# General Guidelines Version 3.18

March 30, 2011


## Part 1: Rating Guidelines ..... 5

1.0 Welcome to the Search Quality Rating Program! ..... 5
1.1 URL Rating Overview ..... 5
1.2 Important Rating Definitions and Ideas ..... 5
1.3 The Purpose of Search Quality Rating ..... 6
1.4 Raters Must Represent the User ..... 6
1.5 Internet Safety Information ..... 7
2.0 Understanding the Query ..... 8
2.1 Understanding User Intent ..... 8
2.2 Task Language and Task Location ..... 8
2.3 Queries with Multiple Meanings ..... 9
2.4 Classification of User Intent: Action, Information, and Navigation – "Do-Know-Go" ..... 9
2.4.1 Action Queries – "Do" ..... 9
2.4.2 Information Queries – "Know" ..... 11
2.4.3 Navigation Queries – "Go" ..... 11
2.4.4 Queries with Multiple User Intents (Do-Know-Go) ..... 12
3.0 The Language of the Landing Page ..... 13
4.0 The Rating Scale ..... 14
4.1 Vital ..... 14
4.1.1 Examples of English (US) Navigation Queries with Vital Pages for the Task Location ..... 14
4.1.2 Examples of Entity Queries with Vital Pages ..... 15
4.1.3 Vital Pages for People Queries ..... 16
4.1.4 Other Important Vital Concepts ..... 17
4.1.5 Vital Pages and Geographic Location ..... 19
4.2 Useful ..... 20
4.2.1 Examples of Useful Pages ..... 20
4.3 Relevant ..... 21
4.3.1 Examples of Relevant Pages ..... 22
4.4 Slightly Relevant ..... 22
4.4.1 Examples of Slightly Relevant Pages ..... 23
4.5 Off-Topic or Useless – On 3/30/2011, this label was changed from **Off-Topic** to **Off-Topic or Useless** ..... 25
4.5.1 Examples of Off-Topic or Useless Pages ..... 25
4.6 Unratable ..... 28
4.6.1 Unratable: Didn't Load ..... 28

4.6.2 Unratable: Foreign Language ..... 31
5.0 Rating: From User Intent to Assigning a Rating ..... 32
5.1 User Intent and Page Utility ..... 32
5.2 Location is Important ..... 33
5.3 Language is Important (This section is for Non-English Task Languages) ..... 34
5.4 Multiple Interpretations ..... 36
5.5 Specificity of Queries and Landing Pages ..... 38
5.6 Common Rating Problems ..... 42
5.6.1 Dictionary or Encyclopedia Results ..... 42
5.6.2 Action vs. Information Intent ..... 43
5.6.3 Queries that Ask for a List ..... 43
5.6.4 Misspelled and Mistyped Queries ..... 47
5.6.5 URL Queries ..... 49
5.6.6 New and Old Pages ..... 53
5.6.7 Search Engine Result Pages ..... 54
5.6.8 Video Landing Pages ..... 59
6.0 Flags ..... 60
6.1 Spam Flag ..... 60
6.2 Pornography Flag ..... 60
6.2.1 Clear Non-Porn Intent ..... 60
6.2.2 Possible Porn Intent ..... 61
6.2.3 Clear Porn Intent ..... 61
6.2.4 Reporting Illegal Images ..... 62
6.3 Malicious Flag ..... 63
6.4 Compatibility between Ratings and Flags ..... 63

**Part 2: URL Rating Tasks with Query Locations ..... 64**

1.0 Query Locations ..... 64
2.0 Location-Specific Rating Task Screenshot ..... 66
3.0 Assigning a Rating When There is a Query Location ..... 67
3.1 When Does the Query Location Matter? ..... 67
4.0 Query Location Rating Examples ..... 69

**Part 3: Rating Examples ..... 76**

1.0 Named Entity Queries ..... 76
2.0 Action Queries ..... 83
3.0 Information Queries ..... 86
4.0 Queries that Ask for a List ..... 89

Proprietary and Confidential – Copyright 2011

**5.0 Rating Examples for Task Locations other than English (US)** ..... 93

## Part 4: Webspam Guidelines ..... 95

**1.0 What is Webspam ?** ..... 95
**1.1 The Relationship between Ratings and Spam** ..... 95
**1.2 Why do Spammers Create Spam Pages?** ..... 95
**1.3 When to Check for Spam** ..... 95
**2.0 Browser Requirement** ..... 96
**3.0 Looking for Technical Signals** ..... 96
**3.1 Hidden Text and Hidden Links** ..... 97
**3.1.1 Apply Ctrl-A to the Landing Page** ..... 97
**3.1.2 Disable CSS** ..... 97
**3.1.3 Disable JavaScript** ..... 98
**3.1.4 View the Source Code** ..... 98
**3.1.5 Look Outside the Normal Viewing Area** ..... 99
**3.2 Keyword Stuffing** ..... 99
**3.2.1 Keyword Stuffing in the URL** ..... 99
**3.3 Sneaky Redirects** ..... 100
**3.3.1 Using "Whois"** ..... 100
**3.4 Cloaking** ..... 101
**3.4.1 JavaScript Redirects** ..... 101
**3.4.2 100% Frame** ..... 101
**4.0 Helpful Webpages vs. Spam Webpages** ..... 101
**4.1 Pages with Copied Content and PPC Ads** ..... 102
**4.1.2 Copied Text and PPC Ads** ..... 102
**4.1.3 Feeds and PPC Ads** ..... 102
**4.1.4 Doorway Pages** ..... 102
**4.1.5 Templates and Other Computer-Generated Pages** ..... 103
**4.1.6 Copied Message Boards** ..... 103
**4.1.7 Recognizing Copied Content** ..... 103
**4.2 Fake Search Pages with PPC Ads** ..... 104
**4.3 Fake Blogs with PPC Ads** ..... 104
**4.4 Fake Message Boards with PPC Ads** ..... 104
**4.5 Copied Content that is NOT Spam** ..... 105
**5.0 Commercial Intent** ..... 105
**5.1 Thin Affiliates** ..... 105
**5.1.1 Recognizing Thin Affiliates** ..... 105
**5.1.2 Not all Affiliates are Thin** ..... 106

5.1.3 Recognizing True Merchants ..... 106
5.2 Pure PPC Pages ..... 106
5.3 Parked (Expired) Domains ..... 107
5.4 Pages with Unhelpful Content and PPC Ads ..... 107
6.0 Phishing Websites ..... 108
7.0 Spam and the Resolving Stage ..... 108
8.0 Conclusion ..... 109

Part 5: Using EWOQ ..... 110

1.0 Introduction ..... 110
2.0 Accessing the EWOQ Rating Interface ..... 110
3.0 Rating ..... 110
4.0 Rating Home Screenshots ..... 111
5.0 Resolving Tasks (Re-rating Unresolved Tasks) / Moderators ..... 116
6.0 Commenting Etiquette ..... 118

Part 6: Quick Guide to URL Rating ..... 120

Part 7: Quick Guide to Webspam Recognition ..... 123

# 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/english_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 you are comfortable with. Please feel free to release rating tasks if they contain unknown or suspicious file formats.

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.

## 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.altavista.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<br><br>Find a calendar that displays holidays<br><br>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 don't 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-usa.com/useugame2007/html/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/ | Reputable 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<br><br>https://www.cia.gov/cia/publications/factbook/geos/sz.html | Travel guide on Switzerland<br><br>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/ | Office homepage of YouTube |
| [ebay], Italian (IT) | Go to the Italian eBay homepage | http://www.ebay.it/ | Official homepage of eBay Italy |
| [harvard admissions], French (FR) | Go to the admissions page on the Harvard website | http://admissions.college.harvard.edu/index.html | Office of Admissions page on the official Harvard 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/Nikon-Electronics/b?ie=UTF8&node=1084298 | The landing page is the "Nikon" page on the target.com website. There are over 30 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/2010/04/03/apple-ipad-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. Don't 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 Wal-Mart 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/<br><br>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.<br><br>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/<br><br>Non-Standard URL: http://www.officedepot.com/index.do | The homepage and target of the query.<br><br>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)<br><br>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/ | Madonna's official homepage |
| 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)<br><br>[Baseball hall of fame], English (US) | http://www.nps.gov/stli/<br><br>http://baseballhall.org/ | Official page on the government website<br><br>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 |
| Government officials | [President Obama], English (US) | http://www.whitehouse.gov/administration/president-obama/ | Official page on the government 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

Queries for famous people, such as [george bush], [Madonna], [david beckham], have obvious dominant interpretations. Queries for common names, such as [bob smith] and [mary jones], which do not have a dominant interpretation, can have no **Vital** result. If you are not sure about a name you don't recognize, try doing query research.

A query for a non famous person can have a **Vital** page if the person is uniquely specified or has a very unusual or unique name so that there is a clear dominant interpretation. For example, Dave Jones is a common English name and the query **[dave jones], English (UK)** can have no **Vital** result because we don't know which Dave Jones the user wants. However, the very specific query **[dave jones codemonkey], English (UK)** does have a clear dominant interpretation.

Homepages, blogs, and social networking pages have become very popular, and many famous and non famous people now have multiple "official" personal pages on the Web. People may have multiple homepages, multiple blogs, and multiple pages on various social networking sites, such as MySpace, Facebook, Friendster, Mixi, LinkedIn, Twitter, YouTube, etc. Official homepages of *all* types are **Vital** for famous people (and for non-famous people who have unusual, uniquely identifiable names).

Social networking pages for small groups of people (such as social clubs or musical bands) are also considered **Vital**.

Social networking pages for companies are **NOT** considered **Vital.**

It can sometimes be difficult to determine if a homepage, blog, or social networking page is official. Usually, official webpages for famous people are "professional" in appearance and are often linked to from the individual's other official pages or from a Wikipedia article about the person. Please use your judgment and have high standards.

Here are some examples:

| Query | URL of the Landing Page | Description | English (US) Vital Page? |
|---|---|---|---|
| [Hillary Clinton], English (US) | http://www.hillaryclinton.com/ | Hillary Clinton's official campaign webpage. Even though the campaign is over, the page still exists for the purpose of accepting contributions to clear up her campaign debt. | **Yes** |
| [Derek Jeter], English (US) | http://newyork.yankees.mlb.com/team/player.jsp?player_id=116539 | Derek Jeter is a famous baseball player who plays for the New York Yankees. This is his webpage on the official Yankee website. | **Yes** |
| [Arianna blog], English (US) | http://www.huffingtonpost.com/ | Arianna Huffington is a famous blogger. This is the homepage of The Huffington Post, a blog and commentary website founded by her. | **Yes** |
| [oprah], English (US) | http://www.oprah.com/ | Oprah Winfrey is a famous talk show host. This is the homepage of Oprah's magazine, radio station, book club, etc. | **Yes** |
| [Lynn Bozof], English (US) | http://www.linkedin.com/pub/dir/?last=bozof&first=lynn | Lynn Bozof is an uncommon name. This is her LinkedIn page. | **Yes** |
| [Dave Smith], English (US) | http://www.davewsmith.com/ | Dave Smith is a common name without a dominant interpretation. A personal webpage for someone named Dave Smith is not **Vital**. | **No** – non-famous people with common names can't have **Vital** pages |
| [Britney Spears], English (US) | http://www.youtube.com/user/britneytv | Britney Spears is a famous singer and celebrity. This is her YouTube Channel page. | **Yes** |

| Query | URL of the Landing Page | Description | English (US) Vital Page? |
|---|---|---|---|
| [green day], English (US) | http://www.greenday.com/ | Green Day is an American rock band. This is the band's official homepage. | **Yes** |
| [green day], English (US) | http://www.myspace.com/greenday | This is Green Day's MySpace webpage. | **Yes** |
| [green day], English (US) | http://www.youtube.com/user/greenday | This is Green Day's YouTube Channel page. | **Yes** |
| [photobucket], English (US) | http://www.myspace.com/photobucket | Photobucket is an online photo sharing company. This is the company's MySpace page. | **No** – social networking pages can only be **Vital** for people, bands, and small groups. They are not **Vital** for companies. |
| [Ford], English (US) | http://www.facebook.com/ford | Ford is an automobile manufacturer. This is the company's Facebook page. | **No** – social networking pages are only **Vital** for people, bands, and small groups. They are not **Vital** for companies. |
| [Sheboygan Press], English (US) | http://twitter.com/sheboyganpress | Sheboygan Press is a newspaper. This is the newspaper's Twitter page. | **No** – social networking pages are only **Vital** for people, bands, and small groups. They are not **Vital** for companies. |
| [sesame street], English (US) | http://www.youtube.com/user/SesameStreet | Sesame Street is a well-known children's TV show. This is the Sesame Street YouTube Channel page. | **No** – social networking pages are only **Vital** for people, bands, and small groups. They are not **Vital** for TV shows. |
| [toyota], English (US) | http://blog.toyota.com/ | Toyota maintains a company blog to communicate with the public. | **No** – company blogs are not **Vital**, unless the blog is specified in the query |

