxom], [us open tennis tournament.org], [www.pizzzzahut.com]

Website Name/Webpage Name Queries - [ebay], [amazon], [yahoo mail]. These queries contain the names of websites or webpages, and the dominant interpretation of the query is the website or webpage. Some website name queries have other meanings, besides the website. For example, [kayak].

Generic Queries – [couches], [diabetes], [quilting]. These are **not** URL queries and they are **not** website name queries. Websites exist that match these queries, but those websites are probably not what users have in mind.

New and Old Pages – The landing page should be rated based on "fit" to the informational need of the query. Some queries demand very recent results, but not all. Most of the time, you need to consider the content of the page rather than the date on the page.

Search Engine Result Pages – Search engine result pages should be rated just like other landing pages: rate the landing page on the basis of how helpful it is for users.

- If the landing page you are given to rate is a search engine page with an empty search box and no results

displayed, then the page has no connection to the query and should get a rating of **Off-Topic or Useless**.
- If the landing page is a set of results from a search engine, the page could be very helpful to users. Depending on how helpful the page would be, ratings can range from **Useful** to **Off-Topic or Useless**. The landing page could be a web search results page, a shopping search results page, a video search results page, an image search results page, etc.

Video Landing Pages – If a query "asks" for a foreign language song, band, film, sporting event, etc., then a video of the song, band, film, sporting is helpful and should not be rated **FL**. If the video is someone talking *about* the song, band, film, or event, it probably cannot be understood and should be rated **FL**.

Flags

Not Spam: Assign this flag if you do not believe deceptive web design techniques were used.
Maybe Spam: Assign this flag if you find a page to be "spammy", but not spam.
Spam: Assign this flag if you believe that the page was designed using deceptive techniques.

Pornography – Assign the **Porn** flag to all porn pages. A page is porn if it has porn content, including porn images, links, text, pop-ups, and/or ads. Please consider user intent when evaluating porn pages:
- **Clear Non-Porn Intent**: If user intent is clearly **not** pornographic, a landing page that serves a porn interpretation of the query and/or has porn for its main content should be rated **Off-Topic or Useless** and assigned a **Porn** flag.
- **Possible Porn Intent**: Some queries have both non-porn and porn interpretations. For example, [girls], English (US) is a "possible porn intent" query: it has both porn and non-porn interpretations. For these queries, please assume that the non-porn interpretation is dominant, even if you think the user is looking for porn. Rate the porn interpretation as a minor interpretation and assign a **Porn** flag.
- **Clear Porn Intent**: For very clear porn queries, where no other intent is possible, assign a rating to the porn landing page using the rating scale without lowering the score. Even though there is porn intent, assign a **Porn** flag. However, please do not assign a **Porn** flag just because the query has porn intent.

Please note that porn stars, porn websites, etc. can have **Vital** pages. Remember to also assign a **Porn** flag.

Malicious: Please assign this flag if:
- You are forced to quit your Firefox browser due to prompts that keep coming back and will not go away.
- There are attempts to download spyware, Trojans, viruses, etc.
Please note that pop-ups that do *not* come back are *not* malicious.

Compatibility between Ratings and Flags: Please be aware that **Unratable** pages can be assigned **Spam, Porn,** and/or **Malicious** flags.

Part 8: Quick Guide to Webspam Recognition

What is Webspam?

Webspam is the term for webpages that are designed by webmasters to trick search engines and direct traffic to their websites. We sometimes refer to webmasters who use deceptive techniques as "spammers".

General Information

- Assign a **Spam** flag if the page uses deceptive techniques, even if it has utility for the user intent.
- Pay-Per-Click (PPC) ads appear on many pages on the Web. Spammers make money when the ads are clicked. **Many pages with PPC ads are NOT spam.**
- Sometimes, spam pages do not have moneymaking links. They are created to change search engine rankings or even do harm to users' computers. They are spam because they use deceptive techniques, even though you cannot see how spammers are making money.
- Do *not* assign a **Spam** flag to a page that is merely annoying, junky, or low quality, such as pages with lots of pop-ups and ads.

