

- 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 9
 - 2.3 Queries with Multiple Meanings 9
 - 2.4 Timeliness 9
 - 2.5 Classification of User Intent: Action, Information, and Navigation – “Do-Know-Go” 10
 - 2.5.1 Action Queries – “Do” 10
 - 2.5.2 Information Queries – “Know” 11
 - 2.5.3 Navigation Queries – “Go” 12
 - 3.0 The Language of the Landing Page 13**
 - 4.0 The Rating Scale 15**
 - 4.1 Vital 15**
 - 4.1.1 Examples of English (US) Navigation Queries with Vital Pages for the Task Location 15
 - 4.1.2 Examples of Entity Queries with Vital Pages 16
 - 4.1.3 Vital Pages for People Queries 16
 - 4.1.4 Other Important Vital Concepts 18
 - 4.1.5 Vital Pages and Geographic Location 19
 - 4.2 Useful 23**
 - 4.2.1 Examples of Useful Pages 23
 - 4.3 Relevant 25**
 - 4.3.1 Examples of Relevant Pages 25
 - 4.4 Slightly Relevant 26**
 - 4.4.1 Examples of Slightly Relevant Pages 26
 - 4.5 Off-Topic 26**
 - 4.5.1 Examples of Off-Topic Pages 27
 - 4.6 Unratable 27**
 - 4.6.1 Unratable: Didn’t Load 28
 - 4.6.2 Unratable: Foreign Language 32
 - 4.6.3 Unratable: Meaningless Query 32

4.6.4	Unratable: Other.....	34
5.0	Rating: From User Intent to Assigning a Rating.....	35
5.1	User Intent and Page Utility.....	35
5.2	Multiple Interpretations.....	36
5.3	Common Rating Problems.....	37
5.3.1	Dictionary or Encyclopedia Results.....	37
5.3.2	Action vs. Information Intent.....	38
5.3.3	Queries That Ask for a List.....	38
5.3.4	Misspelled and Mistyped Queries.....	39
5.3.5	URL Queries.....	41
5.3.6	Search Engine Result Pages.....	43
5.3.7	Video Landing Pages.....	44
5.4	Location is Important.....	46
6.0	Flags.....	48
6.1	Spam Flag.....	48
6.2	Pornography Flag.....	48
6.2.1	Clear Non-Porn Intent.....	48
6.2.2	Possible Porn Intent.....	49
6.2.3	Clear Porn Intent.....	49
6.3	Malicious Flag.....	50
6.4	Compatibility between Ratings and Flags.....	50
Part 2: Rating Examples		51
1.0	Named Entity Queries.....	51
2.0	Action Queries.....	59
3.0	Information Queries.....	61
4.0	Queries that Ask for a List.....	64
5.0	Rating Examples for Task Locations other than English (US).....	68
Part 3: Webspam Guidelines		70
1.0	What is Webspam ?.....	70
1.1	The Relationship between Ratings and Spam.....	70
1.2	Why do Spammers Create Spam Pages ?.....	70
1.3	When to Check for Spam.....	70
2.0	Choosing a Browser.....	71
3.0	Looking for Technical Signals.....	71
3.1	Hidden Text and Hidden Links.....	71
3.1.1	Apply Ctrl-A to the Landing Page.....	72

3.1.2	Disable CSS	72
3.1.3	Disable JavaScript	72
3.1.4	View the Source Code	73
3.1.5	Look Outside the Normal Viewing Area	73
3.2	Keyword Stuffing	74
3.2.1	Keyword Stuffing in the URL	74
3.3	Sneaky Redirects	75
3.3.1	Using “Whois”	75
3.4	Cloaking	75
3.4.1	JavaScript Redirects	76
3.4.2	100% Frame	76
4.0	Helpful Webpages vs. Spam Webpages	76
4.1	Pages with Copied Content and PPC Ads	77
4.1.2	Copied Text and PPC Ads	77
4.1.3	Feeds and PPC Ads	77
4.1.4	Doorway Pages	77
4.1.5	Templates and Other Computer-Generated Pages	78
4.1.6	Copied Message Boards	78
4.1.7	Recognizing Copied Content	78
4.2	Fake Search Pages with PPC Ads	79
4.3	Fake Blogs with PPC Ads	79
4.4	Fake Message Boards with PPC Ads	79
4.5	Copied Content that is NOT Spam	79
5.0	Commercial Intent	80
5.1	Thin Affiliates	80
5.1.1	Recognizing Thin Affiliates	80
5.1.2	Not all Affiliates are Thin	80
5.1.3	Recognizing True Merchants	81
5.2	Pure PPC Pages	81
5.3	Parked (Expired) Domains	81
5.4	Pages with Unhelpful Content and PPC Ads	82
6.0	Spam and the Resolving Stage	83
7.0	Conclusion	83

Part 4: Using EWOQ 84

1.0	Introduction	84
2.0	Accessing the EWOQ Rating Interface	84
3.0	Rating	84
4.0	Rating Home Screenshots	85

5.0 Resolving Tasks (Re-rating Unresolved Tasks) / Moderators	90
6.0 Commenting Etiquette	92
Part 5: Quick Guide to URL Rating	94
Part 6: Quick Guide to Webspam Recognition	97

Part 1: Rating Guidelines

1.0 Welcome to the Search Quality Rating Program!

As a Search Quality Rater, you will evaluate the quality of search engine results by working on many different types of rating projects. These guidelines cover just one type of search quality rating – Query-Page Evaluation, which we usually refer to as URL rating.

Please take the time to carefully read through these guidelines. Once you have become very comfortable with the concepts presented, 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:

- Make sure you understand the query and the intent of the user who typed the query. Occasionally, you may feel confident that you already understand the query; however, most queries require some research. Doing research will help you to understand both the meaning of the query and the user intent. This knowledge will give you an understanding of the kinds of search results that would be helpful to the user who typed the query.
- Examine the URL displayed in the task.
- Click on the URL to evaluate the landing page.

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. As a Search Quality Rater, you should be familiar with the most popular search engines for the task location you are assigned to.

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].

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 the user sees 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 U.S., 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. We refer to other pages on a website as **webpages** or **subpages**.

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: 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.

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 as the user intent.

Topic: The topic of the query is the focus or subject of the query; it is what the query is about. The user typing the query wants 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 the user. 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
<i>Vital</i>	A special rating category used for official webpages.
<i>Useful</i>	A page that is very helpful for most users.
<i>Relevant</i>	A page that is helpful for many users.
<i>Slightly Relevant</i>	A page that is not very helpful for most users, but is somewhat related to the query.
<i>Off-Topic</i>	A page that is not helpful for most users. It is not related to the topic of the query.
<i>Unratable</i>	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 rating 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. As a rater, you will imagine the user clicking on the URL in the task and decide if the landing page is a good search result for him.

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 Internet.**

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

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

Raters 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.

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. 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

<http://encarta.msn.com/>: Microsoft's Encarta encyclopedia

Online dictionaries:

<http://dictionary.reference.com/>

<http://www.wordwebonline.com/>

<http://encarta.msn.com/encnet/features/dictionary/dictionaryhome.aspx>

<http://www.thefreedictionary.com/>

Online thesauruses:

<http://thesaurus.reference.com/>

http://encarta.msn.com/thesaurus_/thesaurus.html

Translation tools:

<http://babelfish.altavista.com/>

<http://www.wordreference.com/>

If the query is meaningless even after doing research, you will assign “**Unratable: Meaningless Query**”. This rating will be discussed in Section 4.6.3.

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 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	Common User Intent
[Fedex], English (US)	Track a package or find a Federal Express location
[calendar], English (US)	Find, customize, and print a calendar for the current month or year Find a calendar that displays holidays. For example, the user may want to know what month and day Easter falls on in the current year. 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	User Intent in the Task Location
[football], English (US)	American football played with a brown oval ball	Recent game scores, game schedules, pictures, team information, etc. for American football in the US.
[football], English (UK)	The game Americans call soccer and that is played with a round ball	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, the fruit, or the music company. We will call these possible meanings **query interpretations**.

Dominant Interpretation: Sometimes, there is one query interpretation which is what most users have in mind. We will call this meaning the *dominant interpretation*. 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 people might want results related to these interpretations.

Minor Interpretations: Sometimes you will find less common interpretations. 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 5.2, we will discuss in detail how to rate pages when the query has multiple interpretations.

2.4 Timeliness

A query may be interpreted differently at different points in time. In 1994, the user who typed [President Bush], English (US) was looking for information on President George H.W. Bush. In 2007, his son George W. Bush is the more likely interpretation. You should always rate according to the current interpretation, unless the query clearly specifies otherwise.

In addition, for some queries, recent, timely results are extremely important. Here are some examples:

- for [academy awards], English (US), the user is looking for the most recent nominations or awards
- for [firefox], English (US), the user is looking for the latest version (unless he specifies an older version)
- for [congressional election results], English (US), the user is looking for results from the most recent election
- for [BCS standings], English (US), the user wants the current week's college football standings

2.5 Classification of User Intent: Action, Information, and Navigation – “Do-Know-Go”

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

- **Action intent** - The user is trying 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: the user wants to do something.
- **Information intent** - The user is trying to find information. These are “**know**” queries: the user wants to know something.
- **Navigation intent** - The user is trying to navigate to a website or webpage. These are “**go**” queries: the user wants to go to a webpage or website.

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.

2.5.1 Action Queries – “Do”

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

The user wants 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.

The only high utility pages for an action query are pages which allow the user to do the activity.

Query	User Intent	URL of High Utility Page	Description of High Utility 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 the user can take
[Beatles poster], English (US)	Purchase a Beatles poster	http://www.allposters.com/-sp/-Posters_i317216_.htm	Page on which to 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

Query	User Intent	URL of High Utility Page	Description of High Utility Page
[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
[calendar], English (US)	Use an online calendar or customize and print a calendar	http://www.timeanddate.com/calendar/ http://www.calendar.yahoo.com http://www.google.com/calendar	Page with current year calendar, which can be customized and printed Pages on which to create and maintain online calendars
[calculate running pace], English (US)	Calculate running pace online	http://www.coolrunning.com/engine/4/4_1/96.shtml	Page with running pace calculator
[text twist], English (US)	Play TextTwist online or download the game	http://get.games.yahoo.com/proddesc?gamekey=texttwist http://www.shockwave.com/gamelandin/g/texttwist.jsp	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://education.yahoo.com/reference/dict_en_es/	Pages on which to translate words between Spanish and English

2.5.2 Information Queries – “Know”

An **information query** seeks information on a topic. The user wants to **know** something; the goal is to find information.

Query	User Intent	URL of High Utility Page	Description of High Utility Page
[Switzerland], English (US)	Find travel and tourism information for planning a vacation or holiday, or find information about the Swiss geography, languages, economy	http://www.lonelyplanet.com/worldguide/destinations/europe/switzerland/ https://www.cia.gov/cia/publications/factbook/geos/sz.html	Travel guide on Switzerland Informative CIA 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	U.S. 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.solutions.uiuc.edu/content.cfm?series=1&item=53&Parents=0%7C5%7C20	Page on a reputable university website with this information

2.5.3 Navigation Queries – “Go”

The intent of a **navigation query** is to locate a specific webpage. The user has a single webpage or website in mind. This single webpage is called the **target** of the query. The user wants to **go** to the target page. Here are some examples:

Query	User Intent	URL of the Target Page	Description of the Target Page
[ibm], English (US)	Navigate to the IBM homepage	http://www.ibm.com/	Official homepage of the IBM Corporation
[youtube], English (US)	Navigate to the YouTube homepage	http://www.youtube.com/	Office homepage of YouTube
[ebay], Italian (IT)	Navigate to the Italian eBay homepage	http://www.ebay.it/	Official homepage of eBay Italy

Please note that sometimes we can think of a query as being a navigation query or an action query, because it is likely that the user wants to **go** to a specific webpage to **do** something. Here is an example:

Query	User Intent	URL of the Target Page	Description of the Target Page
[download firefox], English (US)	Navigate to the Mozilla Firefox download webpage to download the Firefox browser	http://www.mozilla.com/en-US/products/download.html?product=firefox-3.0.1&os=win&lang=en-US	Official Firefox browser download webpage

3.0 The Language of the Landing Page

As a rater assigned to a particular task language and task location, you are expected to be able 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 classified as one of the following:

- **Task Language:** The language to which you have been assigned
- **Acceptable Languages:** 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. Many task locations have no additional acceptable languages.
- **English**
- **Foreign Language:** A language other than the task language, an acceptable language, or English
- **None of the above:** Classification that is appropriate for pages which have no language or which do not load in a way that the language can be evaluated. For example, blank pages, pages with images only, or pages with severe encoding errors should get this classification.

Use the flag that corresponds to your **task language** when:

- The page is entirely in the task language.
- The page has several languages, but there is enough content in the task language to allow you to understand the page.
- The page is on a task language website and the main features of the page (such as navigation, search, menus, links, etc.) are in the task language.
- The primary content of the page is an image or other content (such as a dance video) which can be understood or enjoyed WITHOUT understanding the language, and the page is on a task language website. Please see Section 5.3.7 for more information about rating pages with video pages which have some foreign language components.

Use the flag that corresponds to the appropriate **acceptable language** when:

- The page is entirely in the acceptable language.
- The page has enough content in the acceptable language to allow you to understand the page.

Use the **English** flag when:

- The page is entirely in English.
- The page has enough content in English to allow you to understand the page.

Use the **Foreign Language** flag when:

- The language of the landing page (or the language spoken or sung in a video on the landing page) is in a language other than the task language, an acceptable language, or English. The page may contain a mix of foreign languages or it may be in just one foreign language.

Use the **None of the above** 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.

For raters whose task language is not English: You will sometimes rate English queries, even though your task language is not English. Often the landing pages that you see for English queries will contain some or even a lot of English. When the landing page is from a task language website with task language features, such as navigation and headings, and has both English and task language content, please use the flag for the **task language**. Here is an example: The query is the song [walking on sunshine], Chinese Simplified (CN), and the landing page is a page on a Chinese lyrics website with the lyrics in English. The landing page has Chinese navigation and other features in Chinese. The page should be assigned the flag for the **task language**, not English.

Here are some rules to help you when it is very difficult to decide:

- If you can't decide between the task language, an acceptable language, or English, please choose the flag that corresponds to the **task language**.
- If you can't decide between an acceptable language and English, please choose the flag that corresponds to the **acceptable language**.
- Use **Foreign Language** when the page has very little or no text in the task language, an acceptable language, or in English. If you are having trouble deciding between **Foreign Language** and any of the other flags, please do NOT use the **Foreign Language** flag.

Assigning ratings: You will assign a rating from the rating scale to landing pages that are in the task language, an acceptable language, or English. Occasionally, you will also give a **Vital** rating to a page in another language, if you can tell from the URL of the landing page that it is a **Vital** page. The **Vital** rating is described in Section 4.1 and this specific concept is described in Section 4.1.5.

4.0 The Rating Scale

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

Ratings that you assign from the rating scale should *not* be based *only* on the presence or absence of the query terms on the landing page. Some pages that contain the query terms are not helpful at all, while other pages which do not contain the query terms are actually very helpful.

Ratings should also *not* be based *only* on the appearance of the URL. As a rater, you *must* visit the landing page in *every* rating task, and then rate the page content based on its utility for the user intent of the query.

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 have very high or the highest possible utility. Please note that this is not a requirement for a rating of **Vital**, however. Some **Vital** pages do not have high utility.

We will classify **Vital** pages further in section 4.1.5. First, here are many 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	User Intent	English (US) Vital Page	Description of Vital Page
[Wall Street Journal Online], English (US)	Navigate or go to the official homepage of the online version of the Wall Street Journal	http://online.wsj.com/public/us	Official homepage and target of the query
[yahoo], English (US)	Navigate or go to the official Yahoo homepage	http://www.yahoo.com	Official homepage and target of the query
[yahoo mail], English (US)	Navigate or go to the official Yahoo! Mail login page	http://www.mail.yahoo.com	Office webpage and target of the query
[walmart.com], English (US)	Navigate or go to the official homepage of the Wal-Mart online retail site	http://www.walmart.com/	Official homepage and target of the query

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
Specific Products	[Ipod nano], English (US)	http://www.apple.com/ipodnano/	Official product webpage on manufacturer's site
Famous locations	[Statue of Liberty], English (US)	http://www.nps.gov/stli/	Official webpage on government website
	[Baseball hall of fame], English (US)	http://web.baseballhalloffame.org/	Official homepage on owner's website
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/president/	Official webpage on government website
Blogs	[Freakonomics blog], English (US)	http://freakonomics.blogs.nytimes.com/	Official blog webpage on the New York Times website
Universities	[Harvard], English (US)	http://www.harvard.edu/	Official homepage of Harvard University

4.1.3 Vital Pages for People Queries

Queries for famous people, such as [george bush], [Madonna], and [david beckham], have obvious dominant interpretations. Queries that contain 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. For example, Dave Jones is a common English name and the query **[dave jones], English (UK)** can have no **Vital** result. However, the very specific query **[dave jones codemonkey], English (UK)** does have a **Vital** result: <http://www.codemonkey.org.uk/>.

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 Internet. 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 musical bands are considered **Vital**, but social networking pages for companies usually are not.