### 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.<br><br>Americans with Disabilities Act<br>American Dental Association<br>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/<br>http://www.bn.com<br>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<br>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/<br>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/ |
| [simpsons], English (US) | http://www.simpsons.com/ | This is the "real" official **Vital** page for the query: http://www.thesimpsons.com/index.html |
| [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 doesn't 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 doesn't match the task location and a better version exists. (If a better version for the task location doesn't 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)<br>[Stanford], Chinese (CN)<br>[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)<br>[University of Seville], Chinese (CN)<br>[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)<br>[Microsoft.com], China (CN)<br>[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. |
| [french open website], Spanish (ES)<br>[french open website], Spanish (MX)<br>[french open website], Spanish (AR) | http://www.rolandgarros.com/es_FR/index.html | **Appropriate Vital** | The French Open has three versions of its website: French, Spanish, and English. The landing page is the Spanish version. This page is **Appropriate Vital** for all Spanish-speaking task locations. |
| [bbc], Arabic (EG)<br>[bbc], Arabic (SA)<br>[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. |

| Query | URL | Rating | Description |
|---|---|---|---|
| [United Nations], English (US)<br>[United Nations], Chinese (CN)<br>[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)<br>[Ikea], Chinese (CN)<br>[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)<br>[bbc], Chinese (CN)<br>[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)<br>[ikea], Chinese (CN)<br>[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)<br>[ikea], English (UK)<br>[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 |
|---|---|---|---|
| [is poison oak contagious?], English (US) | Find the answer to this question. This is an information query. | http://www.cincinnatichildrens.org/health/info/allergy-asthma/diagnose/ivy.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<br><br>http://money.cnn.com/quote/quote.html?symb=CSCO<br><br>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. |

| Query | Likely User Intent | Useful Pages | Explanation |
|---|---|---|---|
| [meningitis symptoms], English (US) | Find information on the symptoms of meningitis. This is an information query. | http://www.webmd.com/hw/infection/aa34586.asp<br><br>http://www.nlm.nih.gov/medlineplus/ency/article/000680.htm<br><br>http://www.cdc.gov/meningitis/about/faq.html<br><br>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 and performed by Sting. This is an information query. | http://www.sting.com/discog/?v=so&a=1&id=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=-8691839177585748 79 | 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 isn't 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. |
| [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 and performed by Sting. This is an information query. | http://www.mp3lyrics.org/p/police/every-breath-you-take/<br><br>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) aren't 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 R elevant** 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://www.pregnancysymptoms.net/Basics/info.html | 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 wouldn't 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.<br><br>Note: URLs that contain informational terms like "pregnancy symptoms" should not be rated **Vital**, even when they match the query. |
| [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 wouldn't 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<br>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/. |

| Query | Likely User Intent | Slightly Relevant Pages | Explanation |
|---|---|---|---|
| [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 doesn't display any maps. The page is related to the query but doesn't 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 doesn't 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 4 of the "General Guidelines".

Please note that not all pages with copied content are considered "low quality". The website 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 |
|---|---|---|---|
| [cancer symptoms], English (US | Find information about cancer symptoms | http://cancer-symptoms.org/ | The landing page has information about symptoms of various kinds of cancer, so it is not **Off-Topic or Useless**, but the page is disorganized, the text appears to have been copied from another website, there are many ads, and some of the links don't work. Even though the name of the domain matches the query, the content is low quality. |
| [pain esophagus], English (US) | Find information related to pain in the esophagus | http://www.wrongdiagnosis.com/symptom/esophagus-pain.htm | Even though the title of the landing page matches the query, the page is just superficially relevant. There really isn't much content on the page.<br><br>Clicking the links doesn't take users to helpful information either. In fact, this page links to itself. If you hover your mouse over the links, you will see that they are just ads that are unrelated to the names of the links.<br><br>This page is low quality and many users would not trust this information. |

| Query | Likely User Intent | Slightly Relevant Pages | Explanation |
|---|---|---|---|
| [dvd label maker], English (US) | Download software to make DVD labels | http://wareseeker.com/Graphic-Apps/ronyasoft-cd-dvd-label-maker-1.02.01.zip/413c4193b | The landing page appears to offer DVD label maker 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 doesn't 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. |
| [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 4, 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 3, Webspam Guidelines). |

### 4.5 Off-Topic or Useless – On 3/30/2011, this label was changed from Off-Topic to 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 | **Doesn't 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 doesn't 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/ | **Doesn't 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 doesn't fit the intent of the query. It would be helpful for very few or no users. |

| Query | Likely User Intent | Off-Topic or Useless Pages | Explanation |
|---|---|---|---|
| [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 | **Doesn't 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 doesn't fit the intent of the query. |
| [gmail login], English (US) | Go to the Gmail login page | https://login.yahoo.com/config/login_verify2?&.src=ym | **Doesn't 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 doesn't 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/ | **Doesn't fit the task location:** 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 mentions the **Appalachian Trail** and **long**-term parking near the Amicalola Falls West Ridge Trail. It also displays the words **hot** and **is** 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 / doesn't 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 prostrate treatment. The page is useless and should be rated **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 doesn't 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 doesn't exist. The page is **Off-Topic or Useless**. |
| [how to quit smoking], English (US) | Find information on ways to quit smoking | http://www.getarticles.info/2011/03/blu-cigs-can-be-a-help-for-you-in-quiting-smoke/ | **Gibberish:** The landing page has gibberish text. Read these sentences: *"With each and every pack, one particular USB charger and one wall charger is existing. The two of the sorts permit convenience in charging the products."* 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://www.fashiontrending.info/2z01/dress-me-up-for-our-acq-party/ | **Gibberish:** This landing page also has gibberish text. It is even worse than the previous example. Read this sentence: *"What among one of the most loony point could possibly be the Herve Leger earnings over the internet decreased marketing price promotions, permit me to commit."* 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:

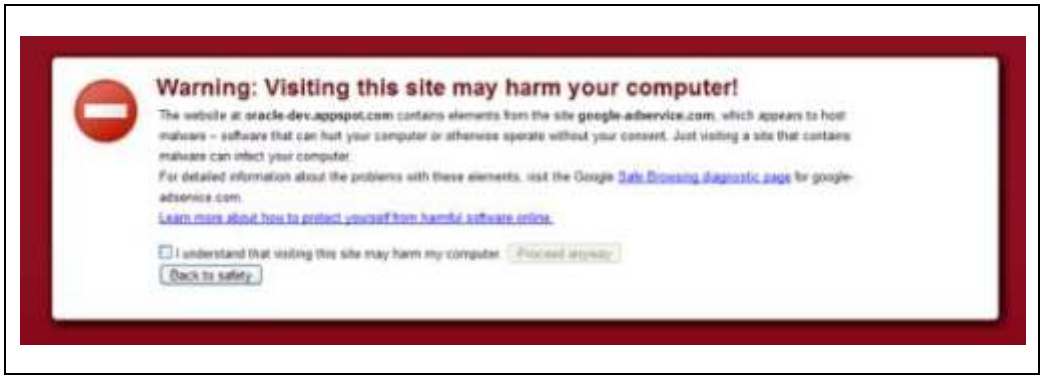
Pages with error messages ***and no other content on the page***
Pages with non-working redirects ***and no other content on the page***
Completely blank pages
Pages with malware warnings, such as "Warning – visiting this web site may harm your computer!"
Pages with certificate acceptance requests

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. |
| [united nations], English (US) | http://disarmament.un.org/wmd/bwc/index.html | "Unable to open http://disarmament.un.org/wmd/bwc/index.html. The Internet site reports that the item you requested could not be found. (HTTP/1.0 404)" | **Didn't Load** | The request cannot be completed. There is no content on the page. |
| [SIAD], English (US) | http://www.siad.org/http%20403%20(forbidden).htm | "You are not authorized to view this page. You might not have permission to view this directory or page using the credentials you supplied." | **Didn't Load** | The page displays a 403 error message. There is no content on the page. |

| Query | URL of the Landing Page | Landing Page Error Message | Rating | Explanation |
|---|---|---|---|---|
| [seonggeo], English (US) | http://www.jungang.or.kr/design05/user/index_intro.php | HTTP 오류 404 - 파일 또는 디렉터리를 찾을 수 없습니다. IIS | **Didn't Load**<br><br>Note: The language of the landing page should be flagged "Foreign Language". | Even though the message is in Korean (HTTP 오류 404), we can tell that the page didn't load **Unratable: Didn't Load**. |
| [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 witherror 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 |
|---|---|---|---|---|
| [snow shoes], English (US) | http://shop.volcom.com/on/demandware.store/Sites-Volcom-Site/default/Search-Show?q=snow+shoes&simplesearch=Go | ?We're sorry, no products were found for your search" snow shoes did you mean "snow shots?." | **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 snow shoes, it is **Off-Topic or Useless.** |
| [shirts], English (US) | http://www.tshirtzetc.com/ | "Website Under Construction. " | **Slightly Relevant** | The landing page is the homepage of a site under construction for Tshirtz Etc, a store located in Rockford, Illinois that does custom printing on t-shirts, sweatshirts, caps, etc. At this time, only a few of the links work, but the page does have the store's street address and contact information. Since the query is so broad and the site is only partially functional, the page would be helpful for few users. |
| [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<br><br>Find a calendar that displays holidays<br><br>Find an online calendar to use | Site on which to make customized, printable calendars: http://www.timeanddate.com/calendar/: **Useful**<br><br>Yahoo calendar: http://calendar.yahoo.com/: **Useful** | Article on the history of different types of calendars: http://astro.nmsu.edu/~lhuber/leaphist.html : **Relevant**<br><br>Basic definitions of the word “calendar”: http://wordnet.princeton.edu/perl/webwn?s=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 isn't 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 isn't 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 Wal-Mart 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 don’t.