Browser Requirement

- Unless told otherwise in the project-specific instructions, you must do ALL of your rating work (including query research) in Firefox. You must not use any other browser for your rating work.
- Mozilla offers a Firefox Add-on called "Web Developer", which provides a special toolbar containing tools helpful in spam detection.

Technical Signals

When evaluating a page for spam, look for these technical signals: hidden text and hidden links: keyword stuffing, sneaky redirects, and cloaking with JavaScript and CSS.

Hidden Text and Hidden Links: Spammers add hidden text and/or hidden links to lure search engines and users to their pages. Hidden text is visible to the search engine, but not to the user who may find it distracting or annoying. Hidden text may be: invisible, in a font color that blends in, in a very tiny font size, or it may be placed on a portion of the page outside the normal viewing area.

Here are techniques for revealing hidden text. Please use the first two techniques on all webpages, since these are quick and easy to do. Please use the other techniques when you are suspicious that the page may be spam.

Apply Ctrl-A: Ctrl-A is the keyboard shortcut for "Select All" for PC users. Hitting the "Ctrl" and "A" keys simultaneously selects all the text on the page and may display hidden text.

Apple computer users will use "⌘" and "A".

Look outside the normal viewing area: Be suspicious of large blank areas on the bottom and far right portions of the page, and scroll through those areas to look for hidden text on those parts of the page.
Disable CSS: Use the Web Developer toolbar to disable CSS and look for hidden text.

Disable JavaScript: Use the Web Developer toolbar or your Firefox browser menu to disable JavaScript. Here are the instructions for disabling JavaScript using your browser menu, in case you do not wish to use Web Developer.
Disabling JavaScript in Firefox:
1. Go to "Tools".
2. Click on "Options".
3. Click on "Content" or "Web Features".
4. To disable JavaScript, make sure the "Enable" box is not checked.
5. Click "OK".

View the Source Code: Another way to reveal hidden text is by looking at the source code of the page. You can use the Web Developer toolbar or your browser toolbar to view the source code. Compare the source code to what you see on page. Sometimes you will see large sections of keyword stuffing in the source code that do not appear on the page. Note: keyword stuffing in the meta tags is not spam.

Keyword Stuffing: Webmasters sometimes load pages with keywords, which may be related or unrelated to the content on the page. Assign a **Spam** flag if you think the number of keywords on the page is excessive and would be annoying to users. Hidden text and keyword stuffing often go together. Hidden text frequently contains keyword stuffing.

Keyword stuffing in the URL: URLs may also contain keyword stuffing. The URLs are computer-generated and have hyphens (dashes) separating the keywords.

Please note: Hidden text is not spam if there is no intention to trick the search engine. If the webmaster "hides" the date of an update, that would not be considered spam.

Sneaky Redirects: We call it a sneaky redirect when a page redirects the user from a URL on one domain to a different URL on a different domain, *with spam intent*

Please note: Not all redirects are sneaky. Redirects to a different page on the same domain are not sneaky. Also, a site might legitimately redirect from one URL to another. After the merger of Compaq and Hewlett-Packard, the Compaq URL automatically redirects to the HP site.

Checking "Who Is" the Domain Owner: When you suspect a page is a sneaky redirect, it is a good idea to check "who is" the owner of the two domains to see if there is a relationship between them. You will do this by going to a "whois" provider to find out "who is" the domain registrant. You will type in the domain names and look at the information provided for each. If you find that the two URLs have the same domain registrant, you will conclude that the page is not spam.

Here are two you can use:
http://www.domaintools.com/
http://whois.mtgsy.net/default.php.

Cloaking: We call it cloaking when the webmaster shows different pages to the search engine and the user. Two cloaking techniques used by spammers are JavaScript redirects and 100% frame.

JavaScript Redirects: Spammers use JavaScript redirects to create two different pages. Looking at the page first with JavaScript enabled and then with JavaScript disabled reveals the differences.

100% Frame: Webmasters sometimes cloak what users see by using frames. Two frames (pages) exist, but one frame takes up 100% of the screen. The user sees one frame (page), but the search engine sees both frames.