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
[Arianna Huffington], English (US)	http://www.huffingtonpost.com/	Homepage of The Huffington Post, a blog and commentary website founded by Arianna Huffington. The blog is not specified in the query, but there is clear user intent to navigate to this webpage.	Yes
[oprah], English (US)	http://www.oprah.com/	Homepage of The Oprah Winfrey Show. The tv show is not specified in the query, but there is clear user intent to navigate to this webpage or get information about this show.	Yes
[Dave Smith], English (US)	http://www.daveusmith.com/	Dave Smith is a common name. A personal webpage for someone named Dave Smith is not Vital .	No - no Vital rating is possible.
[photobucket], English (US)	http://www.myspace.com/photobucket	Photobucket is an online photo sharing company. Photobucket maintains a MySpace webpage. This page does not have the official feel of the company’s website. Most users would not try to navigate to this page or consider it official. Compare this page to http://photobucket.com/ .	No - this page is not Vital for [photobucket].
[toyota], English (US)	http://blog.toyota.com/	Toyota maintains a company blog to communicate with the public. Most users would not have the Toyota blog page in mind when typing this query. The Vital page for the query is http://www.toyota.com/ .	No - this page is not Vital for [toyota].
[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 the Green Day YouTube Channel page.	Yes
[Britney Spears], English (US)	http://www.youtube.com/user/britneytv	This is the Britney Spears YouTube Channel page.	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

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: Americans with Disabilities Act American Dental Association American Diabetes Association Each interpretation has an official homepage, but none is Vital since there is no dominant interpretation.
[Amesbury Park, Toronto], English (CA)	No Vital page currently exists	Amesbury Park is a park in Toronto, Ontario. It does not have an official homepage or webpage.
[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 not one official source for knitting information. No one can own this subject.
[diabetes], English (US)	No Vital page is possible	This is an information query. No person or entity can claim ownership of the query [diabetes]. There is no Vital result page.
[ipod reviews], English (US)	No Vital page is possible	[ipod] is an entity query, but [ipod reviews] is not. This is an information query. The user is 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. The user is 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	User intent	English (US) Vital Pages	Description
[barnes and noble], English (US)	Navigate to the official homepage	http://www.barnesandnoble.com/ http://www.bn.com http://www.books.com	Multiple Vital URLs for the official homepage of this company. These are different domains with the same owner; the landing page is the same for all of these URLs.
[penneys], English (US)	Navigate to the official homepage	http://www.jcpenny.com/jcp/default.aspx http://www.jcpenny.com/jcp/default.aspx	Multiple Vital URLs for the official homepage of this company. These are different domains with the same owner; the landing page is the same for both of these URLs.
[cheaptickets], English (US)	Navigate to the official homepage	http://www.cheaptickets.com/ http://www.cheapticket.com/	Multiple Vital URLs for the official homepage of this company. These are different domains with the same owner; the landing page is the same for both of these URLs.

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, 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 the query. Even though the URL “looks” Vital , it’s not.
[Jennifer Lopez], English (US)	http://www.jenniferlopez.net/	This page is not an official homepage for Jennifer Lopez. This is the “real” official Vital page: http://www.jenniferlopez.com/
[simpsons], English (US)	http://www.simpsons.com/	This is the “real” official Vital page: http://www.thesimpsons.com/index.html

4.1.5 Vital Pages and Geographic Location

The Internet represents many languages and countries around the world. Some companies, corporations, and famous people have just one version of their homepage. Frequently, however, companies, corporations, famous people, and other international entities have many versions of their official website. *All such pages get a **Vital** rating.* Because you will encounter many different types of official pages, please use the following categories of **Vital** to describe them:

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

Sometimes it will be difficult to decide which of these **Vital** ratings to assign. Please don’t struggle with the decision – just use your best judgment. Here are some rules that summarize the examples displayed below and which should help you:

- Use the **Appropriate Vital** rating for pages that are the best and the most helpful for users in the task location. Also use **Appropriate Vital** for pages that match the user request, even if the request is for a different location or language than the task location and task language. If the user specifies a specific language or location in the query, the page that matches that specified language or location should be rated **Appropriate Vital**.
- Use the **International Vital** rating for pages in English that can accommodate users in many different locations. These pages are helpful for some or many users.
- Use **Other Vital** for all other **Vital** pages.

Appropriate Vital – Assign this rating to the official page that best matches the user intent and would be the most helpful to the user in the task location. Here are descriptions of different types of pages that should be rated **Appropriate Vital**.

Appropriate Vital: The page is in the task language and was created for the task location.

Query	Appropriate Vital Page	Description
[rabobank], Dutch (NL)	http://www.rabobank.nl	Official homepage of Rabobank, created for Dutch-speaking users in the Netherlands
[rabobank], German (DE)	http://www.rabobank.de/	Official homepage of Rabobank, created for German-speaking users in Germany
[rabobank], English (AU)	http://www.rabobank.com.au/	Official homepage of Rabobank, created for English-speaking users in Australia

Query	Appropriate Vital Page	Description
[louvre], French (FR)	http://www.louvre.fr/llv/commun/home.jsp?bmLocale=fr_FR	Official homepage of the Louvre Museum (located in Paris, France) in French
[Stanford], English (US)	http://www.stanford.edu/	Official homepage of Stanford University (located in California) in English
[Dove 2000], English (UK)	http://www.dove2000.org.uk/	Official homepage of Dove2000, an organization in the U.K. There is no other interpretation for this query, so this is the dominant interpretation.
[campolongo hospital], Italian (IT)	http://www.campolongohospital.it/	Official homepage of the Campolongo Hospital in Italy.

Appropriate Vital: The page is in the task language, but there is no task location associated with it.

Query	Appropriate Vital Page	Description
[united nations], English (US) [united nations], English (UK)	http://www.un.org/english/	Official homepage of the United Nations, created for English-speaking users
[united nations], Spanish (MX) [united nations], Spanish (ES)	http://www.un.org/spanish/	Official homepage of the United Nations, created for Spanish-speaking users
[united nations], Russian (RU)	http://www.un.org/russian/	Official homepage of the United Nations, created for Russian-speaking users
[amnesty international], Spanish (ES) [amnesty international], Spanish (AR)	http://www.amnesty.org/es	Official homepage of Amnesty International, created for Spanish-speaking users
[amnesty international], Arabic (EG) [amnesty international], Arabic (SA)	http://www.amnesty.org/ar	Official homepage of Amnesty International, created for Arabic-speaking users

Appropriate Vital: The entity is not located in the task location and no page has been created for the task location.

- If the entity has versions of the page in more than one language, please assign **Appropriate Vital** to the page in the **task language**.
- If the entity has versions of the page in more than one language but there is not a version in the task language, please assign **Appropriate Vital** to the page in the **language of the entity**.
- If there is only one version of the page, that page is **Appropriate Vital** for all task locations.

Query	Appropriate Vital Page	Description
[louvre], English (US) [louvre], English (UK)	http://www.louvre.fr/llv/commun/home.jsp?bmLocale=en	Official homepage of the Louvre Museum, created for English-speaking users.
[louvre], French (CA) [louvre], French (MA)	http://www.louvre.fr/llv/commun/home.jsp?bmLocale=fr_FR	Official homepage of the Louvre Museum, created for French-speaking users
[louvre], Japanese (JP)	http://www.louvre.fr/llv/commun/home.jsp?bmLocale=ja_JP	Official homepage of the Louvre Museum, created for Japanese-speaking users

Query	Appropriate Vital Page	Description
[louvre], Spanish (MX) [louvre], Chinese (CN)	http://www.louvre.fr/llv/commun/home.jsp?bmLocale=fr_FR	The Louvre has versions of its homepage in French, English, and Japanese. The homepage in French is Appropriate Vital for all users other than English and Japanese speaking users.
[Stanford], Chinese (CN) [Stanford], Italian (IT)	http://www.stanford.edu/	Stanford University has only one version of its homepage. This page is Appropriate Vital for all task locations and task languages.
[Dove 2000], English (US) [Dove 2000], Russian (RU)	http://www.dove2000.org.uk/	Dove2000 has only one version of its homepage. This page is Appropriate Vital for all task locations and task languages.
[campolongo hospital], Czech (CZ) [campolongo hospital], Spanish (ES)	http://www.campolongohospital.it/	Campolongo Hospital has only one version of its homepage. This page is Appropriate Vital for all task locations and task languages.

Appropriate Vital: The query “asks” for a Vital page, other than the page created for the task location.

Sometimes, you will see a query that “asks” for a *different* Vital page, such as [yahoo jp], English (US) or [yahoo German], English (US). When that happens, we will respect user intent and assign the **Appropriate Vital** rating to the page that answers that request, even if it is not in the task language or associated with the task location.

Query	Appropriate Vital Page	Description
[ebay canada], French (FR)	http://www.ebay.ca/	Because the user “asked” for the eBay Canada page, the correct rating is Appropriate Vital .
[yahoo jp], English (US)	http://www.yahoo.co.jp/	Because the user “asked” for the Japanese Yahoo page, the correct rating is Appropriate Vital .

Appropriate Vital: The entity has a “regional” Vital page, but no task location page. This category can also be used for “regional” Vital pages (such as a “North America” or “Asia Pacific” page, when no task location page exists).

Query	Appropriate Vital Page	Description
[billabong], German (DE)	http://www.billabong.com/eu/	Because there is no Billabong page for Germany, the Europe page is Appropriate Vital .
[Palm], Spanish (EC)	http://www.palm.com/latin/	Because there is no Palm page for Ecuador, the Latin America page is Appropriate Vital .

International Vital - Use this category for the top level of an international site that has multiple versions or for a “choose your language or location” page. These pages are usually in English and are designed for users around the world.

Query	International Vital Page	Description
[Microsoft]	http://www.microsoft.com http://www.microsoft.com/worldwide/de/	http://www.microsoft.com is a top-level page. http://www.microsoft.com/worldwide/ is a “choose your location” page. Please note that http://www.microsoft.com should be rated Appropriate Vital for English (US).
[ebay]	http://www.ebay.com/	http://www.ebay.com/ is a top-level page. Please note that for English (US) users, http://www.ebay.com/ should be rated Appropriate Vital .

Query	International Vital Page	Description
[youtube]	http://www.youtube.com	http://www.youtube.com is a top-level page. Please note that http://www.youtube.com should be rated Appropriate Vital for English (US).
[nokia]	http://www.nokia.com/	http://www.nokia.com/ is a “choose your location” page.
[ikea]	http://www.ikea.com/	http://www.ikea.com/ is a “choose your location” page.
[united nations]	http://www.un.org/	http://www.un.org/ is a “choose your language” page.
[nike]	http://www.nike.com/index.jhtml	http://www.nike.com/index.jhtml is a “choose your language” page.

Other Vital – Use this category for any other type of **Vital** page.

Query	Other Vital Page	Description
[rabobank], Dutch (NL)	http://www.rabobank.de/	The German page is Other Vital for Dutch (NL).
[louvre], English (US)	http://www.louvre.fr/llv/commun/home.jsp?bmlLocale=fr_FR	Because there is an English version of the Louvre homepage, the French page is Other Vital for English (US).
[united nations], Russian (RU)	http://www.un.org/spanish/	The Spanish page is Other Vital for Russian (RU).
[Microsoft], Portuguese (PT)	http://www.microsoft.com/brasil/	The Brazilian Portuguese page is Other Vital for Portuguese (PT).
[ebay], English (US)	http://www.ebay.co.uk/	The UK page is Other Vital for English (US).

You will sometimes rate queries that are exact, perfectly formed, working URLs. For these exact, perfectly formed, working URLs, you will give the **Appropriate Vital** rating to the exact match of the URL, regardless of language or location. For a query to be considered an exact, perfectly formed, working “URL query”, it is not necessary for it to contain the leading “www” or “http://”. For example, [www.baidu.com] and [baidu.com] are both considered to be exact, perfectly formed, working URLs. You will learn more about “URL queries” in Section 5.3.5.

Task Language and Task Location	Appropriate Vital	International Vital	Other Vital
[baidu.com], Chinese Simplified (CN)	http://www.baidu.com/ This is the requested page	None	http://www.baidu.jp/ This is not the requested page, but it is the Japanese homepage of Baidu.
[baidu.com], English (US)	http://www.baidu.com/ This is the requested page	None	http://www.baidu.jp/ This is not the requested page, but it is the Japanese homepage of Baidu.
[baidu.com], Japanese (JP)	http://www.baidu.com/ This is the requested page	None	http://www.baidu.jp/ This is not the requested page, but it is the Japanese homepage of Baidu.

When you truly cannot decide which **Vital** rating to assign, please follow these rules:

- If you feel that you cannot decide between the **Appropriate Vital** rating and one of the other **Vital** ratings, use the **Appropriate Vital** rating.
- If you feel that you cannot decide between the **International Vital** rating and the **Other Vital** rating, use the **International Vital** rating.

4.2 Useful

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

Useful pages are pages you trust. They are from information sources that seem reliable. **Useful** pages are usually high quality: they are well organized and up-to-date. **Useful** information pages should not seem “spammy”.

Useful pages are “as good as it gets” for queries that do not have **Vital** pages. A good search engine should have **Useful** pages in top positions on its pages of search results.

Please note that for queries that are pluralized, a single page or a page with a single item that answers the query can be rated **Useful** if the page or item is a great result. For example, for the query [search engines], English (US), the following should be assigned a rating of **Useful**: <http://www.yahoo.com/>, <http://www.ask.com/>, <http://www.live.com/>, <http://www.google.com/>, etc.

4.2.1 Examples of Useful Pages

Query	User intent	Useful Pages	Explanation
[is poison oak contagious?], English (US)	Find the answer to this question. This is an information query.	http://www.calpoison.org/public/poak.html	Authoritative website answers this question very well
[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 90 reviews for this seafood restaurant
[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
[books on dinosaurs], English (US)	Purchase a book about dinosaurs. This is an action query.	http://shopping.discovery.com/product-58659.html	Discovery Channel site with an award-winning book about dinosaurs for sale
[cisco], English (US)	Find stock quote information for Cisco. This is an information query.	http://finance.yahoo.com/q?d=t&s=CSCO http://money.cnn.com/quote/quote.html?symb=CSCO http://finance.google.com/finance?client=ob&q=CSCO	CSCO is the stock symbol for the Cisco Corporation. These webpages are from well-known websites and are all basically the same, providing the same stock charts, trading information, etc. They are all as “good as it gets” for this query.
[meningitis symptoms], English (US)	Find information on the symptoms of meningitis. This is an information query.	http://www.webmd.com/hw/infection/aa34586.asp http://www.nlm.nih.gov/medlineplus/ency/article/000680.htm http://www.cdc.gov/ncidod/d/bmd/diseaseinfo/meningococcal_g.htm http://www.mayoclinic.com/health/meningitis/DS00118/DSECTION=2	Highly informative pages on authoritative sites. They are all “as good as it gets”.

Query	User intent	Useful Pages	Explanation
[search engine], English (US)	Find a search engine to use. This is an action query.	http://www.yahoo.com http://www.google.com http://www.msn.com http://ask.com	These are all very popular search engines in the U.S. These pages have very high utility.
[search engine], Chinese (CN)	Find a search engine to use. This is an action query.	http://www.baidu.com/ http://www.yahoo.cn/ http://www.google.cn/	These are all very popular search engines in China. These pages have very high utility.
[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	This is a 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.
[academy awards nomination best motion picture of 2006], English (US)	The user wants 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.oscars.org/79academyawards/nomswins.html	This is the official Academy Awards website list of the nominations for all 2006 award categories. It includes the Best Motion Picture nominees. Since this page is not dedicated to just the best motion picture nominations of 2006, and shows nominations for all of the categories, the rating cannot be Vital .
[academy awards nomination best motion picture of 2006], English (US)	The user wants 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/Sections/Awards/Academy_Awards_USA/2007	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.

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	User intent	Useful Pages	Explanation
[stephen colbert], English (US)	Stephen Colbert is a famous comedian. While the homepage of his TV show is Vital for this query, users often look for entertaining Steven Colbert material.	http://video.google.com/videoplay?docid=-869183917758574879	This is a famous presentation in which Stephen Colbert made fun of George Bush and his administration.
[dance video], English (US)	There are many good, entertaining, and popular dance videos on video websites. The user is looking for good or entertaining dance videos.	http://www.youtube.com/watch?v=dMH0bHeiRNQ	This is a popular video of a comedian demonstrating dance styles from previous decades.

For some queries, timeliness is very important. Queries for recent events and recurring events need recent webpages. For example, 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	User Intent	Useful Pages	Slightly Relevant Pages
[Australian Open results], English (US)	Find a page that displays the most recent results for this tennis tournament. This is an information query.	Wikipedia page with the 2008 results: http://en.wikipedia.org/wiki/2008_Australian_Open	Wikipedia page with the 2004 results: http://en.wikipedia.org/wiki/2004_Australian_Open
[Academy Award best picture], English (US)	Find the most recent winner of this award. This is an information query.	Page on the CNN website with this information: http://www.cnn.com/2008/SHOWBIZ/Movies/02/25/oscars/	Page on the official Academy Award website with the 2005 winner of this award: http://awardsdatabase.oscars.org/ampas_awards/DisplayMain.jsp?currentTime=1204074197248
[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 BBC News website with this information: http://news.bbc.co.uk/2/hi/europe/7041082.stm Page on the New York Times website with this information: http://www.nytimes.com/2007/10/13/world/13nobel.html	Page on the BBC website with the 2006 winner of this prize: http://news.bbc.co.uk/2/hi/europe/6047020.stm

4.3 Relevant

A rating of **Relevant** is assigned to pages that have fewer valuable attributes than were listed for **Useful** pages. **Relevant** pages might be less comprehensive, come from a less authoritative source, or cover only one important aspect of the query. **Relevant** pages must have some utility for the user, in addition to being on-topic. These pages are average to good, but are not “as good as it gets.”

4.3.1 Examples of Relevant Pages

Query	User intent	Relevant Pages	Explanation
[seoul, korea], English (US)	Travel to Seoul, or find information about the city	http://www.escortmap.co.kr/english/e_sall.htm	Page with a map of the city of Seoul
[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 isn't good enough to be Useful . There are much better pages on the Web.
[hot dogs], English (US)	Find information about hot dogs, such as recipes	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 U.S. A rating of Useful is also acceptable for this page.
[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 U.S. presidents, including the birthday of Abraham Lincoln. However, Lincoln's birthdate is not prominently displayed and might be a little hard to find.

Query	User intent	Relevant Pages	Explanation
[wii], English (US)	Purchase the wii video game console, find games for the wii, or navigate to the official wii webpage on the Nintendo website.	http://www.amazon.com/Nintendo-RVLSWCUSZ-Wii/dp/B0009VXBAQ	Amazon.com page with new and used wii game consoles for sale through affiliate merchants
[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	Page with one review of this restaurant

4.4 Slightly Relevant

A rating of **Slightly Relevant** is assigned to pages that are generally not helpful, but are still marginally on-topic. **Slightly Relevant** pages may be outdated, too narrowly regional, too specific, too broad, etc. to receive a higher rating. They might have less information or come from a less authoritative source.