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%B0%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. |
| [Nanoscale Materials Tracy Zontek Vol.55, Iss.3, pg.34], Korean (KR) | Find and read a document titled “Nanoscale Materials”, written by Tracy Zontek | http://proquest.umi.com/pqdweb?index=20&did=1985258351&SrchMode=1&sid=1&Fmt=3&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1274393370&clientId=124494 | **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.<br><br>[windows], English, (US): the Microsoft operating system. **Most** users who type this query want results on the Microsoft Windows operating system.<br><br>[amazon], English (US): the popular website www.amazon.com. **Most** users who type this query want to go to the Amazon website.<br><br>[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.**<br><br>[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.<br><br>[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.<br><br>[amazon], English (US): The rainforest or river in South America. **Some** users who type this query could want results about the river or rainforest.<br><br>[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.<br><br>[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.<br><br>[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**<br><br>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.<br><br>[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.<br><br>[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**<br><br>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<br><br>[flowers], English (US): Order flowers online, or learn about types of flowers or find pictures of flowers.<br><br>[credit cards], English (US): Find a credit card company, apply for a card, or compare different brands of credit cards<br><br>[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<br><br>[flowers], English (US): Find a definition of the word "flower"<br><br>[credit cards], English (US): Read an encyclopedia article on the history of credit cards<br><br>[amazon], English (US): Read an encyclopedia article about Amazon.com | **Relevant** to **Off-Topic or Useless**<br><br>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.<br><br>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.<br><br>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~~q70726f63657373696e6774696d653a3e313930302d30312d3031~~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.<br><br>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 isn't **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.<br><br>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.<br><br>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.<br><br>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.<br><br>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.<br><br>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.<br><br>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.<br><br>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.<br><br>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 three models of the Accord: the Accord Sedan, the Accord Coupe, and the Accord Crosstour. | http://automobiles.honda.com/accord/ | **Vital**: The landing page is the official Honda Accord page.<br><br>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 pictures and prominent "Accord" and "Crosstour" links 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.<br><br>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.<br><br>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/<br><br>http://automobiles.honda.com/accord-coupe/<br><br>http://automobiles.honda.com/accord-crosstour/ | **Useful:** The landing pages are the official Accord Sedan, Accord Coupe, and Accord Crosstour pages.<br><br>These landing pages are more specific than the query, but since there are only three Accord models and they are all popular, official pages (or other very helpful pages) for any of the three 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.<br><br>The landing page does not quite fit the query. It has Accords prominently displayed and may be helpful for some users, but we don't 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.<br><br>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.<br><br>The landing page fits the query. |
| | | http://sites.target.com/site/en/spot/page.jsp?title=store_locator_new&ref=nav_storelocator | **Useful or Relevant** – the landing page is the "store finder" page on the Target website.<br><br>The landing page is more specific than the query, but many or some users would be interested in this page. |
| | | http://weeklyad.target.com/target/default.aspx?action=entryflash&ref=sc_iw_l_0_1 | **Useful or Relevant** – the landing page is the "weekly ads" page on the Target website.<br><br>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.<br><br>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.<br><br>The landing page is more specific than the query. Fewer users would be interested in this page. |
| | | http://www.target.com/Boys-Shorts-Clothing-Shoes-Kids/b/ref=sc_iw_r_1_1/178-4746585-1881721?node=16008751 | **Slightly Relevant**– the landing page is the "boys' shorts" page on the Target website.<br><br>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 areoften 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. | http://encarta.msn.com/dictionary_1861607624/e_g_.html | **Useful** or **Relevant** | This is a good explanation of the abbreviation "e.g." 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<br><br>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 don't 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 aren't 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<br><br>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.<br><br>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.<br><br>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.<br><br>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/<br><br>http://www.zaxbys.com/home.aspx<br><br>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.<br><br>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.<br><br>The landing page fits the query. Most users would find this page helpful. |
| | | http://www.gatortots.com/pages/toys-for-babies.htm | **Useful** or **Relevant**– This page has a nice selection of baby toys by category. Gator Tots is not a well-known merchant, but it's a high quality page.<br><br>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.<br><br>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.<br><br>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 4 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.<br><br>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.<br><br>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.<br><br>The landing page doesn't 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 don't specifically know what users want, there are helpful and unhelpful results for this query. | http://www.expedia.com/Hotels<br><br>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.<br><br>These landing pages fit the query. Most users would find these pages helpful. |
| | | http://www.marriott.com/<br><br>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.<br><br>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.<br><br>These landing pages are more specific than the query, but the pages are still helpful for many users. |
| | | http://www.motel6.com/<br><br>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 don't offer as many options in price or features.<br><br>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.<br><br>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.<br><br>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: Don't 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 | Official homepage of Beyonce's website | **Vital** |
| [David Bcekham], English (US) | The only reasonable interpretation is the soccer player named David Beckham. | http://www.davidbeckham.com/ | Official homepage of David Beckham's website | **Vital** |

People queries can be difficult to rateHere 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.<br><br>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/ | Official homepage of Jamie Fox, the guitarist | **Useful** |
| | | http://jamiefoxphotography.com/ | Official homepage of Jamie Fox Photography | **Relevant** or **Useful** |
| | | http://www.jamiefox.net/ | Homepage of Jamie Fox, a web developer | **Relevant** or **Useful** |
| | | http://www.jamiefoxx.com/ | Official homepage 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.<br><br>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.<br><br>Note: Since Michael Jordan is retired from professional basketball, there is no employer /employee relationship between him and the NBA. Therefore, this page can't be **Vital**. | **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".

### 5.6.5 URL Queries

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 don't refer to a particular website or webpage.

Here are some examples of English (US) queries:

| **URL queries** | **Website Name/Webpage Name Queries**<br>(these are "go" queries, with no "URL parts") | **"Generic" Queries** |
|---|---|---|
| [ebay.ca]<br>[amazon.com]<br>[people.com]<br>[bbc.co.uk]<br>[www.dealbook.com]<br>[mail.yahoo.com]<br>[news google.com]<br>[tax form 1040 irs.gov]<br>[rei.com] | [ebay]<br>[amazon]<br>[people]<br>[bbc]<br>[dealbook]<br>[yahoo mail]<br>[google news]<br>[irs 1040 tax form official page]<br>[rei kayak page] | [couches]<br>[diabetes]<br>[weight loss]<br>[tax forms]<br>[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 [imagehosting.us], English (US), both http://imagehosting.us/ and http://www.imagehosting.us/index.php?action=show&ident=1478428 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)<br><br>[yahoo.xcom], English (US)<br><br>[yahoo.co], English (US) | Even though these URLs don't load, it is clear the user wants to go to Yahoo. | **Vital** landing page URL: http://www.yahoo.com/ |
| [simpsons.com], English (US) | In this case, the landing page is spam. It is very likely that the user wants to navigate to www.thesimpsons.com/. However, we will respect the query as written and consider www.simpsons.com to be dominant. | **Vital** landing page URL: http://www.simpsons.com (You will also need to add a **Spam** flag. Please see Part 4 of the "General Guidelines".)<br><br>**Useful** landing page URL: http://www.thesimpsons.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 birdcages.com, 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 kamasutra.com, but that probably isn't 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 weightloss.com, which has helpful, common sense information about losing weight, but users probably aren't 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 couches.com, 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 shouldn't assume the query is a website name.

In other words, just because the query is [couches] doesn't mean that the result http://www.couches.com 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 http://www.greatamericanphotocontest.com/, a website where users post baby pictures which are supposed to be entered in a baby photo contest each month | http://www.greatamericanphotocontest.com/ | The landing page is the target of the query | **Vital** |
| | | http://www.complaintsboard.com/byurl/greatamericanphotocontest.com.html | 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** |
| | | http://www.419legal.org/fraudulent-website/29043-great-american-photo-contest.html | 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** |
| | | http://www.quantcast.com/greatamericanphotocontest.com | 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** |
| | | http://www.killerstartups.com/Site-Reviews/greatamericanphotocontest-com-baby-photo-contest | 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<br><br>Note: When these guidelines were revised, there were many concerns about Facebook privacy and security. | 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://www.quarkbase.com/ratemyprofessors.com | 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 2009 results: http://en.wikipedia.org/wiki/2009_US_Open_Golf_Championship | Wikipedia page with the 2007 results: http://en.wikipedia.org/wiki/2007_U.S._Open_Golf_Championship |
| [golden globe best film drama], English (US) | Find **the most recent** winner of this award. This is an information query. | Page on the BBC website with this information: http://news.bbc.co.uk/2/hi/entertainment/8465435.stm | Page on about.com with the 2006 winner of this award: 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.reuters.com/article/idUSTRE5981JK20091009<br><br>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 hasn't 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** |

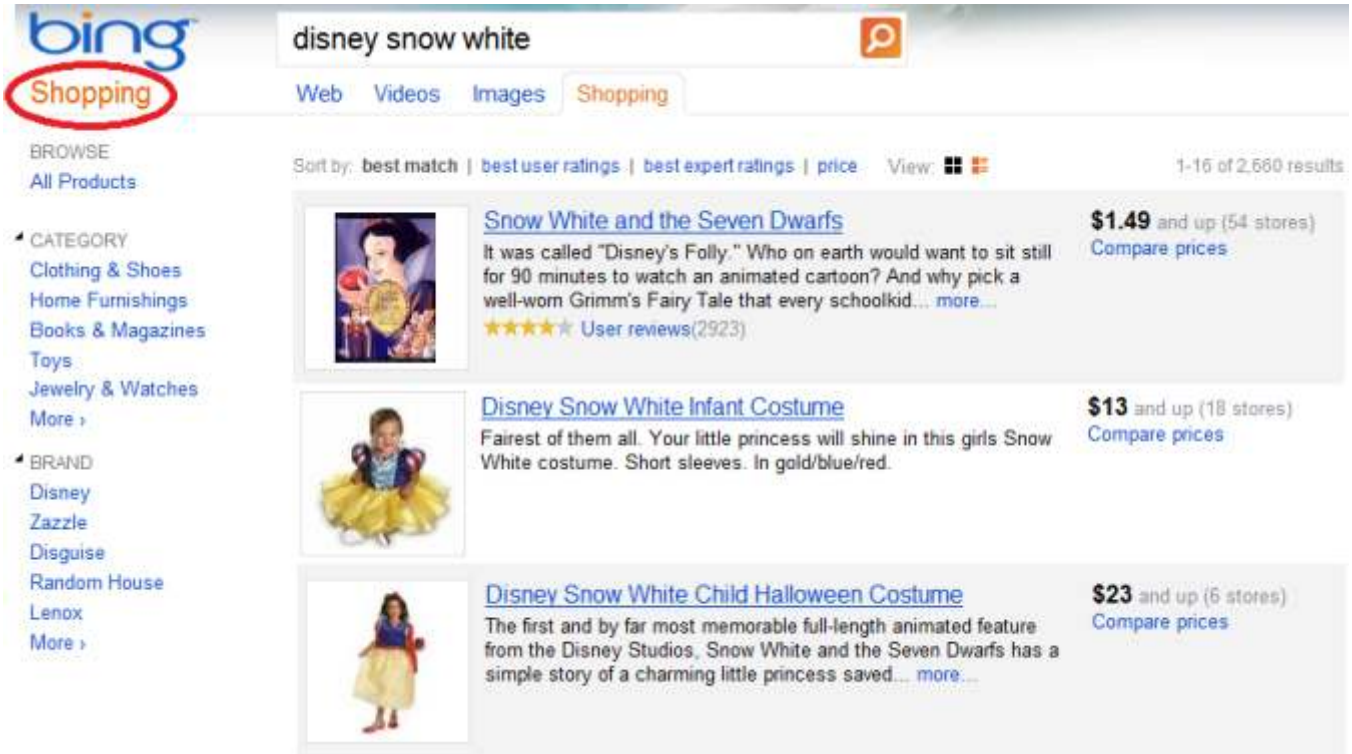
### 5.6.7 Search Engine Result Pages

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.

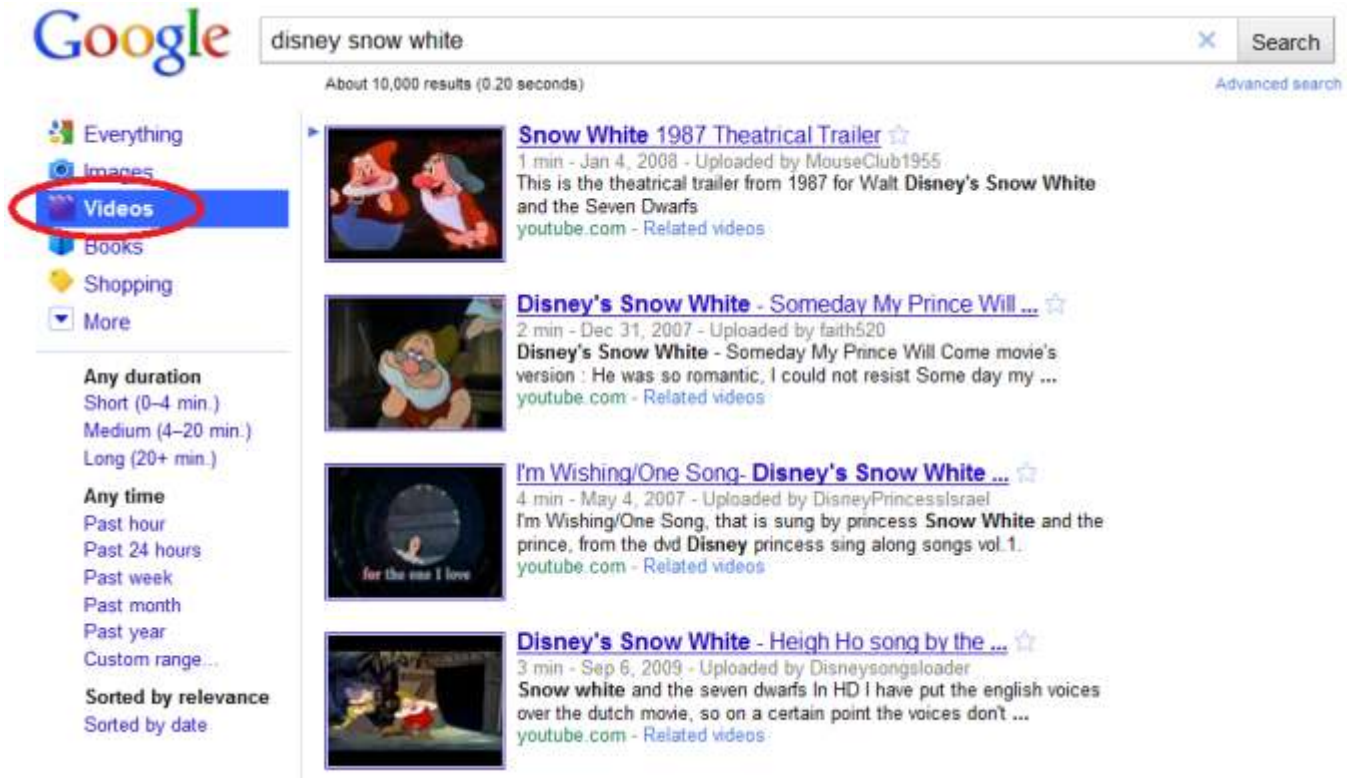
Here are examples of search engine results pages. These are pages users see after entering queries on a search engine.