To look for 100% frame in Firefox, right-click on the page, click "This Frame", and then click "View Frame Info". Compare the URL of the landing page with the URL of the frame. If they are different, you will usually assign a **Spam** flag. It is also sometimes helpful to use "who is" to look at the domain registrants of the pages.

Helpful Webpages vs. Spam Webpages

Search engines want to display webpages that are helpful to users. Some pages with PPC ads are designed to be helpful to users in some way. These pages are not spam. Pages with PPC ads that exist primarily to make money or change search engine rankings are spam.

The following types of pages have content that is helpful to users.
- Pages that allow users to compare prices between merchants are not spam.
- Pages that have *original* product reviews that are helpful to users are not spam.
- Pages with *original* recipes or reviews of non-original recipes are not spam.
- Pages from websites that are designed to help users find lyrics, quotes, proverbs, poems, etc. are not spam.
- Contact information: Pages with physical addresses, phone numbers, maps, etc. are not spam.
- Pages with coupon, discount, and promotion codes that are helpful to users are not spam.

Pages with Copied Content and PPC Ads: Copied content is content copied from another source. Webmasters sometimes use special software to search the Web for content to put on their websites that is related to specific keywords. Content can also be taken from another website using the simple "copy and paste" method.

Copied Text and PPC Ads: Text is often copied from sources like Wikipedia and the Open Directory Project (DMOZ). Even if the webmaster gives credit to Wikipedia for the content, it is considered to be spam.

Feeds and PPC Ads: If a page has a freely available feed (such as a news feed available through RSS or XML) and PPC ads, and is created just to make money, it is spam.

Doorway Pages: Multiple doorway pages, which are created to send users to a common moneymaking page, do not provide meaningful content and are spam.

Templates and Other Computer-Generated Pages: Some websites use templates to mass-reproduce webpages automatically. The content is copied and the pages follow a generic format or pattern. Clicking on links on these pages will usually land you on other pages on the same domain with similar content and links. These pages are spam.

Copied Message Boards: Sometimes you will see copied message boards (user forums) are PPC ads. These pages are spam.

Here are some things you can do that will help you to recognize copied content:
- Search for an exact sentence in the text. Copy and paste a distinctive sentence or piece of text in the search box of a search engine. Put quotation marks around the piece of text. From the search results, you may find where the content originated. If it is original and not copied from another source, it probably was written to be helpful for users.
- Look for PPC ads surrounding the content. Wikipedia and DMOZ do not display ads.
- Become familiar with the format of Wikipedia and DMOZ pages, so you can recognize when their content has been copied.
- Look for suspicious, computer-generated grammar. When it is computer-generated, it often looks like "gibberish". You may also see hyperlinked keywords inside the text.
- Look for URL formatting that suggests that a template was used to create it. Often the URL will display keywords separated by hyphens.
- Try to figure out if the page was created to help users.
- Try to figure out if the page was created by a human or by a machine. Pages created by machines are usually not designed to be helpful and are usually spam.

Fake Search Pages with PPC Ads: A fake search page is a page with a list of links that looks like a page of search results. If you click on a few of the links, you see that the page is just a collection of PPC links disguised as a page of search engine results. Fake search pages sometimes look like parked domains.

Fake Blogs and Fake Message Boards with PPC Ads: Fake blogs and fake message boards have the appearance of real pages, but contain "entries" and "messages" that are nonsensical or copied from another source.

Please note that real, legitimate message boards are sometimes "spammed", which means that someone comes along and puts up posts with PPC ads and/or porn links. We do not assign a **Spam** flag to spammed message boards.

Commercial Intent

Most spam pages have commercial intent. Spammers create pages to make money. If a page exists primarily to make money without sufficient added value for users, the page is spam.

Reminder: Some spam pages do not have obvious moneymaking intent. They are created to change search engine rankings or to do harm to users' computers. They are spam because they use deceptive techniques, even though you cannot see how they are making money.