4.4.1 Examples of Slightly Relevant Pages

Query	User Intent	Slightly Relevant Pages	Explanation
[Louvre], English (US)	Find information about visiting the Louvre or about its art collection, or navigate to the museum's homepage	http://www.ivebeenthere.co.uk/tips/13457	Review of the Café Louvre, a Czech café located in Prague
[BBC], English (US)	Navigate to the homepage of the BBC	http://www.bbc.co.uk/dna/mbfans/forum/F2154398	The "Dundee United" Fans Forum on the BBC website - an unimportant subpage
[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.
[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.
[abe lincoln's birthday], English (US)	Find this specific piece of information	http://www.everythingsl.net/lessons/presidents_day_abraham_lincoln_93511.php	Teacher's lesson plan page for teaching about Abe Lincoln. Page mentions his birthday, but not the year of his birth.
[Citizen Kane DVD], English (US)	Purchase this DVD	http://en.wikipedia.org/wiki/Citizen_Kane	Wikipedia page about the movie, which briefly mentions restoration of the DVD and commentary on it
[britney spears], English (US)	Find current news or pictures related to Britney Spears	http://www.brandchannel.com/features_webwatch.asp?ww_id=76	2004 article about the annulment of Britney's first marriage. This is very old news that would not be of interest to most users.

4.5 Off-Topic

A rating of **Off-Topic** should be given to pages that have no or very little topicality and no or almost no utility. These pages are not helpful for most users. In some cases, there is lack of attention to an important modifier or element of the query. In other cases, the query terms appear on the page in separate places, but the page topic does not match the query topic.

4.5.1 Examples of Off-Topic Pages

Query	User Intent	Off-Topic Pages	Explanation
[Australian Open mens singles result 2008], English (US)	Find a page that displays the 2008 mens singles result for this tennis tournament. This is an information query.	Wikipedia page with the 2004 results: http://en.wikipedia.org/wiki/2004_Australian_Open	Wikipedia page with results from 2004. The query wants results for 2008. Since the page ignores an important element of the query – the year – the rating is Off-Topic .
[Tom Cruise], English (US)	Find information or news about Tom Cruise or purchase a DVD of one of his movies	http://www.ussslater.org/signal/s/vol-3/ss-v3-n4.html	Page that mentions Tom Beeler and Tom Moore and vacation crui ses.
[hotmail login]., English (US)	Navigate to hotmail login page	https://login.yahoo.com/config/oqin_verify2?&.src=yym	Login page for Yahoo! Mail
[german cars], English (US)	Find information about German cars or navigate to official homepage of a German automaker	http://www.subaru.com/	Homepage of Subaru, a Japanese car company
[meningitis symptoms], English (US)	Find information on the symptoms of meningitis	http://www.ifrc.org/WHAT/health/archi/fact/Fmengts.htm	Page about meningitis in Africa with no information about the symptoms of the disease
[earthquakes], English (US)	Find information or news about earthquakes	http://www.yahoo.com/	Search engine page that has no connection to the query. Even though you can issue the query in the search engine and get good results, the rating should be Off-Topic .
[hot dog], English (US)	Find information about hotdogs, such as recipes	http://www.peteducation.com/article.cfm?cls=2&cat=1675&articleid=812	A page about dog houses which happens to display the word “hot” is Off-Topic .
[universities in Europe], English (US)	Find a list of universities in Europe	http://www.indianchild.com/universities_in_india.htm	A page with contact information for universities in India provides no utility.
[Canon SD 1000], English (US)	Purchase or find information on this camera	http://www.dpreview.com/reviews/canonsd300/	Page about the Canon SD 300, a different Canon camera

You will also come across many pages that are so unhelpful (and possibly deceptive) that they should be rated **Off-Topic**. Here is an example: websites that have fake links to download software. Often these are illegal downloads for cracked codes, serials, keygens, etc. When the user types a query in the search box on the landing page, he is presented with many links that appear to be helpful in downloading the desired software. However, the links redirect to other pages that lead to yet other links. Eventually, the user may be redirected back to the same page or to a different page requiring payment for the service. Since the starting page is so unhelpful to the user, it should be rated **Off-Topic**.

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**
- **Meaningless Query:** Query is meaningless even after query research – No rater can reasonably rate this query
- **Other**

Please note that you may assign more than one **Unratable** rating to a page. For example, a page in a foreign language that does not load should be assigned both **Unratable: Didn't Load** and **Unratable: Foreign Language**.

4.6.1 Unratable: Didn't Load

The rules for assigning Unratable: Didn't Load have changed and this entire section was updated on July 1, 2009. Please carefully read this whole section and review the examples.

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.

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

- Pages with error messages (e.g. *generic 404 errors, client errors, web server errors, network timeout errors*) *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!"

See http://en.wikipedia.org/wiki/List_of_HTTP_status_codes for descriptions of different types of error messages.

Landing pages with *customized* 404 error messages (error messages that have been customized by the webmaster to provide information to users) do *not* belong in **Didn't Load** category and should be rated according to their utility.

Here are examples of pages with different kinds of error messages and suggested ratings:

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.
[victorianbc.com], English (US)	http://victorianbc.com/	"Address Not Found. Firefox can't find the server at victorianbc.com."	Didn't Load	The browser could not find the server. There is no content on the page.
[united nations], English (US)	http://disarmament.un.org/wmd/bwc/index.html	"Cannot Complete Request. Additional information about this problem or error is currently unavailable."	Didn't Load	The request cannot be completed. There is no content on the page.
[protein solutions], English (US)	http://www.proteinsolutions.co.uk/	"Please wait whilst you are being redirected....."	Didn't Load	A redirect is promised, but nothing happens. There is no content on the page other than the message.
[mcdonald's monopoly], English (US)	http://monopoly.corsis.com/monopoly07/front.do	"Network Timeout. The server at monopoly.corsis.com is taking too long to respond."	Didn't Load	The server is not responding. 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.
[ikamedia.com], English (US)	http://ikamedia.com/cgi-bin/	"Forbidden. You don't have permission to access /cgi-bin/ on this server. Additionally, a 500 Internal Server Error error was encountered while trying to use an ErrorDocument to handle the request."	Didn't Load	This is a forbidden access error message. There is no content on the page.
[www.religareonline.com], English (US)	http://www.google.com/interstitial?url=http://www.religareonline.com/Fundamental/Advancedsearch.asp%3FReportType%3DHouse%26PRCode%3D10291	"Warning – visiting this web site may harm your computer!"	Didn't Load	Pages with warning messages should be rated Didn't Load .

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	“Warning: ocilogon(): u=hppio OCI_SessionBegin: ORA-01034: ORACLE not available ORA-27101: shared memory realm does not exist IBM AIX RISC System/6000 Error: 2: No such file or directory in D:\jungang_homepage\design05\user\oci8.inc on line 9”	Foreign Language	Even though there are internal system error messages on the landing page, there are working links.
[ufc ring girls video], English (US)	http://en.sevenload.com/videos/bKLWib/UFC-RING-GIRLS-RACHELLE-LEAH-AND-AMBER-NICHO	“Oops! This document is not available. The document you tried to open is not available on this server. Please try to get the content you want by using our default navigation at the top of the sevenload browser.”	Off-Topic	The landing page has content and working links, but it has no utility for the user intent.
[terrifically tacky tape], English (US)	http://www.dickblick.com/zz614/55a/	“Navigation Error - 404 Page Not Found. The page you requested cannot be found. The product you are seeking may have been discontinued.”	Off-Topic	This is a customized 404 error message. Although the landing page has many working links, it has no utility for the user intent.
[Netgear - FS524NA Port Network Switch], English (US)	http://www.l-com.com/Error.aspx?aspxerrorpath=/item.aspx	“We're sorry. The page you requested cannot be found. The page might have been removed, had its name changed, or is temporarily unavailable.”	Off-Topic	In spite of the message, the landing page has many working links, but it has no utility for the user intent.
[news about Lupin Group], English (US)	http://www.lupingroup.com/	“Web Site Under Construction”	Off-Topic	When the query does not have navigation intent, pages that are under construction should be rated according to their utility. The landing page has no information about Lupin Group.
[asi flex], English (US)	http://books.google.com/books?id=ZtmUSbzMml8C&pg=PA299&lpg=PA299&dq=asi+flex&source=web&hl=en&ots=47RnX0tpVQ&sig=eKZySbAlUJsyt5y2sgufz0_tNus	This is a Google Booksearch result. The particular page from the book can't be displayed and shows this message: “You have either reached a page that is unavailable or reached your viewing limit for this book.”	Off-Topic	The booksearch result page loads, but the particular book page can't be viewed. The links on the page work and other pages within the book can be viewed. The rating is Off-Topic because the book has no connection to the query.
[hoover.com], English (US)	http://assist.hoover.com/	“Hoover.com is currently undergoing maintenance. This upgrade affects all areas of the web site. Please bear with us as we perform this upgrade to make your visit to the site a more useful and informative one.”	Off-Topic or Slightly Relevant	This example illustrates that raters really have to rate carefully. Hoover's website at http://www.hoover.com does load, so this page can't be Vital . This is an outdated message. The landing page is on the Hoover domain, but it has no utility to users.

Query	URL of the Landing Page	Landing Page Error Message	Rating	Explanation
[widgetbox], English (US)	http://www.widgetbox.com/widget/tetris	"This widget has been disabled."	Slightly Relevant	In spite of the message, the query has navigation intent and the landing page is on the correct website. There are working links on the landing page.
[capital one], English (US)	http://www.msnbc.msn.com/id/22589062/	"The page you are seeking has expired and is no longer available at msnbc.com."	Slightly Relevant	Even though the article is no longer available, there is live stock information for Capital One on the landing page.
[new yorker], English (US)	http://www.newyorker.com/fact/content?040531fa_fact1	"Error. The page you are looking for could not be found. Try the search box to find the page you were looking for."	Relevant	In spite of the customized message on the page, the landing page has most of the current day's content, sidebars, top navigation links, etc.
[Lennox HC-EAC], English (US)	http://www.lennox.com/products/overview.asp?model=HC-EAC	"The model requested HC-EAC does not exist."	Relevant	The landing page has content and working links. A rating of Relevant is appropriate because the page lets users know that this model no longer exists.
[video], English (US)	http://www.youtube.com/user/RENNET0	"This account is suspended."	Relevant or Useful	The landing page is identical to www.youtube.com except for the message displayed at the top. YouTube is a very popular video website.
[mercedes], English (US)	http://www.mbusa.com/mbusa/ProxyServerlet?/care/client-care/teleaid.do	"We are sorry, there is no MBUSA.com web page matching your request. It is possible you typed the address incorrectly, or that the page no longer exists. Please visit any of the pages below for information about Mercedes-Benz services and products."	Relevant or Useful	In spite of the customized message on the page, the landing page is on the official Mercedes USA website and has links to all vehicles types, build your own car, compare prices, etc.
[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.
[youtube], English (US)	http://www.youtube.com/watch?v=Ct5Zxsf2YeE	"This video is no longer available due to a copyright claim by World Wrestling Entertainment, Inc."	Useful	The landing page is identical to www.youtube.com except that this message is displayed at the top. The page is not an exact duplicate of YouTube's homepage, but it is "as good as it gets".

Query	URL of the Landing Page	Landing Page Error Message	Rating	Explanation
[Amanda Bynes], English (US)	http://www.amandabynes.com/	"Under Construction?"	Vital	This website appears to be owned by Amanda Bynes and was developed in the past: http://web.archive.org/web/20060214153116/http://www.amandabynes.com/ .
[elf yourself], English (US)	http://www.elfyourselves.com/	"The Elves Have Left the Building. Thanks for elfing yourself! Check back next holiday season for more ElfYourself fun!"	Vital	For a few years, OfficeMax (an office supply company) has posted a game on this website during the holiday season. Users visiting the site at other times of the year get a customized message such as the one currently displayed. The landing page is the target page of the query, even when the game is not active.
[yahoomail], English (US)	http://www.yahoomail.com	"Don't have a Yahoo! ID? Signing up is easy. Already have a Yahoo! ID? Sign in."	Vital	Users typing this query expect to land on this page to register or sign in.
[lupingroup.com], English (US)	http://www.lupingroup.com/	"Web Site Under Construction"	Vital	The landing page is the target of this URL query.
[delsey], English (US)	http://www.delseyusa.com/	"Our website is under maintenance. You can still check out our product information."	Vital	The landing page has product information links that are functional. The page should be rated Vital , because this luggage company is the dominant interpretation of the query. The page is just undergoing some redesign.

When you assign **Unratable: Didn't Load** based on one of the types of messages listed above, please copy and paste the message in the comments section of the rating task.

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

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 landing page language flags, 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.
- The page is not clearly **Vital** for the query, based on the appearance of the URL of the landing page.

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.

Here are two situations where you should use the **Vital** rating instead of **Unratable: Foreign Language**:

- The query is a website name or a URL *that exactly matches the landing page URL*, and there is no other reasonable interpretation of the query. Here is an example. The query is [www.baidu.com] and the landing page is <http://www.baidu.com>. The rating for this page is **Appropriate Vital** for any task language. See Section 4.1.5.
- The query is for a famous website (such as Microsoft) and the URL shows that it is an official page for the wrong country. See Section 4.1.5.

4.6.3 Unratable: Meaningless Query

Assign **Unratable: Meaningless Query** when the query is meaningless even after query research, and you believe that no other rater will be able to reasonably rate the query either.

There will be *very few* queries that fall into this category. If you research the query and use good reasoning, you should be able to find a meaning for most queries, even if you don't know exactly what the user is looking for.

Here are the important things to remember:

- This rating should be assigned **very** rarely.
- Do **not** assign this rating to a query that is about a technical or scientific subject that you don't understand.
- Do **not** assign this rating just because the task is difficult to rate.
- If a task is too difficult for you and you can't assign a rating, use the "release" button and let someone else rate the task.
- **Always** research the query. Most queries make sense after doing some research.
- Use more than one search engine to research the query. Some search engines do better with certain types of queries than other search engines do.
- **Always** click the link in the task. The content on the landing page may help you understand the query.
- Queries that are composed of real words or strings of numbers usually have meaning. Remember that users may not necessarily formulate queries the same way you do. Search engines do a great job of finding keywords, and some users take advantage of that by typing queries in unusual ways. Sometimes, it helps to look at sections of a query and try to make sense of it that way.
- Do **not** assign this rating just because there are many interpretations and you can't figure out what the user is looking for.
- Do **not** assign this rating just because the query is broad, vague, or ambiguous. We understand that single word queries are difficult to rate, but they are not unratable. They have meaning; you may just have trouble deciding what the user was trying to accomplish. Be aware that a user may be looking to find the definition of the word or its correct spelling. It is also possible that the letters of the word form an acronym or a stock symbol.
- Do **not** automatically assign this rating when the query is a single letter or a single number. Some single letter or single number queries are meaningless, but not all. You will have to do query research to determine if a specific single letter or single number query is meaningless. Here is an example of a single letter query that is *not* meaningless: [C], English (US). The letter "C" is the stock symbol for the Citigroup, Inc., a large commercial bank. Here are some other meanings for [C]: the programming language, the abbreviation for Celsius, and the chemical element, carbon.
- Some queries have misspellings, but with a little effort you are able to make sense of them. When you see a word or query that doesn't seem to make sense, look more closely to see if it might be misspelled.

- Sometimes if you split a query into parts or merge parts of a query to make new words, it will make sense.
- Pay attention to features on the search engine result page which might help you to understand the query, such as spell corrections, query suggestions, and other features. Different search engines offer different corrections and suggestions, which is another reason to check queries in more than one search engine.

Here are examples of queries that should be rated **Meaningless Query**:

- [????]
- [ar ra ar ra ar ra ar ra]
- [oexzilagnasdggggggggg]

Here are some examples of queries that may be difficult to rate, but should not be rated **Meaningless Query**:

Query	Explanation
10544	Research shows that this query can be a gene ID number, a product number, etc.
"i'm losing my mind"	These are English words. They are not meaningless. There happens to be a song titled "I'm Losing my Mind", but even if such a song didn't exist, the query is not meaningless.
...stay with me the best is yet to be	These are English words. The ellipsis (...) in front of the words indicates that the words are part of a quote or lyrics that the user is looking for.
116 high	This incomplete query is probably part of an address. The best results will probably be rated Slightly Relevant or Off-Topic .
ahhdd	This query could be a misspelling for "adhd" or attention-deficit hyperactivity disorder, or it could be the chemical abbreviation for 2-amino-4-hydroxyl-6-hydroxymethyl dihydropteridine diphosphate.
andrew sons	There is a baritone singer with the name "Andrew Sons". Helpful pages about this interpretation will usually be rated Useful . There are also companies called Andrews & Sons. Helpful pages for these companies will usually be rated Slightly Relevant .
austrillia morgages	This query is probably misspelled. The correct spelling is [australia morgages].
b52h2k2	This query is probably looking for the following zip file: B52h2k2.zip
big 100 in dc	There is a radio station in Washington, D.C. called both "Big 100.3" and "Big 100".
big face funny	These are English words. The user could be looking for images of a big funny or happy face.
books invents	These are English words. The user might be looking for books with the word "invents" in the title, or books about inventions.
bowtrul	This query is probably misspelled. The user is probably looking for Bowtrol, a colon health product.
cfc	There are many interpretations for this query. Queries with meanings should not be rated Meaningless Query .
f	"F" is the stock symbol for the Ford Motor Company, and it also represents the Fahrenheit scale.
faithfully	This is a misspelled query. It should be spelled "faithfully". This is an English word. There is even a popular song called "Faithfully".
gcbro	There is more than one interpretation for this query, including the Gulf Coast BigFoot Research Organization.
gold and brady 2003	There are two authors who write articles and books together named Paul B. Gold and Kathleen T. Brady. The user might be looking for publications from 2003.
google maps satellite youtube bikini	These are English words. The user might be looking for Google Earth (rather than Google Maps) satellite pictures displayed on YouTube of Bikini Island or of girls in bikinis.

Query	Explanation
Gurmail	There are people on the web with Gurmail as their first name. Helpful pages about these people should probably be rated Relevant .
systems	There are many companies with "systems" in their name, such as Cisco Systems. Homepages of such companies should be rated Relevant . The user might also be looking for the meaning of the word in science and/or the organization of entities.
hide suff	This query is probably a misspelling for "hide stuff". Although even this correctly spelled query is ambiguous and hard to rate, these are English words and are not meaningless.
siz zig	If you merge there two "words" and search for [sizzig], you will find that the user is probably looking for sizzling "sizzig" or "sisig", a popular Philippine dish.
tetr	TetR is a gene regulatory network. This might also be an incomplete query, where the user meant to type "tetris".
timemn	If we divide the query into two parts: "time" and "mn", the query could be asking for the time in Minnesota.
urea price	Urea is a product that is sold, and it is reasonable that users might look for its price.
waterworks austrian swimmers	This query is misspelled. Waterworks is a recreational spot in Australia, not Austria.
wine leto	Research shows these two people associated with wine: Jared Leto and Julie Leto.