## Search Results Page

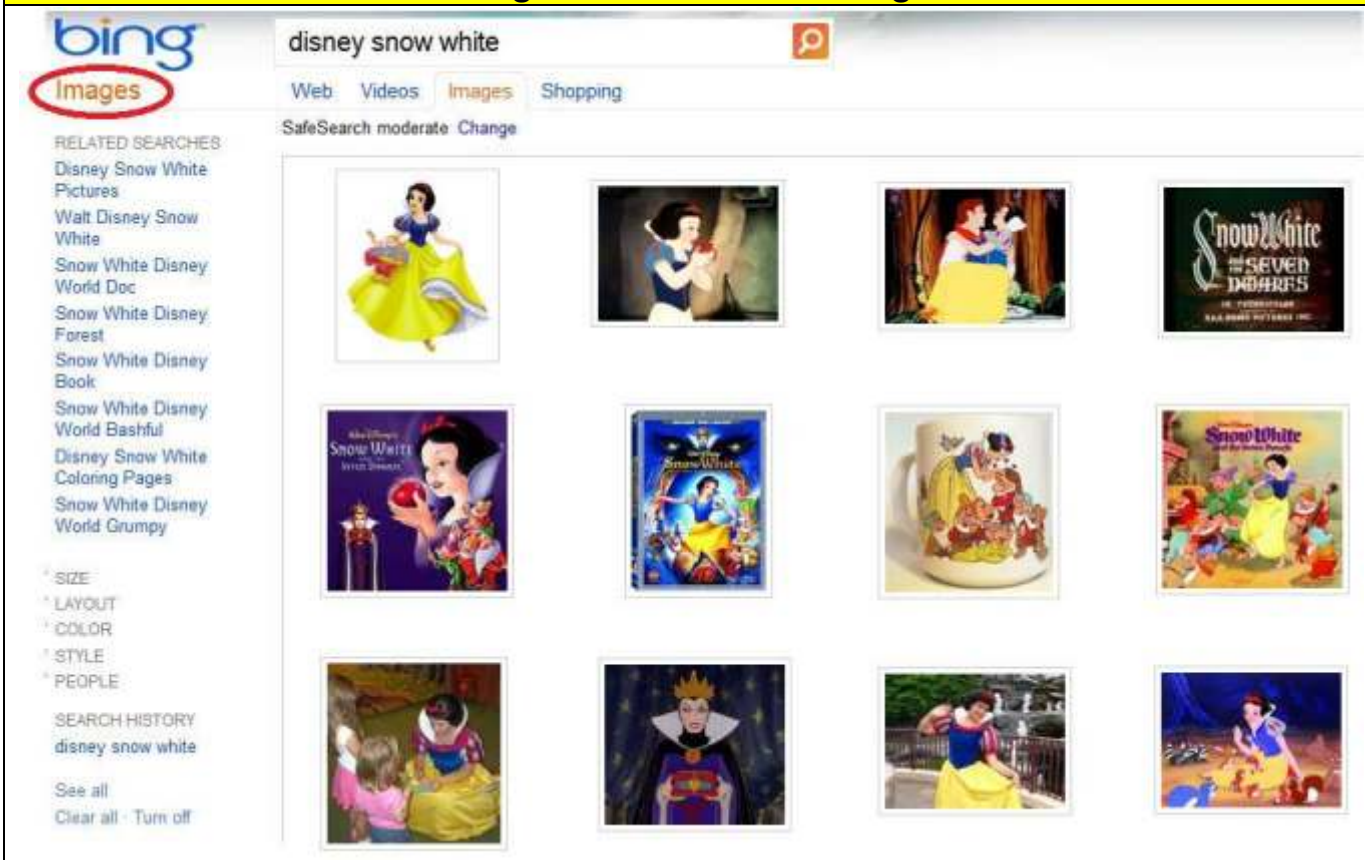
## Shopping Search Results Page



## Video Search Results Page




## Image Search Results Page



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 aren't 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 don't 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 can't 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=7x1xthuk-Iw&feature=related | 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 |
| [beatles live], English (US) | http://www.youtube.com/watch?v=1eyBha-gx2U&feature=related | The query is looking for information about or a video of a Beatles live performance. The landing page has a video of a live performance of the Beatles in Tokyo. | **Relevant** or **Useful** | English |
| [Kasal, Kasali, Kasalo], English (US) | http://www.youtube.com/watch?v=us6Uaewi1mU | 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 4 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 don't 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 porn result 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:

The user intent is clearly not porn, so a porn result 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.

| 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 |

### 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://www.med.umich.edu/1libr/pa/pa_bdiscphy_hhg.htm | **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 information about girls, such as girls fashion, girls names, girls activities, etc. | http://www.ptgirlscouts.org/ | **Relevant** | No |
| [pictures of girls], English (US) | Find information about girls, such as girls fashion, girls names, girls activities, etc. | 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 |

| Query | Likely User Intent | Landing Page | Rating | Porn Flag? |
|---|---|---|---|---|
| [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.pichunter.com/all/cheerleaders.shtml 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 isn't porn, it shouldn't 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 can't 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 send an email with the link to your employer with "Illegal Image" in the subject line. Please do not include images in your email. Please send the link only.

By "link", we are referring to the URL of the image or the URL of the landing page. Please do not send the Task ID URL.

Here is an example of an image URL: http://www.cssnz.org/flower.jpg
Here is an example of a landing page URL: http://www.cssnz.org/flowers.php
Here is an example of what a Task ID URL looks like: https://www.google.com/evaluation/search/rating/task-edit?task=123456789. Please do not send the Task ID URL.

For most project types, please send the landing page URL. For Image Review projects, please send the image URL.

**Please do NOT attach or send images; just send the link only.**

## 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 can’t close a pop up on the page and you are forced to quit your Firefox browser.

# Part 2: URL Rating Tasks with Query Locations

## 1.0 Query Locations

All URL rating tasks have a task location, which is usually the country location.

Some URL rating tasks also have a "query location", which is associated with the geographic location of the user when he or she issued the query. The query location may be a zip code, town, city, city and state, etc. Usually, the query location is automatically detected by the search engine, but may come from the user's stated preferences.

For narrowly defined query locations, such as specific zip codes or towns, the relevant location may extend beyond the specified zip code or town boundaries. Remember that real users are sometimes looking for the nearest stores or restaurants. If those happen to be outside the specified location, that may be acceptable to the user. You will have to use your judgment about what is reasonable.

Here are some important things to know about tasks with query locations:

- You will rate from the perspective of someone living in the query location.
- Local pages (pages associated with the query location) that are helpful should receive high ratings.
- Pages that would be helpful to users in any query location should also receive high ratings.
- When the query is an entity, such as a business, organization, school, etc., and the entity has both an official homepage and official location-specific webpages, a rating of **Appropriate Vital** will apply to both the entity's homepage and the appropriate query location-specific webpage.

**Important:** Sometimes, users specify a location when they type a query. For example, in the query [pizza hut, Marietta, Georgia], the user has specified "Marietta, Georgia" as the location of interest. Some tasks have both a Query Location and a location specified in the query. When this happens, you should rate with respect to the location specified in the query, rather than the Query Location.

Here are examples of three types of tasks:

- The task has a location specified in the query.
- The task has a Query Location.
- The task has both a Query Location and a location specified in the query.

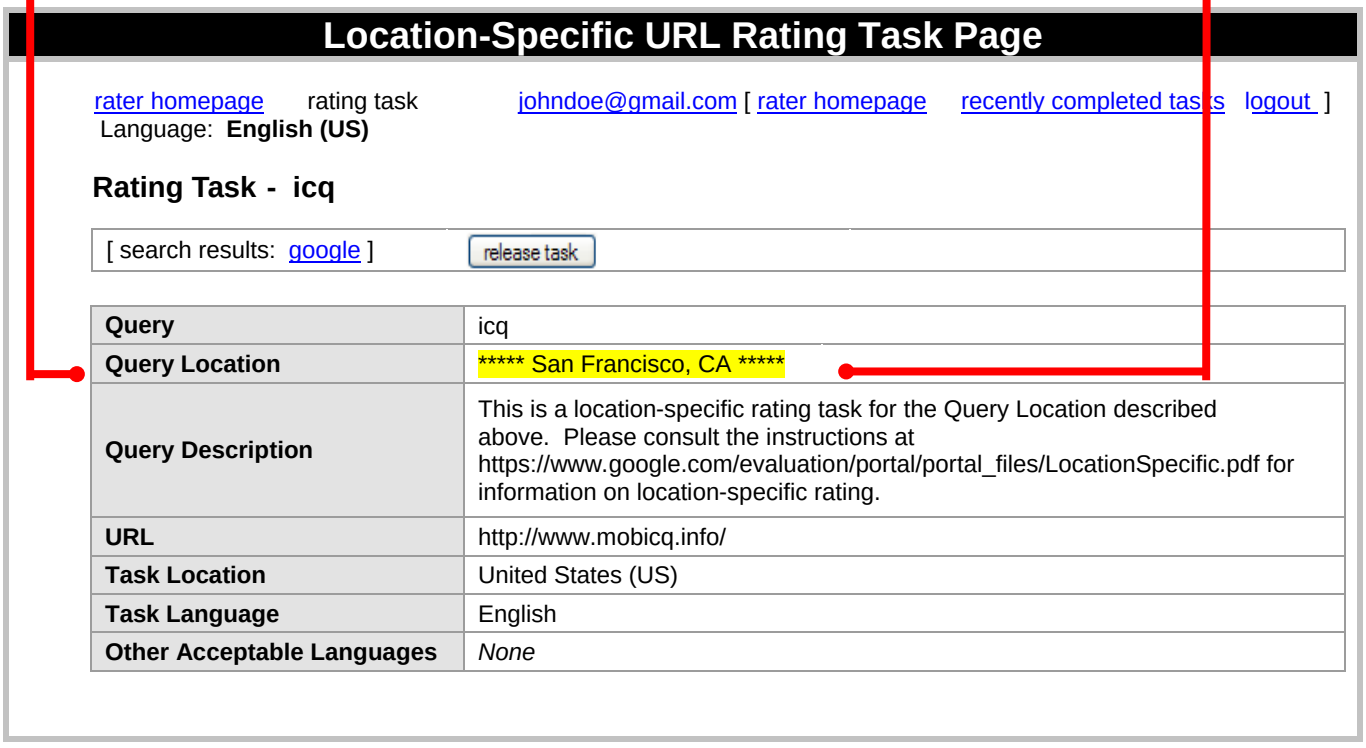
<table>
<tr><th>Task Type</th><th>Screenshot</th><th>Description</th></tr>
<tr>
<td>This is not a location-specific task because it does not have a Query Location.<br><br>Notice, however, that a location is specified in the query.</td>
<td>
<table>
<tr><td>Query</td><td>pizza hut san francisco</td></tr>
<tr><td>URL</td><td>http://www.yelp.com/biz/pizza-hut-san-francisco</td></tr>
<tr><td>Task Location</td><td>United States (US)</td></tr>
<tr><td>Task Language</td><td>English</td></tr>
<tr><td>Other Acceptable Languages</td><td>None</td></tr>
</table>
</td>
<td>The user wants Pizza Hut information for the San Francisco area.</td>
</tr>
<tr>
<td>This is a location-specific task because it has a Query Location.</td>
<td>
<table>
<tr><td>Query</td><td>pizza hut</td></tr>
<tr><td>Query Location</td><td>***** San Francisco *****</td></tr>
<tr><td>URL</td><td>http://www.yelp.com/biz/pizza-hut-san-francisco</td></tr>
<tr><td>Task Location</td><td>United States (US)</td></tr>
<tr><td>Task Language</td><td>English</td></tr>
<tr><td>Other Acceptable Languages</td><td>None</td></tr>
</table>
</td>
<td>The query was issued by a user living in San Francisco.<br><br>We can assume that the user is looking for a Pizza Hut restaurant in San Francisco.</td>
</tr>
<tr>
<td>This is also a location-specific task because it has a Query Location.<br><br>Notice, however, that a location is specified in the query.</td>
<td>
<table>
<tr><td>Query</td><td>pizza hut san francisco</td></tr>
<tr><td>Query Location</td><td>***** New York *****</td></tr>
<tr><td>URL</td><td>http://www.yelp.com/biz/pizza-hut-san-francisco</td></tr>
<tr><td>Task Location</td><td>United States (US)</td></tr>
<tr><td>Task Language</td><td>English</td></tr>
<tr><td>Other Acceptable Languages</td><td>None</td></tr>
</table>
</td>
<td>The query was issued by a user living in New York.<br><br>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 Query Location is New York.</td>
</tr>
</table>