Thin Affiliates: A thin affiliate is a website that earns money from affiliate commissions. Spammers make money when a transaction is completed after the user has clicked through to the merchant's site from their webpages. A thin affiliate typically doesn't add much value compared to many other sources of information on the Web, and often sends the user to another website to complete the actual purchase.

Here are some things you can do to help you determine if a page is a thin affiliate:
- Click buttons on the page, such as a "make a purchase" button. If you are taken to a merchant on a different domain, it is probably a thin affiliate.
- Check the "properties" of images on the page. Right-click on an image and look at "Properties" to see where the image originates. Check to see if the address of the image is the same as the address of the page, or if it is the address of a "real" merchant.
- Look for original content on the page. The quality of an affiliate page or site depends on how much added value, usefulness, or original/additional information is available on the page that is not easily available elsewhere on the Web. If the page has the same "cookie-cutter" text or functionality found on dozens or hundreds of other sites, it is more likely to be spam.
- Use "who is" to look at the domain registrants of the two pages to see if they are the same or different.

Not all affiliates are thin: Some affiliates are created to help users. Anyone can become an "affiliate" of a merchant's site such as Amazon and link to Amazon products. Webmasters may do this to show products they like or to help users find good deals. For example, if the affiliate offers price comparisons, or displays product reviews, recipes, lyrics, etc., it is usually not a thin affiliate. Some websites that offer price comparisons or other helpful shopping features, in addition to the affiliate link, are:

- http://www.shopping.com
- http://www.pricegrabber.com
- http://www.kelkoo.co.uk

Recognizing true merchants: Features that will help you determine if a website is a true merchant include:
- A "view your shopping cart" link that stays on the same website
- A shopping cart that updates when you add items to it
- A return policy with a physical address
- A shipping charge calculator that works
- A "wish list" link, or a link to postpone the purchase of an item until later
- A way to track FedEx orders
- A user forum that works
- The ability to register or login
- A gift registry that works

Please note the following:

- A page does not need to have all of these to be considered a true merchant.
- Yahoo! Stores are true merchants.
- Some true smaller merchants take users to another site to complete the transaction because they use a third party to process the transaction. These merchants are not thin affiliates.

Pure PPC Pages: We refer to pages with PPC ads only (or with PPC ads and very little other content on them) as pure PPC pages. Spammers make money when a link is clicked; no purchase is necessary. Pure PPC pages are spam.

Parked (Expired) Domains
The word "domain" can have two different meanings for raters:
1) "Domain" can refer to the elements in the DNS (Domain Name System), such as .com, org, .uk, .cn, etc. that organize Internet addresses
2) "Domain" can refer to the set of words (URL) that identifies the web address of a specific entity, such as "microsoft.com" or "baidu.cn".

When companies go out of business, are acquired, change their name, or fail to pay their domain registration fee, the domain name "expires" and may be purchased by someone else. Spammers sometimes buy expired or expiring domains and put their own content on the page. Spammers also purchase domains that are similar in spelling to real domains, hoping that users will mistype the domain name or URL and land on their website, which contains PPC ads. All of these types of pages are referred to as parked domains.

A typical parked domain contains some or all of the following:
- A list of sponsored links
- A list of popular categories
- A list of categories that contains the keywords

Here are some ways to identify parked domains:
- Look at the links. All of the links on a parked domain are paid links. There is no original, helpful content on the page.
- Look at the domain name (URL). On a parked domain, the domain name (URL) often has little or nothing to do with the content on the webpage. The links are usually generic and the linked pages are not really associated with the query.
- Look at the page on the Internet Archive. Go to http://www.archive.org/index.php to view the site as it appeared previously, when its original owner maintained it. If the original site was different, it is probably a parked domain.

Pages with Unhelpful Content and PPC Ads: Some pages contain content which was written specifically for spammers. Writers are paid to create articles on a wide range of topics; often the articles are very generic and do not provide a lot of good information, but they are original. You will not find these articles on other webpages. If the content makes sense and appears to be original, please do not assign a **Spam** flag. However, please consider such "superficially relevant" and "shallow" pages to be low quality and unhelpful.