You will use the "release task" button for queries that you personally can't rate, but that other raters may be able to.

Please do not hesitate to use the "release task" button when the query is very technical or scientific or is about something you just can't understand. Studies have shown that raters who have specialized knowledge do a better job rating queries about specialized subject matter than raters without that specialized knowledge. For example, raters with computer science degrees do a better job rating computer science related queries than raters without computer science degrees. If you acquire a task with a query that is about a subject which *really* requires specialized knowledge, or if you *really* feel that you cannot rate the task, you may release it.

4.6.4 Unratable: Other

Assign "**Unratable: Other**" to any other pages that you are not able to evaluate. If you choose "**Unratable: Other**", you must leave a comment to tell us why.

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 rating 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. Some results that are good for raters are not good for users.
- Location is important. Good pages must be location-appropriate.

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. Often, the most difficult queries are short queries.

Here are some examples of user intents behind the query:

Query	Likely User Intent	Vital or Useful Pages	Relevant or Slightly Relevant Pages
[Fedex], English (US)	Track a package or find a Fedex (Federal Express) location	Fedex (Federal Express) homepage: http://www.fedex.com/us/ : Vital	Wikipedia page on Fedex: http://en.wikipedia.org/wiki/FedEx : Relevant
[calendar], English (US)	Find, customize, and print a calendar for the current month or year Find a calendar that displays holidays Find an online calendar to use	Site on which to make customized, printable calendars: http://www.timeanddate.com/calendar/ : Useful Yahoo calendar: http://calendar.yahoo.com/ : Useful	Article on the history of different types of calendars: http://astro.nmsu.edu/~lhuber/leaphist.html : Relevant 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 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 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.
- Only dominant interpretations can have **Vital** pages.
- Pages for common interpretations of the query and reasonable user intents should *not* be lowered in rating.

Here are some examples:

Query Interpretation	Example	Range of Ratings
Dominant Interpretation: Most users would want this interpretation.	<p>[apple], English (US): Apple computers</p> <p>[windows], English, (US): the Microsoft operating system</p> <p>[amazon], English (US): the popular website www.amazon.com</p> <p>[median], English (US): 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 is Useful.</p> <p>[guinea pig], English (US): the small furry animal often kept as a pet. Even though this query has a dominant interpretation, no Vital rating is possible since no one can own this query. The highest possible rating is Useful.</p>	Vital to Off-Topic
Common Interpretation: Many users would want this interpretation.	<p>[apple], English (US): The fruit</p> <p>[windows], English (US): The glass paned windows for a home</p> <p>[amazon], English (US): The rainforest or river in South America</p> <p>[ada], English (US): The American Dental Association or the American with Disabilities Act</p> <p>[mercury], English, (US): The car brand, the planet, or the chemical element</p> <p>[sandals], English (US): The open type of shoe, or the chain of resorts located in the Caribbean Sea with this name</p>	<p>Useful to Off-Topic:</p> <p>There can be no Vital page if the interpretation is not dominant.</p>

Query Interpretation	Example	Range of Ratings
Minor Interpretation: Some users might want this interpretation, but other interpretations are more likely.	[mercury], English (US): The San Jose Mercury News newspaper [median], English (US): Median Technologies, a French company [hot dog], English (US): "Hot Dog", a movie that was in movie theaters in 1984	Relevant to Off-Topic: 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

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 which are less reasonable or less likely should also be lowered on the rating scale.

User Intent	Example	Range of Ratings
Reasonable or likely user intent	[tetris], English (US): Play Tetris (a video game) online, or download the game [flowers], English (US): Order flowers online [credit card], English (US): Find a credit card company and apply for a card, or compare different brands of credit cards [amazon], English (US): Go to Amazon.com.	Vital to Off-Topic
Unreasonable or unlikely user intent	[tetris], English (US): Research the history of Tetris [flowers], English (US): Find a definition of the word "flower" [credit card], English (US): Read an encyclopedia article on credit cards [amazon], English (US): Read an encyclopedia article about Amazon.com	Relevant to Off-Topic: You should lower the rating for unreasonable or unlikely intents

5.3 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. The queries you rate are not written by you, 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.

The following sections discuss some hard rating situations.

5.3.1 Dictionary or Encyclopedia Results

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

Query	Reasonable User Intent	Landing Page	Rating	Reason
[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.”
[banks], English (US)	Find a bank. This is an action query.	http://www.investorwords.com/401/bank.html http://en.wikipedia.org/wiki/Bank	Slightly Relevant	Most English U.S. users know what a bank is. Even an excellent definition or encyclopedia article has little utility.
[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.

5.3.2 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 are action queries, information pages should not be rated above **Relevant**. Think about whether the user wants to know something or do something.

Query	Reasonable 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	Relevant	Most English U.S. users know what an e-card is. The user wants to send an e-card, not read about them.
[text twist], English (US)	Play TextTwist online or download the game. This is an action query.	http://www.answers.com/topic/xttwist?cat=technology	Relevant	The user wants to play the game, not read about it.
[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.

5.3.3 Queries That Ask for a List

Some queries seem to ask for a list. You may see “queries that ask for a list” typed in singular or plural form. For example, the query may be [bank], English (US) or [banks], English (US).

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 the user.

A landing page that provides links to many good individual results can also be very helpful to users. For example, online travel websites, such as www.expedia.com and www.orbitz.com, which allow users to make airline reservations on many different airlines, may be very helpful for the query [airlines], English (US).

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

Query	Reasonable User Intent	Landing Page	Rating	Reason
[bank], English (US)	Find a bank to open an account or get a loan. This is an action query.	http://www.investorwords.com/401/bank.html http://en.wikipedia.org/wiki/Bank	Slightly Relevant	Most English U.S. users know what a bank is. Even an excellent definition or encyclopedia article has little utility.
[banks], English (US)	Find a bank to open an account or get a loan. This is an action query.	http://www.internationalist.com/SERVICE/BANKS/United States.html https://www.wellsfargo.com/	Relevant to Useful	This list of banks is comprehensive and has links to individual banks. This is a well known U.S. bank.
[credit cards], English (US)	Sign up for a credit card; compare different credit card brands before signing up for a card. This is an action query.	https://www.americanexpress.com/ http://www.usa.visa.com/personal/ http://www.mastercard.com/us/gateway.html http://www.discovercard.com/	Useful	These are all very well known and trustworthy credit card companies in the U.S.
[books about sharks], English (US)	The user wants to buy or borrow books about sharks. This is an action query.	http://www.aqua.wisc.edu/waterlibrary/kids/booklist_sharks.html	Relevant	This is a short but good list of books from a library.
[books about sharks], English (US)	The user wants to buy or borrow books about sharks. This is an action query.	http://www.amazon.com/Book-Sharks-Richard-Ellis/dp/0679722106	Relevant to Useful	This is a good book about sharks. Amazon is a popular website and the reviews are helpful. It is a single example, but a search engine result page of great shark books forms a nice list.
[hotels], English (US)	The user wants to make hotel reservations.	http://www.hilton.com http://www.sheraton.com http://www.hotels.com http://www.orbitz.com/	Useful	Hilton and Sheraton are well-known hotel chains. The hotels.com and orbitz.com websites allow users to find information and make reservations at lots of different hotels.

5.3.4 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 the query, [federal expres], English (US), it is reasonable to assume that the user is looking for Federal Express at <http://www.fedex.com/>. 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.

For queries that are **not** obviously misspelled or mistyped, you should assume the user is looking for results for the query as it is spelled. For example, consider the query [mspace], English (US). Many raters may believe that the user intent is to go to the homepage of MySpace, the very popular social networking website. However, there is a company named Microspace Instruments which has an official homepage of www.mspace.com, and there is a web interface called mSpace which has an official homepage of www.mspace.fm. So, the query [mspace], English (US) is not necessarily misspelled.

Similarly, 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 because there is a Las Vegas show producer named John Stuart, whose name exactly matches the spelling of the query.

However, because MySpace and Jon Stewart are very well known, pages for those interpretations are not necessarily **Off-Topic**, since it is possible that a user may have accidentally typed “mspace” instead of “myspace” and “John Stuart” instead of “Jon Stewart”.

Since we want to respect queries as they have been typed by the user, good pages for MySpace and Jon Stewart will not receive as high ratings as good pages for mspace and John Stuart.

Here are some examples:

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
[mspace], English (US)	There are three reasonable query interpretations: Microspace Instruments, a photographic instruments company whose URL is mspace.com; mSpace, a web interface developed by the University of Southampton; and MySpace, the popular social networking website.	http://mspace.com/	Official homepage of Microspace Instruments, a photographic instruments company	Useful
	Because the query is not obviously misspelled and there are websites that exactly match the query, we will assume that Microsoft Instruments and mSpace are the most likely query interpretations and that they are equally likely.	http://mspace.fm/	Official homepage of mSpace, the web interface	Useful
[mspace], English (US)	Because MySpace is extremely popular, we will consider the possibility that the query has been misspelled and consider the MySpace interpretation to be a minor or common interpretation.	http://www.myspace.com/	Official homepage of MySpace, the social networking website	Slightly Relevant or Relevant

Query	Query Interpretation	URL of the Landing Page	Description of the Landing Page	Rating
[Jamie Fox], English (US)	There are two reasonable interpretations for this query: the guitarist named Jamie Fox, and the famous actor named Jamie Foxx. Because the query is not obviously misspelled, we will assume that Jamie Fox, the guitarist, is the dominant interpretation.	http://www.jamiefoxquitar.com/	Official homepage of Jamie Fox, the guitarist	Vital
[Jamie Fox], English (US)	Because Jamie Foxx is such a famous actor and his name might be easily misspelled, we will consider Jamie Foxx to be a minor or common interpretation.	http://www.jamiefoxx.com/	Official homepage of Jamie Foxx, the actor	Slightly Relevant or Relevant

5.3.5 URL Queries

There are many queries that are URLs or look like parts of URLs. We will call these queries “URL queries”. 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 you are rating. Do not just rate URL queries and results based on appearance of the URL.

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] to be the same query as [ibm.com].

Many URL queries are *not* “working” queries. They do not load if you paste them in your browser address bar. 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 with the URL in the query 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.pizzzahut.com] and [www.mcriosoft.com].
- The query has a URL-like format, but contains extra words and/or spaces: [Australian open tennis tournament .com], English (US).
- The query has a mix of words and URLs or website names, such as [walmart.com target.com], English (US).

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 formatted, working URL, please consider that URL to be the dominant interpretation. The **Vital** rating may be given when the URL of the page exactly matches the URL in the query.

If the query is not perfectly formatted 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	Reasonable User Intent	Rating Examples
[www.myspace.com], English (US)	The user wants to go to the MySpace website. The URL is correct.	Vital landing page URL: http://www.myspace.com/
[www.mspace.com], English (US)	The URL www.mspace.com works, so we will assume that the user wants to navigate to www.mspace.com . There is a possibility that this is a misspelled query and that the user wants to go to the very popular site www.myspace.com . However, we will respect the way the query is written and consider MySpace to be a common or minor interpretation.	Vital landing page URL: http://www.mspace.com/ Slightly Relevant to Relevant landing page URL: http://www.myspace.com/
[www.yahoo.c0m], English (US)	Even though the URL doesn't load, it is clear the user wants to go to Yahoo.	Vital landing page URL: http://www.yahoo.com/
[American inventor .com], English (US)	Some users add ".com" to a query to indicate they want a homepage or specific webpage. In this example, the likely user intent is to navigate to the homepage of the popular "American Inventor" TV show (www.americaninventor.tv). However, there is a website named www.americaninventor.com . Because this is not a perfectly formatted working URL and we are not sure about user intent, we will call both americaninventor.com and americaninventor.tv "common" interpretations of this query.	Useful landing page URL: http://www.americaninventor.tv/ Useful landing page URL: http://www.americaninventor.com/
[www.circuitcity.com], English (US)	In this case, the landing page is spam. The URL is a misspelling of the popular electronics site www.circuitcity.com . It is very likely that the user wants to navigate to www.circuitcity.com . However, we will respect the query as written and consider www.circuitcity.com to be dominant.	Vital landing page URL: http://www.circuitcity.com/ (You will also need to add a Spam flag. Please see Part 3 of the "General Guidelines".) Useful landing page URL: http://www.circuitcity.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 3 of the "General Guidelines".) Useful landing page URL: http://www.thesimpsons.com/
[www.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 tax form 1040.	Vital landing page URL: http://www.irs.gov/pub/irs-pdf/f1040.pdf
[toys are us.com], English (US)	There is a well-known U.S. 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/
[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/

5.3.6 Search Engine Result Pages

Another situation that has caused problems for raters in the past is when the landing page is a search engine result page. When we rate URL rating tasks, we assume that the user has typed the query in the “regular” search box of a search engine, and has already experienced seeing a page of search results. We also assume that the page we are evaluating is a search result that a user sees after clicking a link on the page of search results. Here is how to rate search engine result pages:

If the page is a set of generic search results related to the query from a major search engine, this is not a helpful page for the user and should get a rating of **Slightly Relevant**.

However, if the landing page is a set of results from a specialty search (such as a maps, shopping, books, video, images, etc. results page), the page might be very helpful to the user. Ratings will range from **Useful** to **Slightly Relevant**, depending on the utility of the page.

Please note that, if the landing page is a search engine page with an empty search box and no results displayed to evaluate, the page has no connection to the query and should get a rating of **Off-Topic**.

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

Query	Reasonable User Intent	URL of the Landing Page	Rating	Reason
[books about sharks], English (US)	The user wants books about sharks.	http://www.google.com/search?q=books+about+sharks	Slightly Relevant	We will assume that a user has already experienced this kind of search result page.
[books about sharks], English (US)	The user wants books about sharks.	http://books.google.com/books?q=sharks	Useful	This is a specialty search result page which has actual shark books to preview or read.
[books about sharks], English (US)	The user wants books about sharks.	http://www.google.com	Off-Topic	This page isn't topical and has no utility. A user would have to start his search again.
[moby dick book], English (US)	The user wants to buy this book; the user wants information about this book.	Generic search results page: http://www.ask.com/web?q=moby+dick&search=search&qsrc=0&o=0&l=dir	Slightly Relevant	We will assume a user has just done the search on a search engine like Ask.com, so this page has low utility.
[moby dick book], English (US)	The user wants to buy this book; the user wants information about this book	Book search results page: http://books.google.com/books?q=moby+dick&btnG=Search+Books	Relevant or Useful	This is a specialty search result page which displays actual versions of this book to preview or read.
[san francisco restaurants], English (US)	The user wants to find a restaurant to visit in San Francisco	Generic search results page: http://search.yahoo.com/search?p=san+francisco+restaurants	Slightly Relevant	We will assume a user has just done the search on a search engine like Yahoo, so this page has low utility.
[Pizza Hut in Chicago], English (US)	The user wants to find Pizza Hut locations in Chicago.	Maps search results page: http://local.yahoo.com/results;_ylc=X3oDMTEwNTByOW5sBF9TAzI3MTYxNDkEc2VjA2ZwLXRhYgRzbGsDc3Bpcml0?p=pizza+hut&csz=chicago&fr=yfp-t-501&toggle=1&cop=mss&ei=UTF-8	Useful	The page provides a list of Pizza Hut locations in Chicago. There is contact information for every restaurant, as well as a map that displays their locations.

Query	Reasonable User Intent	URL of the Landing Page	Rating	Reason
[jumping shark], English (US)	The user wants information about the term “jumping the shark” that was used on several tv shows; the user wants to see images or videos of a jumping shark.	Video search results page: http://video.search.yahoo.com/search/video?p=jumping+shark&fr=yfp-t-501&toggle=1&cop=mss&ei=UTF-8	Relevant or Useful	This page displays many clickable thumbnails to videos for both user intents.
[Tom Cruise], English (US)	The user wants news or information about Tom Cruise; the user wants to see images or videos of Tom Cruise.	Images search results page: http://images.search.yahoo.com/search/images?p=tom%20cruise&fr=yfp-t-501&toggle=1&cop=mss&ei=UTF-8&tnr=20	Relevant or Useful	This page displays many high quality images of Tom Cruise to look at.
[Tom Cruise], English (US)	The user wants news or information about Tom Cruise; the user wants to see images or videos of Tom Cruise.	Book search results page: http://books.google.com/books?id=5omJUX4YRuEC&pg=PA110&dq=tom+cruise&lr=&sig=ACfU3U3ARUVs1nb-D_7-0Mr44Pq8G_pyOg	Slightly Relevant	This page from a book about mountaineering mentions Tom Cruise. It is unlikely that this page is a good result for the user intent.
[functional programming], English (US)	The user wants information about functional programming.	Scholar search results page: http://scholar.google.com/scholar?q=functional+programming&hl=en&lr=&btnG=Search	Relevant or Useful	This page displays links to many articles in academic journals on functional programming.
[wii console], English (US)	The user wants to buy a wii console.	Shopping search results page: http://www.google.com/products?q=wii+console&sourceid=navclie&nt-ff&ie=UTF-8&rls=GGGL:GGGL:2006-33.GGGL:en&um=1&sa=N&tab=wf	Useful	This page provides links to merchants from which to buy this item. Prices and seller ratings are displayed.

5.3.7 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 always think about user intent and what pages are good for the user. When a video is in a foreign language, but you think that it is what the user is looking for or might expect to see, you should assign a rating from the rating scale. If you can't tell if the video is related to the query, a rating of **Unratable: Foreign Language** is appropriate.

Please note: When evaluating the language of the landing page on a video hosting website such as YouTube or Metacafe, please look at the language of the navigation and other features, not the language of the metadata or any comments written by viewers.