Proprietary and Confidential – Copyright 2011

## 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 Query Location.

| Information | Standard URL Rating Task Page | Location-Specific URL Rating Task Page |
|---|---|---|
| Query Location | Standard URL Rating task home does not have this information. | ***** New York *****<br>***** 90210 *****<br>***** Dallas, TX *****<br>***** 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 |
| **Query Location** | ***** San Francisco, CA ***** |
| **Query Description** | This is a location-specific rating task for the Query Location described above. Please consult the instructions at https://www.google.com/evaluation/portal/portal_files/LocationSpecific.pdf for information on location-specific rating. |
| **URL** | http://www.mobicq.info/ |
| **Task Location** | United States (US) |
| **Task Language** | English |
| **Other Acceptable Languages** | *None* |

## 3.0 Assigning a Rating When There is a Query Location

In some tasks, the query location will be an important consideration in the rating you assign. For example:

**Query:** [IHOP restaurants], English (US)
**Query Location:** Boston, MA
The query location is an important consideration. Users in Boston who type this query are interested in IHOP restaurants in the Boston area, not other locations.

However, in many tasks the query is *not* associated with a specific location and the query location will not be a consideration at all. The rating you assign will be the same rating you would have assigned if the task did not have a query location. For example:

**Query:** [amazon.com], English (US)
**Query Location:** Boston, MA
The query location is not a consideration at all. Amazon.com is a website that is not associated with a specific location.

The query location makes a difference when the landing page would be more helpful to users in some locations than users in other locations.

## 3.1 When Does the Query Location Matter?

Here are some examples that demonstrate when the query location matters and when it doesn't.

| Query | Query Location | URL | Likely User Intent | Does the Query Location Matter in this Example? | Explanation |
|---|---|---|---|---|---|
| [facebook], English (US) | Birmingham, AL | http://www.facebook.com/ | The user in Birmingham, Alabama wants to go to the Facebook website at www.facebook.com. | **No**, because Facebook is a website that is not associated with a specific location. | The landing page is equally helpful to users in Birmingham, Alabama and other locations. It should be rated **Appropriate Vital** for any query location, or if there is no query location specified in the task. |
| [Benihana], English (US) | New York, NY | http://www.benihana.com/ | The user in New York City wants information about the Benihana restaurant in New York City or to go to the Benihana homepage. | **No**, because the homepage of the entity should get an **Appropriate Vital** rating, even if a location-specific webpage exists. | The official Benihana homepage should be rated **Appropriate Vital** for New York City or any other query location, or if there is no query location specified in the task. |

| Query | Query Location | URL | Likely User Intent | Does the Query Location Matter in this Example? | Explanation |
|---|---|---|---|---|---|
| [Benihana], English (US) | New York, NY | http://www.benihana.com/locations/newyorkwest-ny-we | The user in New York City wants information about the Benihana restaurant in New York City or to go to the Benihana homepage. | **Yes**, because users in New York City are interested in Benihana restaurants in New York City, not other locations. | The landing page is the official webpage for the Benihana restaurant located in New York City. It should be rated **Appropriate Vital** for the query location. However, it would be rated **Other Vital** for other query locations or **Slightly Relevant** if there is no query location specified in the task. |
| [ice rink], English (US) | College Station, TX | http://www.arcticwolfice.com/ | The user in College Station, Texas wants information about local ice rinks. | **Yes**, because users in College Station are interested in ice rinks in College Station, not other locations. | The landing page is the official homepage of Arctic Wolf Ice Center, the only ice rink in College Station and therefore the dominant interpretation for this query location. It should be rated **Appropriate Vital** for the query location. However, it should be rated **Off-Topic or Useless** for other query locations or **Slightly Relevant** if there is no query location specified in the task. |
| [weather conditions], English (US) | Las Vegas, NV | http://www.wundergroud.com/US/NV/Las_Vegas.html | The user in Las Vegas, Nevada wants information about local weather conditions. | **Yes**, because users in Las Vegas are probably interested in the weather in Las Vegas, not other locations. | The landing page has information about the current weather conditions in Las Vegas. It should be rated **Useful** for the query location. However, it should be rated **Off-Topic or Useless** for other query locations or **Slightly Relevant** if there is no query location specified in the task. |
| [patriots], English (US) | Concord, MA | http://www.patriots.com/ | The dominant interpretation of this query for the user in Concord, Massachusetts is the New England Patriots football team. The user wants information about the team or to go to the team's official homepage. | **Yes**, because the New England Patriots football team is very popular with users in New England (where Concord, Massachusetts is located). It is highly likely that users in Concord issuing this query have this football team in mind. | The landing page is the football team's official homepage. It should be rated **Appropriate Vital** for the query location. However, it should be rated **Useful** for query locations outside New England, or if there is no query location specified in the task, because it is a common interpretation of the query. |

## 4.0 Query Location Rating Examples

| Query | Query Location | Likely User Intent | URL of the Landing Page | Rating | Explanation |
|---|---|---|---|---|---|
| [benihana's], English (US) | New York City | Find information about the Benihana restaurant in New York City or go to the Benihana homepage. | http://www.benihana.com/locations/newyorkwest-ny-we | **Appropriate Vital** | Benihana is a chain of restaurants. The landing page is the official webpage for the Benihana restaurant located in the heart of New York City. It should be rated **Appropriate Vital** for the New York City Query Location. |
| [benihana's], English (US) | New York City | Find information about the Benihana restaurant in New York City or go to the Benihana homepage. | http://www.benihana.com/ | **Appropriate Vital** | The landing page is the official Benihana homepage. It should be rated **Appropriate Vital** for the New York City query location and all other Query Locations in the US |
| [benihana's], English (US) | 90210 | Find information about the Benihana restaurant in the 90210 zip code location (Beverly Hills, California) or go to the Benihana homepage. | http://www.benihana.com/locations/dallas-tx-da | **Other Vital** | The landing page is the official webpage for the Benihana restaurant in Dallas, Texas. Since the page is on the official Benihana website, it should be rated **Other Vital** for the 90210 zip code Query Location (Beverly Hills, California). |
| [benihana's], English (US) | Chicago | Find information about the Benihana restaurant in Chicago or go to the Benihana homepage. | http://www.benihana.com/locations/lombard-il-lb | **Appropriate Vital** | The landing page is the official webpage for the Benihana restaurant located in Lombard, Illinois, about 25 miles from the heart of Chicago. Because there are no Benihana restaurants located right in Chicago and the Chicago metro area easily extends 25 miles from downtown, this page should be rated **Appropriate Vital** for the Chicago Query Location. |
| [benihana's new york], English (US) | San Francisco | Find information about the Benihana restaurant in New York. | http://www.benihana.com/locations/newyorkwest-ny-we | **Appropriate Vital** | The landing page is the official webpage for the Benihana restaurant in New York City. Although the Query Location is San Francisco, the user specifically wants information about the Benihana restaurant in New York City. It should be rated **Appropriate Vital** for any Query Location in the US. |
| [benihana's new york], English (US) | Chicago | Find information about the Benihana restaurant in New York. | http://www.benihana.com/ | **Appropriate Vital** | The landing page is the official Benihana homepage. Although the query asks for the Benihana restaurant in New York, the official homepage of the Benihana restaurant chain should be rated **Appropriate Vital** for any Query Location in the US. |

| Query | Query Location | Likely User Intent | URL of the Landing Page | Rating | Explanation |
|---|---|---|---|---|---|
| [Outback Steakhouse], English (US) | Chicago | Find information about Outback Steakhouse restaurants in Chicago or go to the Outback homepage | http://www.yelp.com/biz/outback-steakhouse-chicago | **Relevant** | Outback Steakhouse is a chain of restaurants. For the Chicago Query Location, this Yelp landing page with information, a map, reviews, etc. for one of the Outback Restaurants in Chicago is **Relevant**. |
| [Outback Steakhouse], English (US) | San Francisco | Find information about Outback Steakhouse restaurants in San Francisco or go to the Outback homepage | http://www.yelp.com/biz/outback-steakhouse-chicago | **Off-Topic or Useless** | For the San Francisco Query Location, this Yelp landing page with information, a map, reviews, etc. for one of the Outback Restaurants in Chicago is **Off-Topic or Useless**. This page has no utility for San Francisco users. |
| [Outback Steakhouse], English (US) | Chicago | Find information about Outback Steakhouse restaurants in Chicago or go to the Outback homepage | http://www.outback.com/ | **Appropriate Vital** | The landing page is the official Outback Steakhouse homepage. It should be rated **Appropriate Vital** for any Query Location in the US. |
| [information about Bill Gates], English (US) | San Francisco | Find information about Bill Gates | http://en.wikipedia.org/wiki/Bill_Gates | **Useful** | Although this query has a Query Location, it is not associated with a location. This page about Bill Gates should be rated **Useful** for any Query Location in the US. |
| [arizona's rivers], English (US) | Chicago | Find information about the rivers in Arizona | http://geology.com/state-map/arizona.shtml | **Relevant** | Although this query has a Query Location, it is not associated with a location. This page with a map of the rivers in Arizona should be rated **Relevant** for any Query Location in the US. |
| [cabbage patch doll pictures], English (US) | Seattle | Find pictures of Cabbage Patch dolls | http://images.google.com/images?hl=en&q=cabbage%20patch%20doll&sourceid=navclient-ff&rlz=1B3GGGL_enUS321US306&um=1&ie=UTF-8&sa=N&tab=wi | **Useful** | Although this query has a Query Location, it is not associated with a location. This page with many images of Cabbage Patch dolls should be rated **Useful** for any Query Location in the US. |
| [name of Sarah Palin's book], English (US) | Atlanta | Find the name of the book written by Sarah Palin | http://news.yahoo.com/s/ap/20091002/ap_on_en_ot/us_books_palin_cover | **Useful** | Although this query has a Query Location, it is not associated with a location. This Yahoo News page has the title of the book, "Going Rogue", and should be rated **Useful** for any Query Location in the US. |