Here are some rules to follow for some specific cases:

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, rating is very straightforward. You will easily be able to

assign a utility rating and a Language Page Language flag. Questions arise, however, when the query and/or one or more components of the landing page is in a foreign language. Here are some rules to follow:

- If the query is the title of a foreign language song and the landing page has a video of the song in a foreign language, please rate the page based on the relationship between the query and the video, and use the language flag that reflects the written language on the page.
- If the query is the name of a foreign band or foreign artist and the landing page has a music video of the band or artist performing in a foreign language, please rate the page based on the relationship between the query and the video, and use the language flag that reflects the written language on the page.
- If the query is the name of a foreign film and the landing page shows clips from the film in a foreign language, please rate the page based on the relationship between the query and the video, and use the language flag that reflects the written language on the page.
- If the query is the name of a foreign actor and the landing page shows clips from one of the actor’s films in a foreign language, please rate the page and based on the relationship between the query and the video, and use the language flag that reflects the written language on the page.
- If the query is the name of a foreign film or the name of a foreign actor and the landing page has a video **about** the film or **about** the actor, and the people in the video are speaking about the film or actor in a foreign language, please assign **Unratable: Foreign Language**, since you can’t understand what the people are saying. Please use the language flag that reflects the written language on the page.
- If the query is a sport or event and the landing page displays a video pertaining to the sport or event with music using foreign language lyrics, please rate the page based on the relationship between the query and the video, and use the language flag that reflects the written language on the page. The language in the music is not important.
- If the query is a sport or event, or some other type of event and the landing page has a video **about** the sport or **about** the event, and the people in the video are speaking about the sport or event in a foreign language, please assign **Unratable: Foreign Language**, since you can’t understand what the people are saying. Please use the language flag that reflects the written language on the page.

Please note: When evaluating the language of the landing page on a video hosting website such as YouTube or Metacafe, please look at the language of the navigation and other features, not the language of the metadata or any comments written by viewers.

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=fprt0Kxtnfk	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-lw&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

Query	URL of the Landing Page	Description of the Landing Page	Rating	Landing Page Language
[beatles live], English (US)	http://www.youtube.com/watch?v=1eyBhagx2U&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
[beatles live], English (US)	http://www.youtube.com/watch?v=Ou_mlGfimU	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	English
[judy ann santos], English (US)	http://www.youtube.com/watch?v=YO2Iur55_zk	The query is for the popular Philippines actress, Judy Ann Santos. The landing page has a clip from one of her movies.	Slightly Relevant or Relevant	English
[Kasal, Kasali, Kasalo], English (US)	http://www.youtube.com/watch?v=d4IbSU8dW_g	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
[sri lanka cricket], English (US)	http://www.youtube.com/watch?v=Valf_cgF84k	The query is looking for information about Sri Lanka cricket. The landing page has a video which displays cricket in Sri Lanka. There is no speaking in the video, but there is a song sung in a foreign language.	Relevant	English

5.4 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 U.S. 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	Reasonable User Intent	Landing Page	Rating	Reason
[Bridget Jones's Diary], English (US)	Buy a copy of this book or movie	http://www.amazon.com/Bridget-Jones-Diary-Helen-Fielding/dp/014028009X	Useful	This is a good result and the task location matches the query location.
[Bridget Jones's Diary], English (US)	Buy a copy of this book or movie	http://www.amazon.co.uk/Bridget-Jones-Diary-Helen-Fielding/dp/0330375253	Slightly Relevant	For most book purchases, U.S. users would use the U.S. Amazon site. The U.K. site gives prices in pounds, not dollars, and shipping to the U.S. is expensive.

Query	Reasonable User Intent	Landing Page	Rating	Reason
[Cheesecake recipe], English (US)	Find a cheesecake recipe to make	http://allrecipes.com/Recipe/New-York-Cheesecake-III/Detail.aspx	Relevant	The ingredients and measurements are familiar to U.S. residents.
[Cheesecake recipe], English (US)	Find a cheesecake recipe to make	http://www.bbc.co.uk/food/recipes/database/whitechocolatecheese_12412.shtml	Slightly Relevant	The measurements are in metrics and the ingredients are British. Few U.S. residents could make this cake, so there is very little utility.
[human rights violations], English (US)	Find examples or information about human rights violations	http://www.hrw.org/doc?t=chechnya – page about human rights violations in Chechnya on a reputable website http://en.wikipedia.org/wiki/Human_rights_in_the_People's_Republic_of_China - Wikipedia page on human rights violations in China http://www.hrw.org/reports/2007/us0507/ - page about human rights violations at Wal-Mart in the U.S. on a reputable website	Relevant or Useful Relevant or Useful Relevant or Useful	Human rights violations happen around the world in many countries. Most people in the U.S. would be interested in international human rights violations. For this query, results about countries other than the U.S. are just fine. Use your common sense to decide what a user in your location would be interested in.
[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	Slightly Relevant or Off-Topic	For most washing machine purchases, U.S. users would shop in the U.S. It is too expensive to purchase a washing machine in the U.K. and pay to ship it to the U.S., so there is very little or no utility.

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 3 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 deceptive web design techniques, as 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** 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** 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	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	Yes
[how tall is a camel], English (US)	Find answer to this question about camels	<a href="http://www xnxx.com/free/cameltoe-
pictures.php">http://www xnxx.com/free/cameltoe- pictures.php Warning – this page is porn	Off-Topic	Yes
[car pictures], English (US)	Find pictures of cars	<a href="http://www.pbase.com/vanjog/what_life_cont
ain_nudty">http://www.pbase.com/vanjog/what_life_cont ain_nudty Warning – this page is porn	Off-Topic	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 the user is 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.thspankingnews.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 which 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	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	Yes

Query	User Intent	Landing Page	Rating	Porn Flag?
[jenna jameson], English (US)	Find porn pictures of Jenna Jameson or navigate to her official website.	http://tours.clubjenna.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.pornstargals.com/ Warning – this page is porn	Off-Topic	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

6.3 Malicious Flag

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

- You are forced to quit your 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.

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 browser.

Part 2: Rating Examples

In this section, you will see examples of some of the types of queries and landing pages you are likely to encounter, along with suggested ratings. Keep in mind that most queries can be categorized as action, information, or navigation (“do-know-go”), but that many queries fall into more than one category. As you work on URL rating tasks, remember that you must always consider user intent and page utility.

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.)
- Titles of 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	<ul style="list-style-type: none">▪ <i>John McCain is a United States Senator. He is a very well-known politician and there are many pages on the Web about him.</i>
User Intent	<ul style="list-style-type: none">▪ <i>Navigation – The user wants to navigate to an official homepage for John McCain.</i>▪ <i>Information – The user wants information or news about John McCain.</i>
Vital	<ul style="list-style-type: none">▪ John McCain’s official government Senate homepage: http://mccain.senate.gov/▪ John McCain’s official MySpace page: http://www.myspace.com/johnmccain▪ John McCain’s official YouTube page: http://www.youtube.com/johnmccain
Useful	<ul style="list-style-type: none">▪ A comprehensive information page, such as http://en.wikipedia.org/wiki/John_McCain▪ An article with biographical information about John McCain and his complete Senate voting record at http://projects.washingtonpost.com/congress/members/m000303/
Relevant	<ul style="list-style-type: none">▪ A comprehensive information page about John McCain’s <i>father</i>, who was also named John McCain, such as http://en.wikipedia.org/wiki/John_S._McCain%2C_Jr.▪ A page with at least a paragraph of timely information about John McCain▪ A good video with John McCain in it, such as http://www.youtube.com/watch?v=53caXQKTs9Y▪ A page on which to buy a book written by John McCain, such as http://www.amazon.com/Worth-Fighting-John-S-McCain/dp/0375505423▪ A high quality page about an ordinary person named John McCain
Slightly Relevant	<ul style="list-style-type: none">▪ A page about a tax bill proposed by John McCain and another senator in 2003: http://www.nationalcenter.org/TSR102103.html▪ 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▪ A very brief article about an ordinary person named John McCain

	[Nicole Kidman] , English (US)
Query Description	<i>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.</i>
User Intent	<ul style="list-style-type: none"> Navigation – The user wants to navigate to Nicole Kidman’s official homepage. Information – The user wants information, news, video clips, pictures, etc. related to Nicole Kidman.
Vital	<ul style="list-style-type: none"> Nicole Kidman’s official homepage, if one exists. Please be aware that some unofficial sites for celebrities may claim to be official.
Useful	<ul style="list-style-type: none"> Pages that are comprehensive resources for Nicole Kidman, such as http://www.imdb.com/name/nm0000173/. Such pages might include a biography, filmography, pictures, etc. A high quality personal fan page might even be rated Useful. A page of images of Nicole Kidman, such as http://images.google.com/images?q=nicole+kidman+pictures&hl=en&safe=off&rls=GGGL,GGGL:2006-33.GGGL:en&um=1&ie=UTF-8&sa=X&oi=images&ct=title
Relevant	<ul style="list-style-type: none"> A page with at least a paragraph of timely information about Nicole Kidman A video of Nicole Kidman in an ad for Chanel: http://www.youtube.com/watch?v=yTO4FHf8MBs
Slightly Relevant	<ul style="list-style-type: none"> 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	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 and should be assigned a Spam flag: http://www.nicolekidman.org .

	[Erica Hill], English (US)
Query Description	<ul style="list-style-type: none"> <i>Erica Hill is a news anchor for CNN. She has 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 John McCain or Nicole Kidman.</i> <i>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.</i>
User Intent	<ul style="list-style-type: none"> Navigation – The user wants to navigate to an official homepage for Erica Hill, the CNN news anchor. Information – The user wants information or news about Erica Hill, the CNN news anchor.
Vital	<ul style="list-style-type: none"> Erica Hill’s page on the CNN website: http://www.cnn.com/CNN/anchors_reporters/hill.eric.html
Useful	<ul style="list-style-type: none"> “Getting to Know Erica Hill of CNN”, a YouTube video starring Erica Hill, in which she shares a lot of personal and professional information: http://www.youtube.com/watch?v=Ouv2Cydw_YE Quality pages with biographical or good general information on Erica Hill, the CNN news anchor, such as http://en.wikipedia.org/wiki/Erica_Hill The homepage for someone named “Erica Hill”, other than the CNN news anchor, whose homepage would be of interest to a significant number of people
Relevant	<ul style="list-style-type: none"> Any page with at least a paragraph about the CNN news anchor, such as http://www.mediabistro.com/tvnewser/cnn/erica_hill_moving_from_hln_to_cnn_75438.asp The homepage for someone named “Erica Hill”, other than the former CNN news anchor, which would be of interest to a small number of people
Slightly Relevant	<ul style="list-style-type: none"> Lower quality pages about the former CNN news anchor, such as http://www.biocrawler.com/encyclopedia/Erica_Hill The Friendster page of someone named Erica Hill, http://profiles.friendster.com/209275
Off-Topic	<ul style="list-style-type: none"> Pages with the words “Erica” or “Hill” scattered on them, such as this softball box score page that mentions players named <i>Erica</i> Douglas and <i>Sam Hill</i>, http://www.millsaps.edu/athletic/softball/mcs-0001.htm

[A O Smith] , English (US)	
Query Description	A.O. Smith is a company that makes electric motors, water heaters & storage tanks.
User Intent	<ul style="list-style-type: none"> Navigation – the user wants to navigate to the official homepage for the company Action – the user wants to purchase products manufactured by the company
Vital	<ul style="list-style-type: none"> Corporate homepage for A.O. Smith http://www.aosmith.com/
Useful	<ul style="list-style-type: none"> A.O. Smith division webpages at http://www.aosmithmotors.com/ and http://www.hotwater.com/ High quality websites about the A.O. Smith Corp., or sites that sell, distribute, or review A.O. Smith products. A page of current news articles about A.O. Smith
Relevant	<ul style="list-style-type: none"> Helpful subpages on the A.O. Smith website, such as the webpage for investors at http://investor.shareholder.com/aosmith/ Page with many A.O. Smith products for sale, http://www.cshincorporated.com/index.php/cPath/57
Slightly Relevant	<ul style="list-style-type: none"> Outdated article about the A.O. Smith company Less important subpages on the A.O. Smith website, such as: http://www.aosmith.com/governance/financialcode.htm Amazon 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 so unlikely that this page would be helpful to the user who typed the query, Off-Topic is also an acceptable rating.
Off-Topic	<ul style="list-style-type: none"> Article about a singer named Elliott Smith, who was scheduled to perform at a dance called the “A&O Ball”. 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.
User Intent	<ul style="list-style-type: none"> Navigation – the user wants to navigate to the official homepage for the company Information – the user wants information about the store Action – the user wants to make a purchase
Vital	<ul style="list-style-type: none"> Official homepage at http://www.forotherlivingthings.com/
Useful	<ul style="list-style-type: none"> Directory pages with contact information, a map, and reviews about the store, such as: http://www.insiderpages.com/b/3710275102 or http://local.yahoo.com/info-21336044-for-other-living-things-sunnyvale
Relevant	<ul style="list-style-type: none"> Helpful subpages on the website, such as: Contact information: http://www.forotherlivingthings.com/contact_us.php Company information: http://www.forotherlivingthings.com/about_us.php Product information: http://www.forotherlivingthings.com/all-products-c-142.html A directory page with contact information: http://www.zvents.com/sunnyvale-ca/venues/show/125217-for-other-living-things
Slightly Relevant	<ul style="list-style-type: none"> Less important subpages on the website, such as: http://www.forotherlivingthings.com/privacy.php A page about guinea pigs that mentions the For Other Living Things store and has a link: http://community.babycenter.com/journal/wheekergal/685/are_guinea_pigs_the_right_pet_for_your_kids
Off-Topic	<ul style="list-style-type: none"> Page containing 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	<i>There are many companies and people with the name Perkins</i>
User Intent	<ul style="list-style-type: none"> Navigation – the user wants to go to the official homepage of the Perkins Restaurant & Bakery chain, the dominant interpretation Navigation – the user wants to go to the official homepage of another entity with the name Perkins
Vital	<ul style="list-style-type: none"> Official homepage of Perkins Restaurant & Bakery at http://www.perkinsrestaurants.com/
Useful	<ul style="list-style-type: none"> Official homepages of common interpretations for this query, such as: http://perkins.com, homepage of Perkins Engines http://www.perkins.org/, homepage of Perkins School for the Blind
Relevant	<ul style="list-style-type: none"> Helpful subpages on the Perkins Restaurant website: <ul style="list-style-type: none"> http://www.perkinsrestaurants.com/locations, locations subpage http://www.perkinsrestaurants.com/menu, menu subpage Official homepages of less common or minor interpretations, such as: <ul style="list-style-type: none"> http://www.perkinsmedicalsupply.com/, homepage of Perkins Medical Supply, a small company in Vero Beach, Florida http://www.ed.gov/programs/fpl/index.html, homepage of the Federal Perkins Loan Program, which provides student loans to low-income students A recent news article about the Perkins restaurant chain The homepage of someone whose last name is Perkins
Slightly Relevant	<ul style="list-style-type: none"> Less important subpages on Perkins Restaurant website, such as http://www.perkinsrestaurants.com/privacy Outdated news article about the Perkins Restaurant chain
Off-Topic	<ul style="list-style-type: none"> Video of a birthday party at a Perkins Restaurant: http://www.youtube.com/watch?v=TZuvYSOsHug

	[iphone], English (US)
Query Description	<i>The iPhone is a popular mobile smartphone made by Apple.</i>
User Intent	<ul style="list-style-type: none"> Navigation – the user wants to navigate to the product webpage on the Apple website Information – the user wants information (reviews, specifications, features, etc.) on this phone Action – the user wants to purchase this phone
Vital	<ul style="list-style-type: none"> Apple's webpage for the iPhone: http://www.apple.com/iphone/
Useful	<ul style="list-style-type: none"> High quality sites that review or provide comprehensive information on the iPhone, such as http://reviews.cnet.com/smartphones/apple-iphone-8gb-at/4505-6452_7-32309245.html The AT&T FAQ page on the iPhone at http://www.wireless.att.com/learn/iphone-faqs.jsp. AT&T is currently the only carrier that provides service for the iPhone users in the U.S. The AT&T webpage to buy the iPhone: http://www.wireless.att.com/cell-phone-service/specials/iPhone.jsp The Apple iPhone discussion board: http://discussions.apple.com/category.jspa?categoryID=201
Relevant	<ul style="list-style-type: none"> Page with many iPhone many accessories for sale A timely article about the iPhone. A helpful video on the iPhone, such as http://video.google.com/videoplay?docid=7063257898235967577 A Wikipedia article on the iPhone, http://en.wikipedia.org/wiki/Iphone
Slightly Relevant	<ul style="list-style-type: none"> Review on the HTC Touch phone that mentions the iPhone, http://reviews.cnet.com/smartphones/htc-touch-unlocked/4505-6452_7-32465595.html Outdated article on the iPhone Webpage on the Apple Mac Pro, 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.
Off-Topic	<ul style="list-style-type: none"> Webpage about a different type of smartphone, such as http://www.sonyericsson.com/cws/products/mobilephones/overview/p990i

	[Honda Pilot], English (US)
Query Description	<i>The Pilot is a popular Honda SUV.</i>
User Intent	<ul style="list-style-type: none"> Navigation – the user wants to navigate to the product webpage on the Honda website Information – the user wants information (reviews, specifications, features, etc.) about this SUV Action – the user wants to purchase this SUV
Vital	<ul style="list-style-type: none"> Honda's webpage for the Pilot: http://automobiles.honda.com/pilot/
Useful	<ul style="list-style-type: none"> High quality sites that review or provide comprehensive information about the current model of the Honda Pilot, such as http://www.edmunds.com/honda/pilot/review.html The Insurance Institute for Highway Safety (IIHS) page about the Honda Pilot: http://www.iihs.org/ratings/ratingsbyseries.aspx?id=391. Relevant would also be acceptable.
Relevant	<ul style="list-style-type: none"> High quality sites with comprehensive information about other year models of the Honda Pilot, such as: http://autos.aol.com/honda-pilot-2007:8689-overview. If the information is more than a few years old, Slightly Relevant is also appropriate. A timely article about the Honda Pilot. A Wikipedia article on the Honda Pilot, http://en.wikipedia.org/wiki/Honda_Pilot
Slightly Relevant	<ul style="list-style-type: none"> Shopping page for Pilot headlights and fog lights: http://shopping.yahoo.com/s:Headlights:4168-Brand=Pilot Amazon.com page with Honda Pilot repair manual for sale: http://www.amazon.com/Honda-Pilot-Acura-MDX-Haynes/dp/1563926903
Off-Topic	<ul style="list-style-type: none"> High quality webpage about the Honda Civic: http://www.edmunds.com/honda/civic/review.html

	[Chicago], English (US)
Query Description	<i>Chicago is a big city in the United States.</i>
User Intent	<ul style="list-style-type: none"> Navigation – the user wants to navigate to the official Chicago government website Action – the user wants to make travel plans and reservations Information – the user wants travel and tourism information or general information about Chicago <p><i>When a city is a major travel destination, it is likely that the query is looking for information. However, if the city (or county, state, country, etc.) has an official government webpage, that page should get a Vital rating.</i></p>
Vital	<ul style="list-style-type: none"> The official homepage for the city of Chicago http://egov.cityofchicago.org/
Useful	<ul style="list-style-type: none"> A comprehensive set of hotel listings or other travel & tourism information that offers utility to the user, such as http://www.choosechicago.com/ High quality pages about Chicago: its history, climate, travel, culture, public transportation, etc., http://www.lonelyplanet.com/worldguide/usa/chicago An excellent blog or collection of personal information, which would offer utility to someone visiting the city, such as http://www.gochicagocard.com/blog/ A comprehensive collection of high quality images of the city of Chicago, http://images.google.com/images?q=chicago&sourceid=navclient-ff&ie=UTF-8&rls=GGGL.GGGL:2006-33.GGGL:en&um=1&sa=N&tab=wi A high quality map of the city, such as http://travel.yahoo.com/p-map-191501928-map_of_chicago_il-i Official homepage of Chicago, the band, http://www.chicagotheband.com/

	[Chicago], English (US)
Relevant	<ul style="list-style-type: none"> The homepage for the main regional newspaper, Chicago Tribune, at http://www.chicagotribune.com/. The homepages of large, prominent entities associated with the word 'Chicago' in the minds of many people, 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. Video of the band "Chicago" performing in concert, such as http://video.google.com/videoplay?docid=7153058075771593413&q=chicago+video&total=145976&start=0&num=10&so=0&type=search&plindex=7
Slightly Relevant	<ul style="list-style-type: none"> Local weather forecasts for Chicago, http://www.wunderground.com/US/IL/Chicago.html Homepages of universities or businesses in the Chicago area that are not as closely associated with the city, such as Northwestern University, http://www.northwestern.edu/ Homepages of other newspapers that cover the Chicago area, but are not the main 'newspaper of the city', such as http://www.chicagoweeklynews.com/
Off-Topic	<ul style="list-style-type: none"> Webpage of the summer music program at Northwestern University (a university located just outside Chicago), http://www.music.northwestern.edu/summer/ Video of the Blue Brothers performing the song, "Sweet Home Chicago", http://www.youtube.com/watch?v=FcPmBKl_rB8

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 the user. For example, a hotel affiliate page with a list of Chicago hotels may be assigned a rating **Relevant**, but also receive a **Spam** flag.

	[Nevada], English (US)
Query Description	<i>Nevada is one of the states in the United States.</i>
User Intent	<ul style="list-style-type: none"> <i>Navigation – the user wants to navigate to the official Nevada government website</i> <i>Information – the user wants general information about Nevada or travel and tourism information</i>
Vital	<ul style="list-style-type: none"> The official homepage for the state of Nevada: http://www.nv.gov/
Useful	<ul style="list-style-type: none"> The state of Nevada's official travel and tourism website: http://travelnevada.com/ High quality, comprehensive pages about Nevada: http://en.wikipedia.org/wiki/Nevada High quality travel and tourism pages for Nevada, such as http://travel.yahoo.com/p-travelguide-191501966-nevada_vacations-i_ylc=X3oDMTFka28zOGNuBF9TAzI3NjY2NzkEX3MDOTY5NTUzMjUEc2VjA3NzcC1kZXN0BHNsawN0aXR sZQ--
Relevant	<ul style="list-style-type: none"> The 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/ Pages with facts about Nevada: http://www.leg.state.nv.us/general/FACTS.cfm and http://www.nv.gov/new_KidsHomework.htm Wikipedia page with links to other Wikipedia pages for many Nevada cities: http://en.wikipedia.org/wiki/List_of_cities_in_Nevada
Slightly Relevant	<ul style="list-style-type: none"> IMDB page for a movie titled "Nevada Smith": http://www.imdb.com/title/tt0060748/. Off-Topic is also acceptable. Homepage of the Nevada Republican Party: http://www.nevadagop.org/ Outdated article about an election in Nevada.
Off-Topic	<ul style="list-style-type: none"> Homepage for the UCMT Family of Schools, which has massage therapy schools in Utah, Nevada, Arizona, and Colorado: http://www.ucmt.com/

	[white house] , English (US)
Query Description	The residence of the President of the United States is called the White House.
User Intent	<ul style="list-style-type: none"> ▪ Navigation – the user wants to navigate to http://www.whitehouse.gov. ▪ Information – the user wants information about the White House
Vital	http://www.whitehouse.gov
Useful	<p>The President’s webpage on the official U.S. government White House website at http://www.whitehouse.gov/president/.</p> <p>Other important subpages on the official White House website, such as:</p> <ul style="list-style-type: none"> ▪ Contact Information subpage: http://www.whitehouse.gov/contact/ ▪ News subpage: http://www.whitehouse.gov/news/ ▪ Daily Briefings subpage: http://www.whitehouse.gov/news/briefings/ <p>The Wikipedia page on the White House, which has a link to the official website: http://en.wikipedia.org/wiki/White_House</p> <p>The homepages of common or minor interpretations:</p> <ul style="list-style-type: none"> ▪ Homepage of the White House Restaurant in Laguna Beach, California : http://www.whitehouserestaurant.com/ ▪ Homepage of a city in the state of Tennessee: http://www.cityofwhitehouse.com/ ▪ Homepage of an inn called White House: http://www.whitehouseinn.com/
Relevant	<p>Most relatively important subpages on the official U.S. government White House website, such as:</p> <ul style="list-style-type: none"> ▪ Kids subpage: http://www.whitehouse.gov/kids/ ▪ History and Tours subpage: http://www.whitehouse.gov/history/life/video/index.html <p>Important subpages on the official website of common or minor interpretations:</p> <ul style="list-style-type: none"> ▪ City government subpage on the official White House, Tennessee website: http://www.cityofwhitehouse.com/index.asp?NID=62 ▪ “About the Inn” subpage on the official White House inn website: http://www.whitehouseinn.com/about.html
Slightly Relevant	<p>Unimportant subpages on the official U.S. government White House website, such as the 2007 Easter Egg Roll page at http://www.whitehouse.gov/easter/2007/</p> <p>Unimportant subpages on the official website of common or minor interpretations, such as:</p> <ul style="list-style-type: none"> ▪ Citizen Request Tracker subpage on the official White House, Tennessee website: http://www.cityofwhitehouse.com/requesttracker.asp ▪ Partner Sites subpage on the official White House inn website: http://www.whitehouseinn.com/partner_sites.html <p>Homepages of very minor interpretations:</p> <ul style="list-style-type: none"> ▪ Homepage of The White House Federal Credit Union: http://www.whcu.org/index.aspx ▪ Homepage of the Eisenhower House, which was the summer white house while Eisenhower was president: http://www.whitehouseinn.com/partner_sites.html
Off-Topic	<ul style="list-style-type: none"> ▪ A page about removing white house paint from brown boots: http://www.answerbag.com/q_view/507910

[whitehouse.gov] , English (US)	
Query Description	<i>This is a special type of query, which we refer to as a URL query. This URL is the official homepage for the White House.</i>
User Intent	<ul style="list-style-type: none"> • Navigation – the user wants to navigate to http://www.whitehouse.gov.
Vital	<ul style="list-style-type: none"> • http://www.whitehouse.gov
Useful	<ul style="list-style-type: none"> • The President's webpage on the official White House website at http://www.whitehouse.gov/president/ , which is very similar to the White House homepage, and possibly matches the user intent.
Relevant	<p>Any of the following relatively important webpages on the official White House website. Many other subpages on the official website could also be considered Relevant.</p> <ul style="list-style-type: none"> • Contact Information subpage: http://www.whitehouse.gov/contact/ • News subpage: http://www.whitehouse.gov/news/ • Daily Briefings subpage: http://www.whitehouse.gov/news/briefings/ • Kids subpage: http://www.whitehouse.gov/kids/ • History and Tours subpage: http://www.whitehouse.gov/history/life/video/index.html
Slightly Relevant	<ul style="list-style-type: none"> • The Wikipedia page on the White House, which has a link to the official website: http://en.wikipedia.org/wiki/White_House • Unimportant subpages on the official website, such as the 2007 Easter Egg Roll page at http://www.whitehouse.gov/easter/2007/
Off-Topic	<ul style="list-style-type: none"> • 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, the user is 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: the user wants 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 survey
- Take an online personality test

	[adobe reader download], English (US)
Query Description	<i>Adobe Reader software allows the user to view and print PDF files.</i>
Likely User Intent	<ul style="list-style-type: none"> ▪ Action – the user wants to download Adobe Reader
Vital	<ul style="list-style-type: none"> ▪ The page on the Adobe site on which to download Adobe Reader: http://www.adobe.com/products/acrobat/readstep2.html
Relevant	<ul style="list-style-type: none"> ▪ The Adobe homepage: The Adobe homepage: http://www.adobe.com/ ▪ A page on a reputable website that has 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
Slightly Relevant	<ul style="list-style-type: none"> ▪ A Yahoo! Answers page that tells what Adobe Reader does, and which has a link to Adobe: http://answers.yahoo.com/question/index?qid=1005111000036
Off-Topic	<ul style="list-style-type: none"> ▪ A page about the Omea Reader, a free RSS reader: http://www.jetbrains.com/omea/reader/

	[text twist], English (US)
Query Description	<i>TextTwist is a popular computer game that can be played online or downloaded.</i>
Likely User Intent	<ul style="list-style-type: none"> ▪ Action – the user wants to play the game online or download it (for free or for a fee)
Vital	<ul style="list-style-type: none"> ▪ None possible
Useful	<ul style="list-style-type: none"> ▪ A page on which a user can play or download the game, such as http://get.games.yahoo.com/proddesc?gamekey=texttwist
Relevant	<ul style="list-style-type: none"> ▪ An article which contains tips for playing the game, such as http://videogames.lovetoknow.com/wiki/Text_Twist_Tips_and_Strategies
Off-Topic	<ul style="list-style-type: none"> ▪ A page on which to download Tetris, a different computer game.

	[take an online personality test], English (US)
Query Description	<i>Personality tests help people to understand their behavior and can help them learn what type of career they might be suited for</i>
Likely User Intent	<ul style="list-style-type: none"> Action – the user wants to take an online personality test for free or for money
Vital	<ul style="list-style-type: none"> None possible
Useful	<ul style="list-style-type: none"> An online personality test, based on the famous Myers-Briggs Type Indicator that identifies 16 distinct personality types, such as http://www.humanmetrics.com/cgi-win/Jtypes2.asp or http://www.kisa.ca/personality/
Relevant	<ul style="list-style-type: none"> A very short online personality test, based on the famous Myers-Briggs personality test, at http://www.personalitytype.com/quiz.html The website of a company that offers the Myers-Briggs Type Indicator online for a fee, and offers clients many kinds of reports based on test results. The company's clients include many well-known U.S. corporations. http://www.knowyourtype.com/
Slightly Relevant	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<i>There are videos on the web of dogs using skateboards</i>
Likely User Intent	<ul style="list-style-type: none"> Action – the user wants to watch a video of a skateboarding dog
Vital	<ul style="list-style-type: none"> None possible
Useful	<ul style="list-style-type: none"> YouTube and Metacafe pages that have videos of talented skateboarding dogs, 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/ Tyson's homepage that displays a video of him riding his skateboard: http://www.skateboardingbulldog.com/ Tyson's MySpace homepage that displays a video of him on his skateboard: http://www.myspace.com/tysonskateboardingbulldog
Relevant	<ul style="list-style-type: none"> Videos of less talented dogs using skateboards: http://www.metacafe.com/watch/925757/barney_the_skateboarding_dog/ http://uk.youtube.com/watch?v=nhE9Y1tEwQw&NR=1 http://uk.youtube.com/watch?v=tlx-AdIR7ew
Slightly Relevant	<ul style="list-style-type: none"> A video of a skateboarding dog made out of clay: http://www.youtube.com/watch?v=WVUoTiqp7qo
Off-Topic	<ul style="list-style-type: none"> A video of a dog doing other amazing tricks, but not skateboarding, such as http://www.videojug.com/film/lord-of-dogtown-buddy-the-amazing-surfing-dog http://video.google.com/videoplay?docid=5202848730472933222&q=dog+water+skiing&total=70&start=0&num=10&so=0&type=search&plindex=5

3.0 Information Queries

When typing an information query, the user is trying to find information. These are “**know**” queries: the user wants to know something. For many information queries, it would be difficult to imagine a user intent other than looking for information. Here are some examples of information queries.

Please note that for several of these information queries, a page exists that warrants a rating of **Vital**. The user is looking to find information and these **Vital** pages give the user exactly what he is looking for on the official, authoritative website associated with the query. Even when user intent is information and the information can be found on many pages on the Web, a **Vital** rating is sometimes possible.

	[retina and laser surgery], English (US)
Query Description	<i>Laser surgery can be performed on the retina to treat a variety of retinal problems.</i>
User Intent	<ul style="list-style-type: none"> Information – the user wants information about laser surgery for the retina
Vital	<ul style="list-style-type: none"> None possible
Useful	<ul style="list-style-type: none"> Pages from high quality sources providing information on laser surgery for the retina, http://www.nei.nih.gov/health/retinaldetach/index.asp Newsgroups which are focused on the subject and provide utility to the user
Relevant	<ul style="list-style-type: none"> Individual retinal laser surgery practitioner websites that provide information on the topic, such as http://www.socalretina.com/html/procedures.html Wikipedia page on eye surgery, which discusses many types of eye surgery, including laser retina surgery, http://en.wikipedia.org/wiki/Eye_surgery Yahoo! Answers page that discusses this topic, http://au.answers.yahoo.com/answers2/frontend.php/question?qid=20070724160757AAHmLJy Article on diabetic retinopathy that discusses laser treatment, http://www.solomoneyeassociates.com/procedures/diabetic_eye_treatment.htm
Slightly Relevant	<ul style="list-style-type: none"> Forum that discusses fellowships, training programs, and the lifestyle of doctors performing this type of surgery, http://forums.studentdoctor.net/archive/index.php/t-67862.html
Off-Topic	<ul style="list-style-type: none"> Sites about cosmetic laser surgery on other parts of the body, not including the eyes, such as http://www.fda.gov/FDAC/features/2000/300_laser.html Sites about 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	<i>Used coffee grounds do not need to be thrown away; there are many uses for them.</i>
User Intent	<ul style="list-style-type: none"> Information – the user wants information about uses for coffee grounds
Vital	<ul style="list-style-type: none"> None possible
Useful	<ul style="list-style-type: none"> Pages (including FAQs and newsgroup 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> Online directory listing for a restaurant called “Coffee Grounds” in Tempe, Arizona, http://phoenix.citysearch.com/profile/1701833/tempe_az/coffee_grounds.html

	[HTML lessons], English (US)
Query Description	<i>HTML stands for HyperText Markup Language, the markup language for the creation of most webpages.</i>
User Intent	<ul style="list-style-type: none"> Information – the user wants pages that provide instructions for learning HTML.
Vital	<ul style="list-style-type: none"> None possible
Useful	<ul style="list-style-type: none"> Pages that offer lessons, step-by-step instructions, or tutorials for learning HTML, such as http://www.utexas.edu/learn/html/ or http://www.w3schools.com/html/default.asp
Relevant	<ul style="list-style-type: none"> Pages that offer short tutorials on using HTML
Slightly Relevant	<ul style="list-style-type: none"> A Wikipedia page with good information about HTML and links to tutorial pages: http://en.wikipedia.org/wiki/HTML
Off-Topic	<ul style="list-style-type: none"> Pages that offer lessons or tutorials for learning XML, not HTML, such as http://www.w3schools.com/xml/default.asp An article that discusses HTML 5, a major upgrade to HTML, 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	<i>South Beach is a section of Miami Beach, Florida. Collins Avenue is a major street that runs through Miami Beach.</i>
User Intent	<ul style="list-style-type: none"> Information – The user needs a map of South Beach that displays Collins Avenue.
Vital	<ul style="list-style-type: none"> None possible
Useful	<ul style="list-style-type: none"> Map that shows the South Beach area of Miami Beach, and identifies Collins Avenue, such as http://www.miamibeach411.com/maps_south_beach.html or http://www.qmcb.com/visitors/SouthBeachMap.pdf
Slightly Relevant	<ul style="list-style-type: none"> Map that shows the South Beach area of Miami Beach, but does not identify Collins Avenue, http://miami.citysearch.com/profile/map/11344117/miami_beach_fl/south_beach.html Wikipedia page on 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	<ul style="list-style-type: none"> Map finder page in which you can type “□Collins ave, south beach, fl” and get a map of the area, such as http://maps.yahoo.com/

	[international telephone codes], English (US)
Query Description	<i>Every country has a country calling code (dialing prefix) that is dialed before the telephone number when calling that country.</i>
User Intent	<ul style="list-style-type: none"> Information - The user wants a list of the country calling codes
Vital	<ul style="list-style-type: none"> None possible
Useful	<ul style="list-style-type: none"> Pages that provide a comprehensive set of international calling codes, such as http://www.thephonebook.bt.com/publisha.content/en/search/international/index.publisha or http://en.wikipedia.org/wiki/List_of_country_calling_codes A page that describes how to dial an international call and provides a link to a page with a list of country calling codes, http://www.wiktel.com/standards/howdial.htm
Relevant	<ul style="list-style-type: none"> Pages with international telephone codes, but for Europe only, http://www.europe.org/dialingcodes.html
Slightly Relevant	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> A page with areas codes for the U.S. and Canada, http://www.consumer.att.com/global/english/usa_codes.html

	[enable javascript ie], English (US)
Query Description	"ie" is an abbreviation for Internet Explorer, which is Microsoft's Web browser.
User Intent	<ul style="list-style-type: none"> Information - The user wants to learn how to enable JavaScript in Internet Explorer.
Vital	<ul style="list-style-type: none"> The page from Microsoft's website that tells you how to enable JavaScript in Internet Explorer at http://support.microsoft.com/gp/howtoscript
Useful	<ul style="list-style-type: none"> Other reputable websites that tell you how to enable JavaScript in Internet Explorer, such as http://kb.iu.edu/data/ahqx.html
Off-Topic	<ul style="list-style-type: none"> Websites that tell you how to enable JavaScript in browsers other than Internet Explorer, such as http://kb.iu.edu/data/aet.html

	[Louvre visiting hours], English (US)
Query Description	The Louvre is a famous museum in Paris.
User Intent	<ul style="list-style-type: none"> Information - The user wants the museum's visiting hours.
Vital	<ul style="list-style-type: none"> Visiting hours page on the site of the Louvre at http://www.louvre.fr/llv/pratique/horaires.jsp?bmLocale=en
Useful	<ul style="list-style-type: none"> A page from a reputable travel website that provides visiting hours and other useful information http://www.frommers.com/destinations/paris/A25285.html
Relevant	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> A page from a museum guidebook that displays the hours, but in 24-hours time (which U.S. users are less familiar with). Relevant is also acceptable for this page. http://www.europeanmuseumguide.com/museumInfo.php?museumid=115
Off-Topic	<ul style="list-style-type: none"> General travel information about Paris with a brief mention of the Louvre, but no reference to visiting hours, http://www.tripadvisor.com/Tourism-g187147-Paris_Ile_de_France-Vacations.html Wikipedia page on the Louvre, which does not provide visiting hours or even have a link to a page with visiting hours. This page ignores an important element of the query – a request to find the 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 the user.