| Query | Query Location | Likely User Intent | URL of the Landing Page | Rating | Explanation |
|---|---|---|---|---|---|
| [susan boyle], English (US) | New York City | Watch a video or find information about Susan Boyle | http://www.youtube.com/watch?v=RxPZh4AnWyk | **Useful** | Although this query has a Query Location, it is not associated with a location. This YouTube video of Susan Boyle performing should be rated **Useful** for any Query Location in the US. |
| [buy going rogue online], English (US) | Miami | Purchase the book "Going Rogue" online | http://www.borders.com/online/store/TitleDetail?sku=0061939897 | **Useful** | Although this query has a Query Location, it is not associated with a location. Users in any Query Location in the US would find this Borders.com page to be **Useful**. |
| [the independent], English (US) | San Francisco | Go to the official homepage of The Independent, a popular music venue in San Francisco | http://www.theindependentsf.com/ | **Appropriate Vital** | Users in San Francisco have a different intent for this query than users in other locations because there is a popular music venue in San Francisco with this name. For the San Francisco Query Location, the landing page is **Appropriate Vital**. |
| [the independent], English (US) | New York City | Go to the official homepage of The Independent, a popular music venue in San Francisco or the official homepage of The Independent, the well-known and widely-read British newspaper | http://www.theindependentsf.com/ | **Relevant** | This query is not associated with a location for NYC users. For the NYC Query Location, this landing page is **Relevant** because it satisfies one of the common interpretations of the query for users in any Query Location outside the San Francisco area. |
| [the independent], English (US) | New York City | Go to the official homepage of The Independent, a popular music venue in San Francisco or the official homepage of The Independent, the well-known and widely-read British newspaper | www.independent.co.uk/ | **Appropriate Vital** | This query is not associated with a location for NYC users. For the NYC Query Location, this landing page is **Appropriate Vital** because the newspaper is the dominant interpretation outside San Francisco. |
| [the independent], English (US) | San Francisco | Go to the official homepage of The Independent, a popular music venue in San Francisco or the official homepage of The Independent, the well-known and widely-read British newspaper | www.independent.co.uk/ | **Useful** | The official homepage of the well-known and widely-read British newspaper is **Useful** for the San Francisco Query Location. |
| [Louie's 106], English (US) | New York City | Find information about or the homepage for Louie's 106, a restaurant in Austin, Texas | http://www.louies106.net/ | **Appropriate Vital** | There is only one Louie's 106 restaurant, and it is located in Austin, Texas. The homepage of this restaurant should be rated **Appropriate Vital** for any Query Location in the US. |

| Query | Query Location | Likely User Intent | URL of the Landing Page | Rating | Explanation |
|---|---|---|---|---|---|
| [DMV New York], English (US) | San Francisco | Go to the official homepage of the Department of Motor Vehicles in New York State | http://www.nydmv.state.ny.us/ | **Appropriate Vital** | The landing page is the official homepage of the Department of Motor Vehicles in New York State and is **Appropriate Vital** for any Query Location in the US. |
| [DMV New York], English (US) | San Francisco | Go to the official homepage of the Department of Motor Vehicles in New York State | http://dmv.ca.gov/ | **Off-Topic or Useless** | The landing page is the official homepage of the Department of Motor Vehicles in California. The DMV offices in New York and California are separate entities. The correct rating is **Off-Topic or Useless**. |
| [Museum of Modern Art], English (US) | San Francisco | Go to the official homepage of the Museum of Modern Art in San Francisco. | http://www.sfmoma.org/ | **Appropriate Vital** | There are two well-known museums in the US with this name. The landing page is the official homepage of the Museum of Modern Art in San Francisco. It is highly likely that the San Francisco Museum of Modern is the target of the query. The correct rating is **Appropriate Vital** for this Query Location. |
| [Museum of Modern Art], English (US) | San Francisco | Go to the official homepage of the Museum of Modern Art in San Francisco. | http://www.moma.org/ | **Useful or Relevant** | There are two well-known museums in the US with this name. The landing page is the official homepage of the Museum of Modern Art in New York City. It is highly likely that the San Francisco Museum of Modern Art is the target of the query instead, but it is possible that users in San Francisco are interested in the New York museum. The correct rating is **Useful** or **Relevant** for the San Francisco Query Location. |

| Query | Query Location | Likely User Intent | URL of the Landing Page | Rating | Explanation |
|---|---|---|---|---|---|
| [Museum of Modern Art], English (US) | Chicago | Go to the official homepage of the Museum of Modern Art. | http://www.sfmoma.org/<br><br>http://www.moma.org/ | **Useful** | There is no Museum of Modern Art in Chicago. Users in Chicago may be interested in either the San Francisco or New York Museum of Modern Art. Both of these official homepages should be rated **Useful** for the Chicago Query Location. |
| [Museum of Modern Art san francisco], English (US) | None | Go to the official homepage of the Museum of Modern Art in San Francisco. | http://www.sfmoma.org/ | **Appropriate Vital** | As specified in the query, the user is interested in the Museum of Modern Art in San Francisco. The landing page is the official homepage of the Museum of Modern Art in San Francisco and is **Appropriate Vital**. |
| [Museum of Modern Art new york], English (US) | None | Go to the official homepage of the Museum of Modern Art in New York. | http://www.sfmoma.org/ | **Off-Topic or Useless** | As specified in the query, the user is interested in the Museum of Modern Art in New York. The landing page is the official homepage of the Museum of Modern Art in San Francisco and is **Off-Topic or Useless**. |
| [Bar None restaurant], English (US) | San Francisco, CA | Find information about or the homepage for the Bar None restaurant/bar in San Francisco. | http://www.barnonenyc.com/ | **Other Vital** | There are many restaurants and bars with the name Bar None in the US. Some of them have the same parent company; others do not. The homepage for this Bar None in New York City should be rated **Other Vital**, since it is part of the same chain as the Bar None in San Francisco, but is not the restaurant the user in the San Francisco Query Location is looking for. |
| [Bar None restaurant], English (US) | San Francisco, CA | Find information about or the homepage for the Bar None restaurant/bar in San Francisco. | http://www.clubzone.com/c/2829/bar-none-nottingham | **Off-Topic or Useless** | The landing page is for a Bar None restaurant in Nottingham, England. This restaurant is unrelated to the Bar None chain of restaurants in the US and the landing page should be rated **Off-Topic or Useless.** This page has no utility for users in San Francisco. |

| Query | Query Location | Likely User Intent | URL of the Landing Page | Rating | Explanation |
|---|---|---|---|---|---|
| [Shear Bliss], English (US) | San Francisco, CA | Find information about or the homepage for the Shear Bliss beauty salon in San Francisco. | http://www.shearblissnyc.com/ | **Off-Topic or Useless** | There are Shear Bliss hair salons in multiple cities in the US. The landing page is for the Shear Bliss salon in New York. It should be rated **Off-Topic or Useless** for the San Francisco Query Location. These hair salons are not part of a chain and this page has no utility for users in the San Francisco Query Location. |
| [Walgreens], English (US) | None | Find information about Walgreen's pharmacies. | http://www.yelp.com/search?ns=1&rpp=10&find_loc=atlanta&find_desc=walgreens | **Slightly Relevant** | There is no Query Location. The Yelp page has lots of information on Walgreens pharmacies in the Atlanta area. It's not helpful to most users. |
| [Walgreens], English (US) | Atlanta, GA | Find information about Walgreen's pharmacies in Atlanta, GA | http://www.yelp.com/search?ns=1&rpp=10&find_loc=atlanta&find_desc=walgreens | **Useful** | The Query Location is Atlanta, GA. This Yelp page with lots of information about Walgreens pharmacies in the Atlanta area would be helpful for most users in the Query Location. |
| [Walgreens], English (US) | Atlanta, GA | Find information about Walgreen's pharmacies in Atlanta, GA | http://en.wikipedia.org/wiki/Walgreens | **Relevant** | Although the task has a Query Location and the user probably wants to find information about Walgreen's pharmacies in Atlanta, it is also possible that users in the Atlanta Query Location are looking for general information about the company. |
| [Walgreens Atlanta, Georgia], English (US) | Atlanta, GA | Find information about Walgreen's pharmacies in Atlanta, GA | http://en.wikipedia.org/wiki/Walgreens | **Off-Topic or Useless** | The Query Location is Atlanta, GA and the user has specified Atlanta in the query. The user definitely wants to find information about Walgreen's pharmacies in Atlanta. This page with general information about Walgreens is not helpful. |
| [mono], English (US) | Philadelphia, PA | Find information about the mononucleosis | http://www.myspace.com/monojp | **Useful** | Although this query has a Query Location, it is probably not associated with a location. The most likely user intent is to find information about the disease, mononucleosis. However, it is also possible that users in any Query Location are looking for information about the band, MONO. Since the landing page is the band's official MySpace page, it should be rated **Useful**. |

| Query | Query Location | Likely User Intent | URL of the Landing Page | Rating | Explanation |
|---|---|---|---|---|---|
| [mono], English (US) | Philadelphia, PA | Find information about the mononucleosis | http://www.webmd.com/a-to-z-guides/infectious-mononucleosis-topic-overview | **Useful** | Although this query has a Query Location, it is probably not associated with a location. The most likely user intent is to find information about the disease, mononucleosis. The landing page is a highly informative page on an authoritative medical website. |
| [Waterford], English (US) | Gainesville, FL | Purchase Waterford china or crystal, or go to the official Waterford homepage | http://www.waterford.com/ | **Appropriate Vital** | The crystal and china company at http://www.waterford.com/ is the dominant interpretation for the query. This query has a Query Location, but it might not be associated with a location. Although there are businesses with Waterford in their name in Gainesville, Florida, The official homepage for Waterford should be rated **Appropriate Vital**. |
| [Waterford], English (US) | Gainesville, FL | Purchase Waterford china or crystal, or go to the official Waterford homepage | http://www.waterfordtitle.com/ | **Useful** | Although the Waterford china company is the dominant interpretation for the query, it is very possible that users in the Gainesville Query Location are looking for local businesses with Waterford in their name. The official homepage of Waterford Title company in Gainesville is **Useful**. |
| [Waterford], English (US) | Gainesville, FL | Purchase Waterford china or crystal, or go to the official Waterford homepage | http://www.waterfordbank.com/ | **Slightly Relevant** | Although the Waterford china company is the dominant interpretation for the query, there is a slight possibility that users in the Gainesville Query Location are looking for local businesses in other locations with Waterford in their name. The official homepage of Waterford Bank in Waterford, Ohio should be rated **Slightly Relevant.** |

# Part 3: 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.)

| | **[John McCain], English (US)** |
|---|---|
| ***Query Description*** | *John McCain is a United States Senator. He is a very well-known politician and there are many pages on the Web about him.* |
| ***Likely User Intent*** | ***Know*** *– Users want information or news about John McCain*<br>***Go*** *– Users want to go to an official page for John McCain* |
| **Vital** | John McCain's official government Senate homepage: http://mccain.senate.gov/<br>John McCain's official MySpace page: http://www.myspace.com/johnmccain<br>John McCain's official YouTube page: http://www.youtube.com/johnmccain |
| **Useful** – helpful for most users | Quality pages with biographical or good general information, such as this Wikipedia page about Senator John McCain: http://en.wikipedia.org/wiki/John_McCain<br>An article with biographical information about John McCain and his complete Senate voting record at http://projects.washingtonpost.com/congress/members/m000303/ |
| **Relevant** – helpful for many or some users | Quality pages with biographical or good general information about Senator John McCain's ***father***, who is also named John McCain: http://en.wikipedia.org/wiki/John_S._McCain,_Jr. **Slightly Relevant** is also acceptable.<br>A timely article about Senator John McCain.<br>A video with Senator John McCain in it, such as http://www.youtube.com/watch?v=53caXQKTs9Y<br>A page on which to buy a book written by Senator John McCain, such as http://www.amazon.com/Worth-Fighting-John-S-McCain/dp/0375505423 |
| **Slightly Relevant** – helpful for few users | A page about a tax bill proposed by Senator John McCain and another senator in 2003: http://www.nationalcenter.org/TSR102103.html<br>A page of photos of the USS John S. McCain, a naval destroyer named after John McCain's grandfather at http://www.navsource.org/archives/05/01056.htm<br>An article about an ordinary person named John McCain. |

| | [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*<br>***Go*** *– Users want to go to an official page for Nicole Kidman* |
| **Vital** | Nicole Kidman's official homepage, if one exists. Please be aware that some unofficial sites for celebrities may claim to be official. |
| **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.<br>A very high quality personal fan page<br>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<br>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. |

| | [Erica Hill], English (US) |
|---|---|
| ***Query Description*** | *Erica Hill is a news anchor for The Early Show on CBS. She previously worked on the following CNN shows: "Anderson Cooper 360", "CNN Headline News", and "Prime News". Although she is a fairly well-known news anchor, you would not expect to find as many high quality pages about her on the Web as you would for Senator John McCain or Nicole Kidman.*<br>*The first name "Erica" and the last name "Hill" are fairly common names. You would expect to find other people named Erica Hill in the world.* |
| ***Likely User Intent*** | ***Know*** *– Users want information or news about Erica Hill, the CBS news anchor*<br>***Go*** *– Users want to go to an official page for Erica Hill, the CBS news anchor* |
| **Vital** | Erica Hill's page on the CBS website:<br>http://www.cbsnews.com/stories/2008/09/22/earlyshow/bios/main4468573.shtml |
| **Useful** – helpful for most users | Quality pages with biographical or good general information about Erica Hill, the CBS news anchor, such as http://en.wikipedia.org/wiki/Erica_Hill |
| **Relevant** – helpful for many or some users | Homepage of an Erica Hill fansite: http://www.ericahill.org/. Since her biography on the page hasn't been updated, **Slightly Relevant** is also acceptable.<br>Short article about Erica Hill:<br>http://blogs.orlandosentinel.com/entertainment_tv_tvblog/2010/01/erica-hill-moving-from-cnn-to-news-reader-spot-on-cbs-early-show.html<br>Helpful page about a different person named "Erica Hill", who is less well-known and would be of interest to some or few people. **Slightly Relevant** is also acceptable. |
| **Slightly Relevant** – helpful for few users | Lower quality pages about the CBS news anchor, such as<br>http://www.biocrawler.com/encyclopedia/Erica_Hill<br>Outdated pages about the CBS news anchor, such as<br>http://www.cnn.com/CNN/Programs/anderson.cooper.360/blog/2008/01/erica-hill-cometh.html |
| **Off-Topic or Useless** – helpful for very few or no users | Pages with the words "Erica" or "Hill" scattered on them, such as this softball box score page that mentions players named *Erica* Douglas and Sam *Hill*,<br>http://gomajors.com/news/2009/7/9/GEN_0709093159.aspx?path=general |

| | [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*<br>***Do*** *– Users want to purchase products manufactured by the company*<br>***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/<br>Pages that sell, distribute, or review multiple A.O. Smith products. **Relevant** may also be acceptable, depending on how helpful the page is.<br>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/<br>A current news article about A.O. Smith<br>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<br>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<br>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 | Article about a singer named Elliott Smith, who was scheduled to perform at a dance called the "A&O Ball".<br>http://media.www.dailynorthwestern.com/media/storage/paper853/news/2002/05/02/Campus/Ao.Ball.Signs.On.A.Second.Headliner-1909814.shtml |

| | [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*<br>***Do*** *– Users want to make a purchase*<br>***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<br>A directory page with contact information: http://www.zvents.com/sunnyvale-ca/venues/show/125217-for-other-living-things<br>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<br>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*<br>***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<br>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 these 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<br>Wikipedia article about Perkins restaurant<br>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<br>Outdated news articles about the Perkins restaurant<br>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*<br>***Know*** *– Users want information (reviews, specifications, features, etc.) about the iPhone*<br>***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/<br>The Apple Store page on the Apple website: http://store.apple.com/us<br>The iPhone page of the Apple Store: http://store.apple.com/us/browse/home/shop_iphone/family/iphone?mco=OTY2ODA2OQ<br>High quality sites that review or provide comprehensive information on the iPhone, such as http://www.cnet.com/apple-iphone.html<br>The AT&T page where users can purchase the iPhone: http://www.att.com/wireless/iphone/<br>The Apple iPhone discussion board: http://discussions.apple.com/category.jspa?categoryID=201 |
| **Relevant** – helpful for many or some users | Page with many iPhone many accessories for sale<br>A timely article about the iPhone<br>A helpful video about the iPhone, such as http://www.youtube.com/watch?v=IpQ9RESJnWM<br>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<br>Outdated article on the iPhone<br>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*<br>***Know*** *– Users want information (reviews, specifications, features, etc.) about the Honda Pilot*<br>***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/<br>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<br>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.<br>A relatively short article about the current year's Honda Pilot<br>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<br>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*<br>***Know*** *- Users want general information about Nevada or travel and tourism information*<br>***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/<br>High quality, comprehensive pages about Nevada: http://en.wikipedia.org/wiki/Nevada<br>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/<br>Pages with facts about Nevada: http://www.leg.state.nv.us/general/FACTS.cfm and http://www.nv.gov/new_KidsHomework.htm<br>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.<br>Homepage of the Nevada Republican Party: http://www.nevadagop.org/<br>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*<br>***Know*** *– Users want travel and tourism information or general information about Chicago*<br>***Go*** *– Users want to navigate to the official Chicago city government website*<br><br>*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<br>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<br>An excellent blog or collection of personal information, which would be helpful to someone visiting the city, such as http://www.gochicagocard.com/blog/<br>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<br>A high quality map of the city, such as http://travel.yahoo.com/p-map-191501928-map_of_chicago_il-i<br>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/.<br>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.<br>YouTube Channel page of Chicago's official tourism site: http://www.youtube.com/user/explorechicago<br>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<br>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/<br>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/<br>Video of the Blue Brothers performing the song, "Sweet Home Chicago", http://www.youtube.com/watch?v=Tlou_2lMLAc |

Note: Major cosmopolitan cities are preferred targets for spammers, especially hotel affiliates. Such results should be flagged as **Spam**, even if they are related to the query and helpful to users. For example, a hotel affiliate page with a list of Chicago hotels may be assigned a rating **Relevant**, but also receive a **Spam** flag.

| | [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*<br>***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/<br>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)<br>Wikipedia page about the White House: http://en.wikipedia.org/wiki/White_House<br>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/<br>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<br>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<br>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*<br>***Know*** *– Users want information about Adobe Reader*<br>***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<br>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=tIx-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<br>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<br>Wikipedia page on eye surgery that discusses many types of eye surgery, including laser retina surgery: http://en.wikipedia.org/wiki/Eye_surgery<br>Yahoo! Answers page on the topic of the query: http://au.answers.yahoo.com/answers2/frontend.php/question?qid=20070724160757AAHmLJy<br>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/<br>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*<br>***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<br>An article that discusses HTML 5, a major upgrade to HTML, but doesn't 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<br>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<br>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*<br>***Know*** *– Users want to learn how to enable JavaScript in Internet Explorer*<br>***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*<br>***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.europeanmuseumguide.com/museumInfo.php?museumid=115 |
| **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<br>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*<br>***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.<br>http://www.americanexpress.com/<br>http://www.usa.visa.com/personal/<br>http://www.mastercard.com/us/gateway.html<br>http://www.citicards.com/cards/wv/home.do<br>http://www.discovercard.com/<br><br>Pages on reputable sites that offer credit card comparisons, such as:<br>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<br>Pages on reputable sites with information about credit cards, such as<br>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<br>http://www.unionplus.org/benefits/money/card.cfm<br>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*<br>***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):<br>http://www.citibank.com/<br>https://www.bankofamerica.com/<br>http://www.chase.com/<br>Website with links to banks in the United States, organized by state:<br>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<br>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.<br>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*<br>***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):<br>http://www.schwinnbike.com/usa/eng/<br>http://www.trekbikes.com/us/en/<br>http://www.specialized.com/us/en/bc/home.jsp<br>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<br>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/<br>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*<br>***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):<br>http://www.orbitz.com/<br>http://www.expedia.com/<br>http://www.travelocity.com/<br>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):<br>http://www.united.com/<br>http://www.aa.com/<br>http://www.usairways.com/<br>https://www.southwest.com/<br>The Federal Aviation Administration's page of links to US airline companies:<br>http://www.fly.faa.gov/FAQ/Airline_Links/airline_links.jsp<br>Wikipedia page with links to airlines that operate in the United States:<br>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.<br>http://www.alitalia.com/us_en/?no<br>http://www.jal.co.jp/en/<br>Wikipedia page that contains a list of airlines, organized by continent and country:<br>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,<br>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*<br>***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):<br>http://www.radisson.com/<br>http://www.hilton.com/<br>http://www.marriott.com/<br><br>Homepages of online hotel and travel companies that allow users to make reservations with many different hotel chains:<br>http://www.hotels.com/ http://www.orbitz.com/<br>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.<br>http://www.bedandbreakfast.com/<br>http://www.bbonline.com/<br><br>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*<br>***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.talkingcities.co.uk/london_pages/shopping_womensfashion.htm Such pages might include maps, pictures, addresses, descriptive information, price ranges, store hours, etc.<br>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/<br>IBM Spain webpage: http://www.ibm.com/es/es/<br>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/<br>Wikipedia article about IBM India: http://en.wikipedia.org/wiki/IBM_India<br>2008 news article about IBM India: http://www.tradingmarkets.com/.site/news/Stock%20News/1930596/ |
| **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/<br>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.<br>Wikipedia article about the football magazine: http://en.wikipedia.org/wiki/Match_magazine<br>Wikipedia article about the dating company: http://en.wikipedia.org/wiki/Match.com<br>Wikipedia article about matches that people use to light a fire: http://en.wikipedia.org/wiki/Match<br>"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/<br>France Sephora homepage: http://www.sephora.fr/<br>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<br>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.<br>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/<br>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.<br>Women's page: http://femmes.orange.fr/<br>News page: http://actu.orange.fr/<br>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 4: 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 ref er 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 aren’t 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.

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 only 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 can’t 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 won't be noticed.
It may be formatted in a very, very small font size (e.g., 1-point) so that it won't 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 don't 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**”.<br>2. Click on “**Options**”.<br>3. Click on “**Content**” or ”**Web Features**”.<br>4. To disable JavaScript, make sure the ”**Enable**” box is not unchecked.<br>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**”.<br>2. Click on “**Page Source**”.<br>or<br>1. Right click on the page.<br>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.<br>2. Click "**This Frame**".<br>3. Click "**View Frame Info**".<br>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. |

27

| **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.

#### 4.1.2 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<br>Spam URL: http://www.nationmaster.com/encyclopedia/magnetite | Screenshot Example |
| DMOZ Example | DMOZ URL: http://www.dmoz.org/Computers/Security/<br>Spam URL: http://contentguarder.com | Screenshot Example |

#### 4.1.3 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.

#### 4.1.4 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. Here is an example: http://www.limosnationwide.com/. This page 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/ | Screenshot Example |
| California page URL | http://www.hair-removal-hair-laser.com/ca.html | |
| Florida page URL | http://www.hair-removal-hair-laser.com/fl.html | |
| 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 | |

### 4.1.5 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 | |

### 4.1.6 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.

### 4.1.7 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 doesn’t 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 don’t 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<br>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 only 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 only to make money, 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 can't 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 only 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 no additional information and does not try to help 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.** Affiliate pages that include original content in addition to the affiliate link are not 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 whose in Section 3.3.1.

### 5.1.2 Not all Affiliates are Thin

Some affiliates are created to help users. Anyone can become an "affiliate" of merchant sites such as Amazon and link to Amazon products. Webmasters may do this to show products they like or to help users find a good deal.

For example, if the affiliate offers price comparison functionality, or displays product reviews, recipes, lyrics, etc., it is usually not a thin affiliate, and, therefore, not spam. Some websites that offer price comparisons or other helpful shopping features, in addition to the affiliate link, are:

| | | |
|---|---|---|
| http://www.shopping.com/ | http://www.nextag.com/ | http://www.kelkoo.co.uk/ |
| http://www.pricegrabber.com/ | http://www.bizrate.com/ | http://www.ciao.it/ |
| http://www.dealtime.com/ | http://www.mysimon.com/ | http://www.dooyoo.it/ |

### 5.1.3 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 don't provide a lot of good information, but they are original. You won't 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<br>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 **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 don't 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).

## 8.0 Conclusion

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 don't 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 5: 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

There are two different ways to access the EWOQ URL rating interface:

1) Rater Hub: Click on the "Start Rating Now" link in the upper left corner of the Rater Hub homepage. This link will take you to your Rater Homepage.