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	<i>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.</i>
User Intent	<ul style="list-style-type: none"> Action – the wants to sign up for a credit card online
Vital	None possible
Useful	<ul style="list-style-type: none"> Since the user has not specified a particular credit card association or financial institution, the homepage of any of the well-known credit card associations or issuers of credit cards in the United States would be Useful, such as: <ul style="list-style-type: none"> http://www.americanexpress.com/ http://www.usa.visa.com/personal/ http://www.mastercard.com/us/gateway.html http://www.citicards.com/cards/wv/home.do http://www.discovercard.com/ A reputable site that offers credit card comparisons, such as http://www.quicken.com/banking_and_credit/bankrates/?pindex=0&sindex=0&cindex=0&posted=1&psindex=0&SUBRATES=CC&RATES=c1.var&STATE=0&CITY=0
Relevant	<ul style="list-style-type: none"> Pages with information about how credit cards work, such as http://www.howstuffworks.com/credit-card.htm Reputable pages with information and credit card debt advice, such as http://www.ftc.gov/bcp/online/pubs/credit/repair.shtm
Slightly Relevant	<ul style="list-style-type: none"> The credit card application page for a credit card that requires union membership, such as http://www.unionplus.org/benefits/money/card.cfm The credit card application page for a company that issues cards to permanent Australian residents only, http://virginmoney.com.au/credit_card/
Off-Topic	<ul style="list-style-type: none"> University webpage that advises paying tuition bill without a credit card, http://www.emich.edu/finaid/tuition_without_creditcards.html

	[banks], English (US)
Query Description	<i>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.</i>
User Intent	Action – the user wants to open a bank account
Vital	None possible
Useful	<ul style="list-style-type: none"> ▪ Since the user has not specified a particular bank, the homepage of any of the well-known banks in the United States would be Useful, such as: <ul style="list-style-type: none"> ▪ https://www.wellsfargo.com/ ▪ http://www.citibank.com/ ▪ https://www.bankofamerica.com/ ▪ http://www.chase.com/ ▪ Website with links to banks in the United States, organized by state: http://www.thecommunitybanker.com/bank_links/
Relevant	<ul style="list-style-type: none"> ▪ Official government webpage that displays contact information for the U.S. Federal Reserve Banks, http://www.federalreserve.gov/fraddress.htm ▪ The homepage of a small regional bank, which serves communities in that region, http://www.albanybank.com/
Slightly Relevant	<ul style="list-style-type: none"> ▪ The homepage of a bank in another country, such as http://www.barclays.co.uk/ ▪ Outdated article on bank interest rates, http://money.cnn.com/magazines/moneymag/moneymag_archive/2004/12/01/8192192/index.htm
Off-Topic	<ul style="list-style-type: none"> ▪ 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	<i>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, recumbent, etc.</i>
User Intent	<ul style="list-style-type: none"> ▪ Action – the user wants to purchase a bike ▪ Information – the user wants to find information about bikes
Vital	None possible
Useful	<ul style="list-style-type: none"> ▪ Since the user has not specified a particular bike manufacturer, the homepage of any of the well-known bike manufacturers would be Useful, such as: <ul style="list-style-type: none"> ▪ http://www.trekbikes.com/ ▪ http://www.specialized.com/ ▪ https://www.raleighusa.com/ ▪ http://www.schwinnbike.com/ ▪ A website with a wide range of bikes for sale, http://abikestore.com/ ▪ Reputable websites with a comprehensive list of bike reviews or information about different kinds of bikes ▪ The homepage of “Bicycling”, a magazine with information about bikes, including reviews of bikes and bike gear, training tips, maintenance tips, etc., http://www.bicycling.com/
Relevant	<ul style="list-style-type: none"> ▪ Pages with information about how bikes work , such as http://www.howstuffworks.com/bicycle.htm
Slightly Relevant	<ul style="list-style-type: none"> ▪ The “privacy policy” subpage on the Trek website, http://www.trekbikes.com/us/en/general/privacy_policy/ ▪ Homepage of ConferenceBike, manufacturer of a bike that can be ridden by seven riders, http://www.conferencebike.com/
Off-Topic	<ul style="list-style-type: none"> ▪ 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	<i>There are many airline companies that operate in the United States and throughout the world.</i>
User Intent	<ul style="list-style-type: none"> ▪ Action – the user wants to purchase an airline ticket ▪ Information – the user wants to find airlines from which to purchase an airline ticket
Vital	<ul style="list-style-type: none"> ▪ None possible
Useful	<ul style="list-style-type: none"> ▪ Since the user has not specified a particular airline, the homepage of any of the well-known U.S. airline companies would be Useful. Here are some examples (there are many others): <ul style="list-style-type: none"> ▪ http://www.united.com/ ▪ http://www.aa.com/ ▪ http://www.usairways.com/ ▪ https://www.southwest.com/ ▪ Online travel companies that offer flights on numerous airlines: <ul style="list-style-type: none"> ▪ http://www.orbitz.com/ ▪ http://www.expedia.com/ ▪ http://www.travelocity.com/ ▪ The Federal Aviation Administration’s page of links to US airline companies: http://www.fly.faa.gov/FAQ/Airline_Links/airline_links.jsp ▪ Wikipedia page with links to airlines that operate in the United States: http://en.wikipedia.org/wiki/List_of_airlines_of_the_United_States
Relevant	<ul style="list-style-type: none"> ▪ The homepage of a major airline not based in the U.S.: <ul style="list-style-type: none"> ▪ http://www.alitalia.com/ ▪ http://www.jal.co.jp/en/ ▪ Wikipedia page that contains a list of airlines, organized by continent and country: http://en.wikipedia.org/wiki/List_of_airlines
Slightly Relevant	<ul style="list-style-type: none"> ▪ A two-year old article that discusses rumors about mergers between U.S. airline companies.
Off-Topic	<ul style="list-style-type: none"> ▪ The homepage of an company that gives airplane tours of the Grand Canyon, http://www.airgrandcanyon.com/

	[hotels], English (US)
Query Description	<i>There are many hotel companies that operate in the United States and throughout the world.</i>
User Intent	<ul style="list-style-type: none"> ▪ Action – the user wants to make a hotel reservation ▪ Information – the user wants to find information about hotels before making a reservation
Vital	<ul style="list-style-type: none"> ▪ None possible
Useful	<ul style="list-style-type: none"> ▪ Since the user has not specified a particular hotel, the homepage of any of the well-known hotel chains companies would be Useful. Here are some examples (there are many others): <ul style="list-style-type: none"> ▪ http://www.bestwestern.com/ ▪ http://www.radisson.com/ ▪ http://www.hilton.com/ ▪ http://www.marriott.com/ ▪ Online hotel and travel companies that allow users to make reservations with many different hotel chains: <ul style="list-style-type: none"> ▪ http://www.hotels.com/ ▪ http://www.orbitz.com/ ▪ http://www.expedia.com/ ▪ http://www.travelocity.com/
Relevant	<ul style="list-style-type: none"> ▪ Websites that allow users to make reservations with many different bed and breakfast inns: <ul style="list-style-type: none"> ▪ http://www.bedandbreakfast.com/ ▪ http://www.bbonline.com/ ▪ Wikipedia page with general information about hotels: http://en.wikipedia.org/wiki/Hotels
Slightly Relevant	<ul style="list-style-type: none"> ▪ Page about hotel chains in India: http://www.indfy.com/hotel-chains-of-india/
Off-Topic	<ul style="list-style-type: none"> ▪ The homepage of an company that gives airplane tours of the Grand Canyon, http://www.airgrandcanyon.com/

	[London Boutiques], English (US)
Query Description	<i>Boutiques are small specialty shops.</i>
User Intent	<ul style="list-style-type: none"> ▪ Action – the user wants to shop at a boutique in London ▪ Information – the user wants information about boutiques in London
Vital	<ul style="list-style-type: none"> ▪ None possible
Useful	<ul style="list-style-type: none"> ▪ 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. ▪ 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	<ul style="list-style-type: none"> ▪ A blog site with reviews of London boutiques, such as http://samlane100.blogspot.com/2007/11/london-boutiques_19.html ▪ A review of an individual London boutique, with address and contact information, such as http://www.frommers.com/destinations/london/S27883.html
Slightly Relevant	<ul style="list-style-type: none"> ▪ Outdated article about the opening of a London boutique, http://www.fashionunited.co.uk/Content_by_Mail/Received_content/Louis_Vuitton_Opens_New_London_Boutique_200407221926/
Off-Topic	<ul style="list-style-type: none"> ▪ An article about boutiques in Paris, http://www.iht.com/articles/1998/03/13/shop.t.php

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

	[IBM], English (IN)
Query Description	<i>IBM (International Business Machines) is a multinational computer technology company with offices around the world.</i>
User Intent	<ul style="list-style-type: none"> Navigation – the user wants to go the IBM India website.
Appropriate Vital	<ul style="list-style-type: none"> IBM India webpage: http://www.ibm.com/in/
International Vital	<ul style="list-style-type: none"> “Choose your country/region and language” IBM webpage: http://www.ibm.com/planetwide/select/selector.html
Other Vital	<ul style="list-style-type: none"> IBM Australia webpage: http://www.ibm.com/au/en/ IBM Spain webpage: http://www.ibm.com/es/es/ IBM China webpage: http://www.ibm.com/cn/zh/
Useful	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> India IBM contact information page: http://www.ibm.com/contact/in/ Wikipedia article about IBM India: http://en.wikipedia.org/wiki/IBM_India 2008 news article about IBM India: http://www.tradingmarkets.com/.site/news/Stock%20News/1930596/
Slightly Relevant	<ul style="list-style-type: none"> 2004 news article about IBM India: http://www.itworld.com/041007ibmopenindia
Off-Topic	<ul style="list-style-type: none"> Homepage of HP India: http://welcome.hp.com/country/in/en/welcome.html

	[Match], English (UK)
Query Description	<i>There are two equally likely interpretations for this query for U.K. users: Match, the online dating company and Match, the British football magazine</i>
User Intent	<ul style="list-style-type: none"> Navigation – the user wants to go either http://uk.match.com/ or http://www.matchmag.co.uk/
Vital	<ul style="list-style-type: none"> Since neither interpretation is clearly dominant, no Vital is possible.
Useful	<ul style="list-style-type: none"> U.K. Match dating company webpage: http://uk.match.com/ Homepage of Match, the football magazine: http://www.matchmag.co.uk/ Homepage of Match, a research collaboration between five leading UK universities: http://www.match.ac.uk/ . Relevant is also acceptable.
Relevant	<ul style="list-style-type: none"> “Choose your country/language” Match dating company webpage: http://uk.match.com/international/index.aspx U.S. Match dating company homepage: http://www.match.com/ . Slightly Relevant is also acceptable. Wikipedia article about the football magazine: http://en.wikipedia.org/wiki/Match_magazine Wikipedia article about the dating company: http://en.wikipedia.org/wiki/Match.com Wikipedia article about matches that you use to light a fire: http://en.wikipedia.org/wiki/Match “Match of the Day” football webpage on the BBC website: http://news.bbc.co.uk/sport1/hi/football/match_of_the_day/default.stm
Slightly Relevant	<ul style="list-style-type: none"> Careers webpage for the dating company which shows jobs in the U.S.: http://uk.match.com/careers/index.aspx
Off-Topic	<ul style="list-style-type: none"> 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	<i>Sephora is a beauty supply company that sells products online and in stores around the world.</i>
User Intent	<ul style="list-style-type: none"> Navigation – the user wants to go the Sephora website.
Appropriate Vital	<ul style="list-style-type: none"> Canada Sephora webpage: www.sephora.com/canada
International Vital	<ul style="list-style-type: none"> “Choose your country” Sephora webpage: http://www.sephora.com/international.jhtml
Other Vital	<ul style="list-style-type: none"> U.S. Sephora homepage: http://www.sephora.com/ France Sephora homepage: http://www.sephora.fr/ Italy Sephora homepage: http://www.sephora.it/
Useful	<ul style="list-style-type: none"> Canada Sephora Store Locator webpage: http://www.sephora.com/help/stores/allStores.jhtml?country=canada. Relevant is also acceptable.
Relevant	<ul style="list-style-type: none"> Yelp map/review page with information about the Toronto Sephora store: http://www.yelp.ca/biz/sephora-beauty-canada-toronto Amazon.ca page with Sephora beauty guide book for sale: http://www.amazon.ca/Sephora-Ultimate-Makeup-Beauty-Authority/dp/0061466409 Wikipedia article about Sephora: http://en.wikipedia.org/wiki/Sephora
Slightly Relevant	<ul style="list-style-type: none"> Checkout page on Canada Sephora website: https://www.sephora.com/secure/arc20/richCheckout.jhtml;jsessionid=ZXBKWD2KQ0NBICV0KRTQQAQ
Off-Topic	<ul style="list-style-type: none"> Homepage for FabaoCanada, a different Canadian beauty supply company: http://www.fabaocanada.com/

Part 3: Webspam Guidelines

1.0 What is Webspam ?

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

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

1.1 The Relationship between Ratings and Spam

In the “Rating Guidelines”, you learned that landing pages are rated according to their utility to the user 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. It is possible for a page to receive a very high rating – even a **Vital** rating – and also be assigned a **Spam** flag.

1.2 Why do Spammers Create Spam Pages ?

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

- **Pay-Per-Click (PPC)** ads: Spammers get paid each time an ad is 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 money-making 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 Choosing a Browser

As a rater, you are not asked to use a specific browser or to use more than one browser. However, many raters have found it very helpful to use more than one browser and, in particular, to use Mozilla Firefox, especially when looking for spam.

Different browsers sometimes display search results differently. One browser may display a page perfectly, while another browser doesn’t load the page at all or loads it in a way that is difficult to evaluate. Therefore, it is sometimes helpful to open a page in a second browser.

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. Please note that recent versions of Internet Explorer also allow you to add and use tabs.

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	
http://www.h5.dion.ne.jp/~cozmo/	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	
http://www11.asphost4free.com/portale/donne-focose.html	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 Internet Explorer and in Firefox; however, it takes more steps and more time than using the Web Developer toolbar:

If you are using Internet Explorer:	If you are using Firefox:
<ol style="list-style-type: none"> 1. Go to "Tools". 2. Click on "Internet Options". 3. Click the "Security" tab. 4. Click on "Custom level" 5. Scroll down to the "Scripting" section. To disable JavaScript, make sure "Disable" is selected under "Active Scripting". 6. Click "OK". 	<ol style="list-style-type: none"> 1. Go to "Tools". 2. Click on "Options". 3. Click on "Content" or "Web Features". 4. To disable JavaScript, make sure the "Enable" box is not unchecked. 5. Click "OK".

Disabling JavaScript to reveal hidden text	
http://wordlists.waraxe.us/estonian-names.txt	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.

If you are using Internet Explorer:	If you are using Firefox:
<ol style="list-style-type: none"> 1. Go to "View". 2. Click on "Source". <p style="text-align: center;">or</p> <ol style="list-style-type: none"> 1. Go to "Page". 2. Click on "View Source". 	<ol style="list-style-type: none"> 1. Go to "View". 2. Click on "Page Source". <p style="text-align: center;">or</p> <ol style="list-style-type: none"> 1. Right click on the page. 2. Click on "View Page Source".

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

Viewing Source Code to find hidden text	
http://www.regency-uk.com/	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	
http://woefkesranch.com	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	http://pegyxar.20six.fr/pegyxar/rss	Screenshot Example
Fake Blog Example	http://homologauto.com/Foro/login.php/?pill-item=847	Screenshot Example
Computer-Generated Text Example	http://dameomda.isuisse.com/	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	
http://www.agentpedia.com/Coach-house-motor-home.html	Screenshot Examples
http://infopirate.org/bm_used-car-loans	
http://infopirate.org/bm_car-loans-people-bad-credit	

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://supportaj.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	
http://www.twa.com redirects to http://www.aa.com/aa/homePage.do	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 both Internet Explorer and Firefox:

If you are using Internet Explorer:	If you are using Firefox:
<ol style="list-style-type: none">1. Right-click on the page.2. Click "Properties".3. 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.	<ol style="list-style-type: none">1. Right-click on the page.2. Click "This Frame".3. Click "View Frame Info".4. Compare the URL of the frame with the URL of the page. If they are different, the page is probably 100% framed, and should be flagged as Spam.

100% Frame Example	
URL of the page: http://www.neoobe.com/856/animaturk.com.yonlendir.html URL of the frame: http://www.animaturk.com/	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/>), which allow the distribution of their content and may even encourage it, is still considered to be copied content.

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

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

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

4.1.3 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 U.S. 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 money-making 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	<ul style="list-style-type: none">▪ http://groups.google.com/group/katafalak/web/blog-cheap-trackback-url-zyprexa▪ http://groups.google.com/group/katafalak/web/arizona-zyprexa-lawyer▪ http://groups.google.com/group/katafalak/web/zyprexa-side-effects-lawsuit	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://nzeland.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	
<ul style="list-style-type: none">▪ http://www.agipello.info▪ http://www.curriculum-vitae.com▪ http://top-medpills.com/search.php?q=Phentermine▪ http://search.ug/search.php?q=dell▪ http://sketchers.org	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. They often contain keyword stuffing, which will be described in Section 3.1.2. 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	
<ul style="list-style-type: none">▪ http://www.cosmicscripts.com/boards/message/mainboard.html▪ http://www.priyablue.com/msg/	Screenshot Examples

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

Spammed Message Boards: Spammed message boards are different from fake message boards. A spammed message board is a real message board with real posts and real responses, but which has been spammed by posts with PPC ads and/or Porn links. 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 money-making 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 money-making 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 whois in Section 3.1.2.

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.pricegrabber.com/>
<http://www.dealtime.com/>

<http://www.nextag.com/>
<http://www.bizrate.com/>
<http://www.mysimon.com/>

<http://www.kelkoo.co.uk/>
<http://www.ciao.it/>
<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	Thin affiliate (expired domain) URL: http://www.pinecrestcampground.com/	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	
URL: http://letgo.servetown.com/	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 it was maintained by its original owner. 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	
<ul style="list-style-type: none">▪ http://www.mcays.com/▪ http://www.google.com/▪ http://www.knitting.com/	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. Some of this content is written specifically for spammers by writers who 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 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	
<ul style="list-style-type: none">▪ http://super-choice.blogspot.com/2005/06/super-calculator.html▪ http://www.impotence-erectile-dysfunction.com/viagra_drug_the_little_blue_pill.htm	Screenshot Examples

6.0 Spam and the Resolving Stage

It is not uncommon for tasks to go into the “resolving” stage because raters disagree on whether a page should be assigned **Unratable: Didn’t Load** or a rating from the rating scale and a **Spam** flag. The disagreement occurs because raters see different pages when they click on the link in the task. These differences may be due to timing, or due to browser and browser 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. Open the page in a different browser. If you are working in Internet Explorer, try opening the page in Firefox, Safari, Opera, etc., or vice versa.
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).

7.0 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. Please try not to overuse this 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 4: 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:

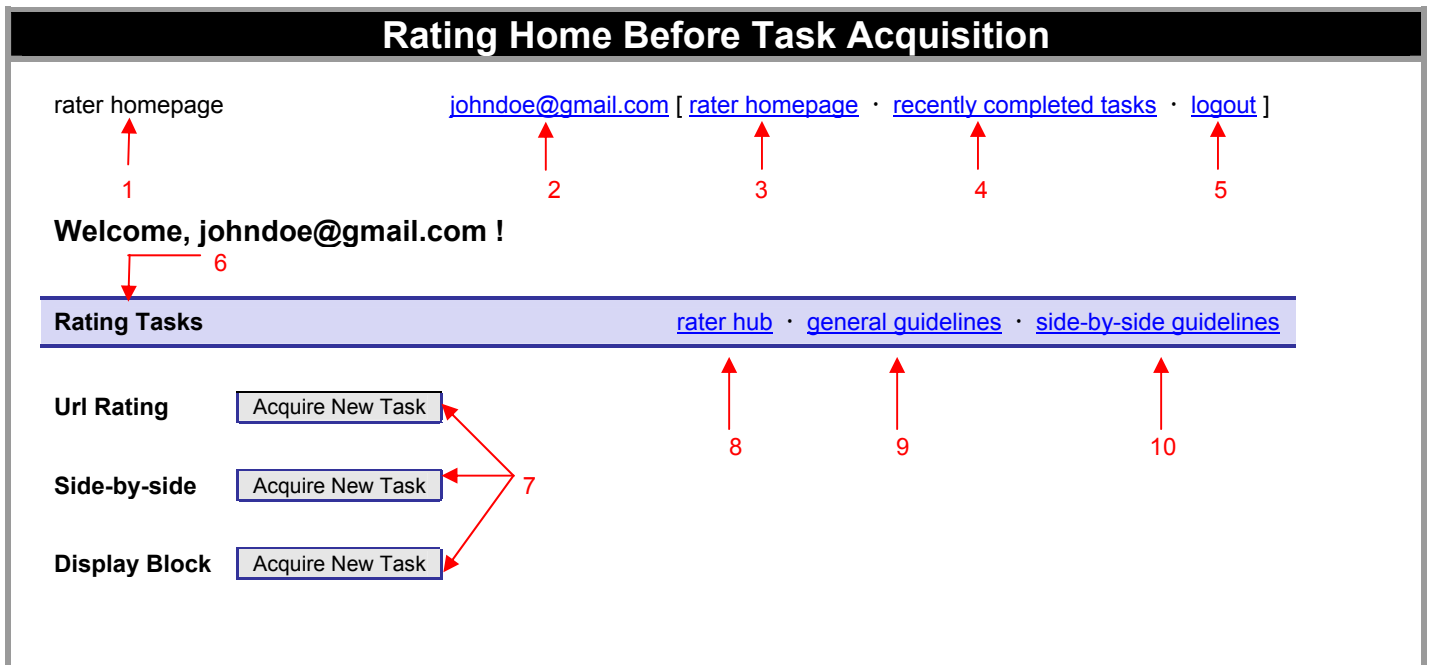
- 1) Rater Hub: Click on the "Start Rating Now" link in the upper left corner of the [Rater Hub](#) homepage. This link will take you to your [Rater Homepage](#).
- 2) Go to this link - <https://www.google.com/evaluation/search/rating/home>

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)



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

rater homepage [johndoe@gmail.com](#) [[rater homepage](#) · [recently completed tasks](#) · [logout](#)]

Welcome, johndoe@gmail.com !

Rating Tasks [rater hub](#) · [general guidelines](#) · [side-by-side guidelines](#)

You have a **URL Rating** task in your queue, please continue .

↑ 11

12 ↓

Resolving Tasks

Resolving tasks in your queue:

Task ID	Status	Language	Query	URL	Last Modified	Expires	Rating
1234567	Unresolved	English (US)	hawaii	http://www.hawaii.gov	2/20/2008	2/20/2008	Off-Topic
7654321	Unresolved	English (US)	sea turtle	http://www.turtle.com	2/21/2008	2/21/2008	Vital

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.

The red numbers represent the following:

1. **rater homepage**
This text shows that you are at the **Rater Homepage**.
2. **rater homepage → rating task**
This shows your location in the EWOQ system; in our screenshot, the display shows the path from the **rater homepage** to the current **Rating Task** page.
3. **johndoe@gmail.com**
Your Gmail account.
4. **rater homepage**
Click on this link to go to the **Rater Homepage**.
5. **recently completed tasks**
Click on this link to change ratings on tasks completed in the last several minutes. Currently, the option to change ratings on recently completed tasks only applies to **Side-by-Side** and **Url Rating** tasks.
6. **logout**
Click on this link to end your EWOQ session. Please logout to end your EWOQ session.
7. **search results**
EWOQ provides you with links to search engines commonly used in your task location. Clicking these links automatically displays search results for the query in the search engine you select.
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 (and possibly reassigned to you) via a randomized process based on availability and priority.

This table displays the differences between the “release task” button and the “Unratable: Meaningless Query” rating:

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.	Yes
Unratable: Meaningless Query	You do not believe any rater will be able to rate 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 four 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.
- You are required to leave a comment if you choose **Unratable: Other**.

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 one 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 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, or when you choose the **Unratable: Other** rating. 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 unresolved task now appears in bold text, please click on it to participate in the resolving process as soon as it is convenient for you to do so.

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](#) · [yandex](#)] · [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

	Rater	Last Modified	Rating	Spam	Flags
1	Rater 2	3/14/08 10:36 AM	Slightly Relevant	Maybe Spam	
	Rater 3	3/12/08 9:02 AM	Off-Topic	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	Spam	Pornography
	Rater 5	3/14/08 6:36 PM	Relevant	Not Spam	

Comments on this Rating

	Comment	Rater	Timestamp
3	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.

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)
AV	(Appropriate Vital)	DL	(Unratable: Didn't Load)
IV	(International Vital)	FL	(Unratable: Foreign Language)
OV	(Other Vital)	MQ	(Unratable: Meaningless Query)
Usf	(Useful)	Mal	(Malicious)
Rel	(Relevant)	PPC	(pay-per-click)
SR	(Slightly Relevant)	LP	(landing page)

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 my browser to continue working
- Page downloaded Trojan on my computer
- My anti-virus software detected a virus

8. Brief comments to confirm your rating in the resolving stage are always appreciated:
 - “Still DL for me.”
 - “Confirming Usf: it’s the best result I could find.”

Part 5: 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.

The 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 the user.

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 or dictionary (such as <http://dictionary.reference.com> or <http://www.wikipedia.org>) and/or do Web research. If the query is meaningless after doing research, you will assign “**Unratable: Meaningless Query**”, a rating that will be discussed later.

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 the user 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, the fruit, or the music company. We call these possible meanings “**query interpretations**”.

Dominant Interpretation: The one query interpretation that is what most users want. For [windows], English (US), the Microsoft operating system is the dominant interpretation.

Common Interpretations: Sometimes, there is no dominant interpretation. For [mercury], English (US), the car, the planet, and the chemical are common interpretations.

Minor Interpretations: Sometimes you will find less common interpretations. For [mercury], English (US), the Mercury Marine Insurance company is a minor interpretation.

Timeliness: A query can be interpreted differently at different points in time. In 1994, the user who typed [President Bush], English (US) was looking for information on President George H.W. Bush. In 2008, 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 fit into more than one type of user intent.

Action Intent: The user wants to accomplish a goal or engage in an activity. These are “**do**” queries: make a purchase, download software, play a game, print a calendar, send flowers, watch a video, copy an image, etc.

Information Intent: The user wants information. These are “**know**” queries: the user wants to know something.

Navigation Intent: The user wants to navigate to a website or webpage. These are “**go**” queries: the user wants to go to or visit a website or webpage.

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. For example, a blank page, a page with images only, or a page with severe encoding errors.

When it is hard to decide, please follow these rules:

- If choosing between task language, acceptable language, or English: choose the flag for the **task language**.
- If choosing between an acceptable language or English: choose the flag for the **acceptable language**.
- Use **Foreign Language** when the page has little or no text in the task language, an acceptable language, or English. If you are having trouble deciding between **Foreign Language** and any other flag, please do **not** use the **Foreign Language** flag.

The Rating Scale

The Rating Scale rating options are: **Vital**, **Useful**, **Relevant**, **Slightly Relevant**, **Off-Topic**, and **Unratable**.

A rating of **Vital** 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 webpage associated with that entity, e.g. [ipod nano], English (US) and <http://www.apple.com/ipodnano/>.

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: Some individuals and other international entities have many versions of their official website, which *all get a Vital rating*. You will use the following categories to describe the pages:
28

Appropriate Vital: Use this for official pages that are the best and most helpful for users in for the task location. The page may be created for users in the task location, or it may be the official page of the entity when no page has been specifically created for the task location. For example, <http://www.stanford.edu> is **Appropriate Vital** for all task locations. This rating can also be used for “regional” **Vital** pages, **Vital** pages that only have a language associated with them and no country, or when the query “asks” for a specific **Vital** page.

International Vital: Use this for either the top level of an international site or a “choose your language or location” page. These pages are designed for users around the world.

Other Vital: Use this for an official page that is specific for a different location or is in a language other than the task language. These pages are not created for users in the task location or for international users; they are created for users someplace else.

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. However, this is not a requirement. Some **Vital** pages do not have high utility.
- Not all entities have official websites or webpages.
- Information queries cannot have **Vital** pages.
- Often, the URL of the official homepage of an entity contains the query terms. For example, the **Vital** page for [ibm], English (US) is <http://www.ibm.com>. However, the **Vital** rating should not be assigned on the basis of the URL alone. Some URLs that “look” **Vital** are not. www.diabetes.com cannot be **Vital** for the query [diabetes], English (US) because this is an information query and no one can claim ownership of 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 pages are very helpful for most users. They often have some or all of these characteristics: comprehensive, highly satisfying, high in quality, authoritative, well-organized, entertaining and/or recent (such as breaking news on a topic). **Useful** pages are “as good as it gets” for queries that do not have **Vital** pages. A good search engine should have **Useful** pages in top positions on pages of search results.

Relevant pages 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.

Slightly Relevant pages are generally not helpful, but are still marginally on-topic. They may be outdated, too narrowly regional, too specific, too broad, etc. to get a higher rating. They may have less information and come from an even less authoritative source.

Off-Topic pages are not helpful for most users. They have no or very little topicality and almost no utility.

- In some cases, there is lack of attention to an important modifier or element of the query.
- The query terms may appear on the page in separate places, but the page topic does not match the query.
- If navigation to helpful content is very difficult, **Off-Topic** is an appropriate rating.

Unratable: Assign to pages that you are unable to evaluate. These are the **Unratable** categories that you will use: **Didn't Load**, **Foreign Language**, **Meaningless Query**, and **Other**. You may assign more than one of these to a task.

Unratable: Didn't Load: Assign this rating to:

- Pages with *generic* 404 errors, client errors, web server errors, network timeout errors, 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.”

Unratable: Foreign Language: Assign this rating when the landing page:

- Is not in any of the following: the task language, an acceptable language, or English.
- Is not clearly **Vital** for the query, based on the appearance of the URL of the landing page.

Unratable: Meaningless Query: Assign this rating *very rarely* - only when the query is meaningless even after query research, and you believe that no other rater will be able to reasonably rate the query either. Do NOT use this category for queries that you personally can't rate because they are too scientific, technical, vague, broad, etc., or if there are multiple interpretations and you can't figure out what the user is looking for. You will use the “release task” button for queries that you personally can't rate, but that other raters may be able to.

Unratable: Other: Assign this to any other pages that you are not able to evaluate and leave a comment explaining why.

Important notes:

- The query terms do not have to appear on the page for a page to have utility. For example, for [Britney Spears], English (US), a photo of Britney Spears has utility even if the query terms do not appear on the page.
- If you are unsure between two ratings, go with the lower rating. If you are unsure among three ratings, use your best judgment.
- Although we don't base ratings only on the URL, it is sometimes helpful to look at the URL when rating in these cases: for spam identification, to notice redirects, and for identification of some **Vital** pages.

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. The queries you rate are not written by you, 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 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. Please think about user intent. For the query [banks], English (US), think about whether it is likely the user is looking for webpages of specific banks or an encyclopedia article about banks.

Queries That Ask for a List: You can think of the results on a search engine result page as a list. Sometimes the best results for "queries that ask for a list" are the best results from that list. For [banks], English (US), homepages of major U.S. banks are **Useful**.

Misspelled Queries: You will notice that some queries are misspelled or mistyped. For **obviously** misspelled queries, you should base your rating on user intent, not necessarily on the query exactly as it is spelled. For [federal expres], English (US), it is reasonable to assume that the user is looking for the homepage of Federal Express. For queries that are **not** obviously misspelled, you should assume the user is looking for results for the query as it is spelled.

URL Queries: These are queries that are URLs or look like parts of URLs (e.g. without "www." or "http://" in the URL). If the URL is a perfectly formatted, working URL, please consider that URL to be the dominant interpretation and give a **Vital** rating when the URL of the page exactly matches the URL in the query. If the query is not perfectly formatted 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.

Search Engine Result Pages: When we rate URL rating tasks, we assume that the user has typed the query in the regular search box of a search engine, and has already experienced seeing a page of web search results. We also assume that the page we are evaluating is a search result that a user sees after clicking a link on the page of search results. Here is how to rate search engine result pages:

- If the page is a set of generic web search results from a major search engine, this is not a helpful result for the user and should get a rating of **Slightly Relevant**.
- If the page is a set of results from a specialty search (such as a map, shopping, book, video, etc. page), it could be very helpful to the user. Ratings will range from **Slightly Relevant** to **Useful**, depending on the utility of the page.
- If the landing page is a search engine page with an empty search box and no results displayed to evaluate, it has no connection to the query; the rating should be **Off-Topic**.

Flags

Spam – You must decide if the page is spam by looking for the spam signals that you learned about in the "Webspam Guidelines", Part 3 of this document. You will assign one of the flags to each URL rating task:

- **Not Spam:** Assign this flag if you do not believe that a page has been designed using deceptive web design techniques.
- **Maybe Spam:** Assign this flag if you find a page to be "spammy", but not spam.
- **Spam:** Assign this flag if you believe that a page has been designed using deceptive techniques.

Pornography – Please 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** 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.

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 quite your 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 6: 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 money-making links. These spam pages 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 they 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.

Choosing a Browser

- Because different browsers display pages differently, it is sometimes helpful to open a page in a second browser.
- Mozilla offers a Firefox Add-on called “Web Developer”, which provides a special toolbar containing tools helpful in spam detection. Two of the most helpful tools on the toolbar allow you to easily disable CSS and JavaScript.

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 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.

If you are using Internet Explorer:

1. Go to “Tools”.
2. Click on “Internet Options”.
3. Click the “Security” tab.
4. Click on “Custom level”.
5. Scroll down to the “Scripting” section. To disable JavaScript, make sure “Disable” is selected under “Active scripting”.
6. Click “OK”.

If you are using 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 the user. 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 within 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. Here are several you can use:
<http://www.domaintools.com/>
<http://whois.mtgsy.net/default.php>.

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.

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 IE, right-click on the page and then click “Properties”. 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 are spam.

The following types of pages have content that is helpful to users. Pages with this type of content are usually not spam.

- 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 which 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 copied text.

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 that are created to send users to a common money-making 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 it 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 you will see keywords contained in the URL separated by hyphens.
- Try to figure out if it seems like 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 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 a spammed message board.

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 money-making 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.

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 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 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 t
- 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 close 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 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 it was maintained by its original owner. If the original site was different, it is probably a parked domain.

Pages with Unhelpful Content and PPC Ads: Some webpages with content are created just for the purpose of putting ads on them. Some of this content is 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 the articles on other webpages. If the content makes sense and appears to be original, please do not assign a **Spam** flag.