2) Go to this link - https://www.google.com/evaluation/search/rating/home

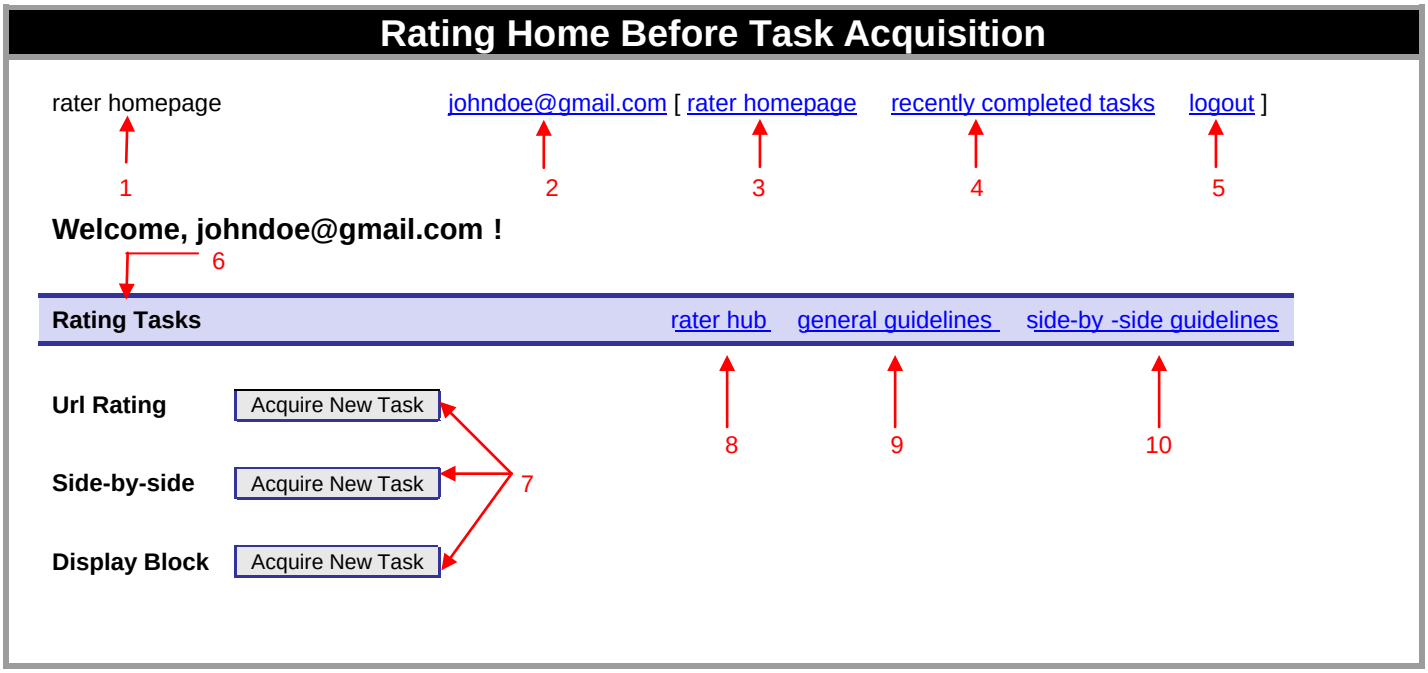
You will supply your Gmail user ID and password for authentication.

## 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)

## 4.0 Rating Home Screenshots




**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.

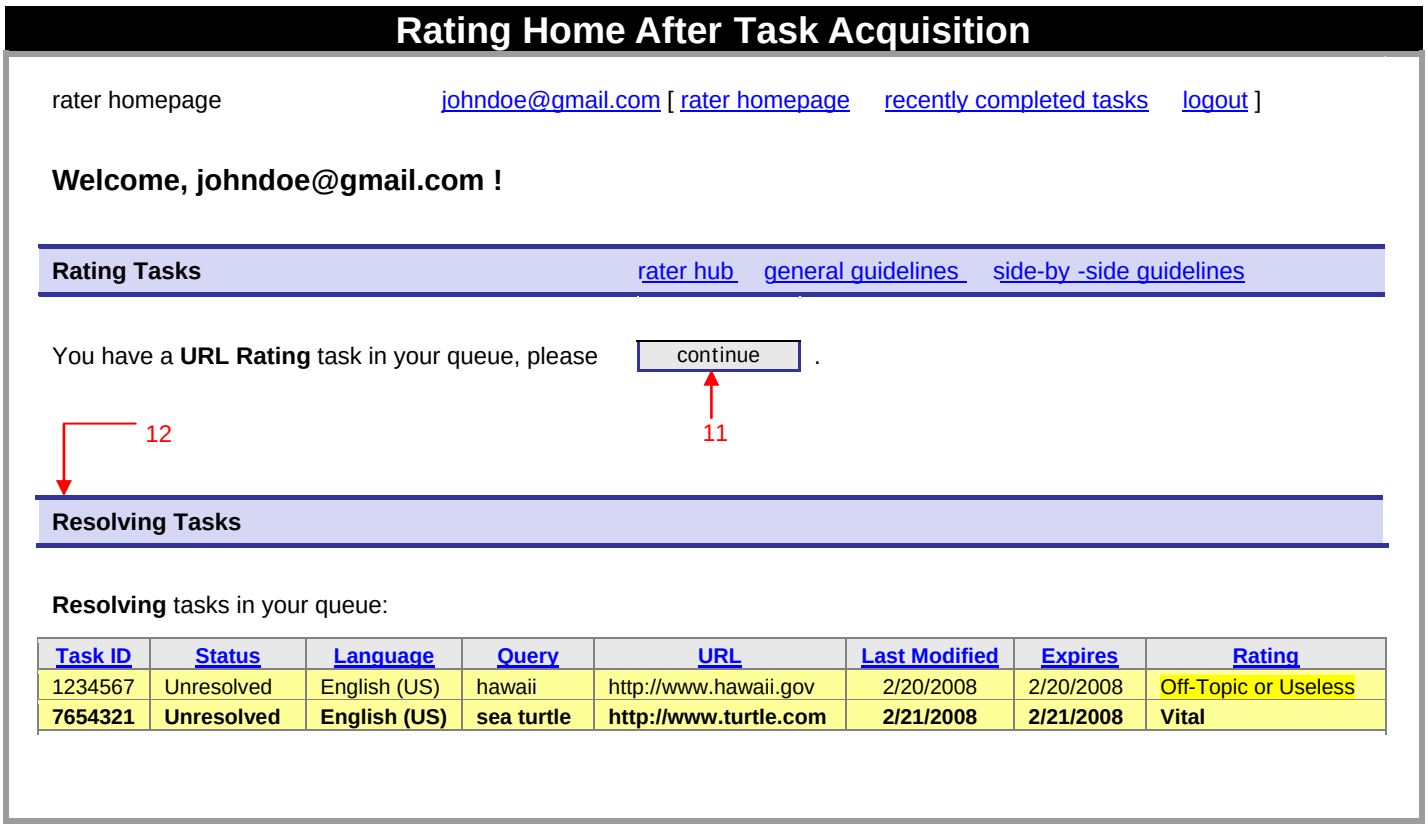
7. **Acquire New Task**
   Click this button to acquire a new task. The new Rater Homepage will allow you to acquire only one task from one of the project types displayed on your Rater Homepage. When tasks are available, you will see buttons for up to three different project types displayed. Please click 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. **rater hub**
   Click on this link to access the Rater Hub. This is the primary resource page, which supports the quality-rating program. This page contains Frequently Asked Questions (FAQs), News & Updates, Helpful Suggestions, Rater Training Tools, etc.

9. **general guidelines**
   Click on this link to read the "General Guidelines".

10. **side-by-side guidelines**
    Click on this link to read the "Side-by-Side Rating Guidelines".

## Rating Home After Task Acquisition




**The red numbers represent the following:**

11. **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.

12. **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

rater homepage (1) rating task (2) johndoe@gmail.com [ rater homepage (3) (4) recently completed tasks (5) logout (6) ]

**Rating Task - icq** (7)

[ search results: google ] release task (8) general guidelines (9) rater hub (10)

| | |
|---|---|
| (11) **Query** | Icq |
| (12) **Query Description** | *This field is present only if there is a description for the query.* |
| (13) **URL** | http://www.mobicq.info/ |
| (14) **Task Location** | Ukraine (UA) |
| (15) **Task Language** | Ukrainian |
| (16) **Other Acceptable Languages** | Russian |

## URL RATING

| | |
|---|---|
| (17) **Rating** Choose one | **Vital** (choose one geographical location) (18)<br>**Appropriate Vital**<br>**International Vital**<br>**Other Vital**<br>**Useful**<br>**Relevant**<br>**Slightly Relevant**<br>**Off-Topic or Useless**<br>**Unratable** (19)<br>Didn't Load<br>Foreign Language |
| (20) **Landing Page Language** Choose one | Ukrainian<br>Russian<br>English<br>Foreign Language<br>None of the above |
| (21) **Spam** Choose one | Not Spam<br>Maybe Spam<br>Spam |
| (22) **Other Flags** Choose all that apply | Pornography<br>Malicious |
| (23) **Comment** | |

Cancel (24) Save Draft (25) Submit (26)

Proprietary and Confidential – Copyright 2011

**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 **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. **rater hub**
    Click on this link to go to the Rater Hub.

11. **Query**
    Make sure you understand the query. Please research the query to learn about its meaning and the user intent behind it.

12. **Query Description**
    This field is present only if there is a description for the query. Currently, only a minority of queries carry a description. Query descriptions are entered by administrators. These descriptions may advise you that the query has been known to bring up a particular type of result and offer tips on how to rate this type of result. Some descriptions tell you which interpretation of the query should have the most weight. You may not agree with the query description. If so, be sure to make a comment explaining why you disagree.

13. **URL**
    This is the URL that you will click to view the landing page.

14. **Task Location**
    The location associated with the task.

15. **Task Language**
    The language associated with the task.

16. **Other Acceptable Languages**
    Please refer to the "Rating Guidelines" for information on acceptable languages.

17. **Rating**
    Please refer to the "Rating Guidelines" for information on each rating category.

18. **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.

19. **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.

20. **Landing Page Language**
    Please refer to the "Rating Guidelines" for information on selecting the landing page language.

21. **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**.

**22. Other Flags**

Please choose **Pornography** and/or flags when appropriate.

**23. 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.

**24. 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** ."

**25. Save Draft**

This button is only available to people taking the rating exam. Exam takers may use "Save Draft" to retain ratings on tasks they want to revisit before submitting their exam.

**26. 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

rater homepage  rating task  johndoe@gmail.com [ rater homepage  recently completed tasks  logout ]

## Rating Task - icq

[ search results: google ]  release task  general guidelines  rater hub

| | |
|---|---|
| **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* | |
| 2 → 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 | |

## 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.

Proprietary and Confidential – Copyright 2011

### 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 or other rating information from the Rater Hub. 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.”

Proprietary and Confidential – Copyright 2011

# Part 6: 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 that you are comfortable with. 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. T he car, the planet, and the chemical are common interpretations for [mercury], English (US).

**Minor Interpretations**: Sometimes you will find less common interpretations. Mercury Ma rine InsuranceCompany 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 Pr eside 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 aspecific 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**:

**Famous vs. Common**: Queries for famous people such as [Madonna] have obvious dominant interpretations and can have **Vital** pages. Queries for ordinary people with common names, such as [bob smith] cannot.

**Multiple Personal Pages**: Some famous people have multiple "official" personal pages. All such pages should be rated **Vital**. Use your judgment to decide if a page is "official".

**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 doesn't match the task location, and a better version exists. (If a better version for the task location doesn't 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 don't.

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 no 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 can't 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 aren't 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 can't 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 porn result 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 7: 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 can't 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 several 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 only 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 only to make money, 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 can't see how they are making money.

**Thin Affiliates**: A thin affiliate is a website that earns money from affiliate commissions. It exists only to make money. The spammer shows content from other "real" merchant or travel sites, such as Amazon or Orbitz. When users click on links to buy products or make reservations, they are redirected to the "real" merchant page (e.g. Amazon or Orbitz).

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. Affiliate pages that include original, helpful content in addition to the affiliate link are not 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 don't provide a lot of good information, but they are original. You won't 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.