# A Study of HTML Title Tag Creation Behavior of Academic Web Sites

**by Alireza Noruzi***

## Abstract

The HTML title tag information should identify and describe exactly what a web page contains. This paper analyzes the *Title element* and raises a significant question: "Why is the title tag important?" Search engines base search results and page rankings on certain criteria. Among the most important criteria is the presence of the search keywords in the title tag. This research concentrates on exploring the retrieval results of Google in retrieving web pages without the title tag. More than one million of academic web pages are found to be untitled, that is, they have not used the title tag.

**Keywords**: HTML; Metadata; Title tag; Search engine; Universities

## Introduction

HyperText Markup Language (HTML) is the *lingua franca* for publishing hypertext on the Web. According to the HTML standard, an HTML page is composed of three parts:

(1) a line containing HTML version information, (2) a declarative header section (delimited by the HEAD element), and (3) a BODY, which contains the page's actual content.[1] The HEAD element contains information, also called metadata, about a web page, such as its title, keywords, description, language, and place of publication that may be useful to search engines, and other metadata that are not considered page content. Search engines do not generally consider elements that appear in the HEAD as page content. However, they may make information in the HEAD available to web users through the search process.

The TITLE element is situated within the <TITLE> and </TITLE> tags, residing inside the HEAD element. For example, the title of the homepage of the University of Tehran is:

```
<HTML>
<HEAD>
<TITLE>University of Tehran</TITLE>
... other HEAD elements ...
</HEAD>
<BODY>
... document body text ...
</BODY>
</HTML>
```

Every HTML page must have a TITLE element in the HEAD section. Web authors should use the TITLE element to identify the contents of a page. Since users often consult pages out of context, authors should provide context-rich titles. Thus, instead of a title such as "Introduction", which doesn't provide much contextual background, authors should supply a title such as "Introduction to Medieval Bee-Keeping".[2]

The information contained in a title tag appears as a header or label at the top of the screen in the reverse bar of a browser window and in the tab bar for multi-tab browser displays to present obviously the title of the web page that is being viewed; in most browsers, such as Internet Explorer, it is usually white text with a blue background.

* Alireza Noruzi is a faculty member of the Dep. of Library and Information Science, University of Tehran, Tehran, Iran <nouruzi (at) gmail.com>

Noruzi, A. (2007). A Study of HTML Title Tag Creation Behavior of Academic Web Sites. *Journal of Academic Librarianship*, 33 (4): 501-506.

When a web user surfs on the Web and saves (or "bookmarks") the address of a web page/site that s/he comes across, it is the title that appears in the menu of *Favorites* or *Bookmarks*. Thus, the title of a web page/site should be informative as to its contents. It is frustrating to web users to change the title of the bookmark if it is not informative. Unfortunately, not all webmasters put a meaningful title on their web pages, and some do not put any title at all.

Search engines, such as Google and Yahoo, "may monitor data maintained or generated by a user, such as *bookmarks*, *favorites*, or other types of data that may provide some indication of documents favored by, or of interest to, the user. [The purpose is] to analyze over time a number of bookmarks/favorites to which a document is associated to determine the importance of the document".[3] The assumption is that the majority of bookmarks generally seem to point toward authoritative sources. The title tag is also used as the default file name when saving the page to disk in browsers such as Internet Explorer. When printing a web page, the text in the title tag appears in the upper left corner of the sheet. So it should be informative and succinct, as well as descriptive to identify the contents of the page.

Search engines generally consider two criteria in indexing and ranking web pages, namely *on-page* (i.e. HTML codes, title tag and the content of the page) and *off-page* (i.e. back-links) criteria. Obviously, the title tag is one of the most important factors in achieving high ranking on search engine results. For reasons of accessibility and searchablity, search engines make the content of the TITLE element available to users and they truncate long titles in search results, because of the limitations in using the title tag characters. Thus, important keywords should be used early in the title to help search engines and web users identify the main subject of the page, and also to avoid truncating the words appear at the end by search engines that use short titles.

Search engines also use the text included in the title tag as a clickable link to the page/site in response to a search query in the search results. Moreover, search engines may penalize web sites that repeat a keyword in the title tag of web pages (considered spam). The information in the title tag is searchable by web users because search engines, such as Yahoo, AltaVista and AlltheWeb, support the *TITLE command*. For example,

title:Metadata

title:How to Use Meta Tags

All search engines consider the title tags in web pages. In fact, the title tag is the most powerful positioning element for them in indexing and ranking web pages. When a robot or spider retrieves a page and passes it to its indexing program, it will start with the words in the title and begin looking for those words in the text on the web page. Robots consider the title of the page to be the most telling description of the content of a page. In response to a search query, a search engine looks at the title first, and if it finds a keyword in the title, the search engine will rank the page above other pages which only have the keyword in the main body of text. Generally web users never open a web page unless the title or subject makes it absolutely clear what the page is about.

The title tag should have an accurate and specific entry for every item. This is important for two reasons: first, because search engine algorithms give it significant weight, and second, because users see the contents of the title tag highlighted in their search results. The title tag is the only effective metadata element in HTML, and the primary means of making metadata functional via the Web.[4] Users have to browse numerous web pages in their search results to identify relevant items. It therefore follows that the title tag is vital for web pages.

If the title tag is not present in a page, the page is shown as "Untitled" on search engines' results (see Table 1). Unlike '*heading*' which usually appears in large bold text further down a web page, '*title*' is not typically rendered in the text of a page itself.

The scientific movements generated by universities, such as the increased setting up of digital repositories, the institutional self-archiving of peer-reviewed papers, theses and dissertations, the movement to publish open access (OA) journals by universities, etc., are compelling reasons why the title tag is important to academic libraries and universities.

The objectives of this study are:

1. to demonstrate the importance of the TITLE tag;
2. to compare the amount of untitled web pages on the Google search engine;
3. to compare untitled web pages of universities around the world; and
4. to quantify the amount of English meaningless expressions used in page titles on the Web.

## Literature review

The results of a study by Tunender and Ervin[5] showed that by better utilizing the HTML title tag, web designers can increase the chance of retrieval of their sites. Perhaps the best retrieval can be achieved by planting multiple terms within the HTML title tag for their web sites. Keywords inserted between title tags are also considered a positive factor for ranking. Christopher[6] argued that Excite, Infoseek and AOL Search display only 70 characters from the title tag. Some search engines show more characters, others show fewer characters, but 70 characters is a good limit to target for page titles.

"The page's HTML title tag is most important. Failure to put target keywords in the title tag is the main reason why perfectly relevant web pages may be poorly ranked […]. The titles should be relatively short and attractive, like newspaper headlines".[7] "One of the main rules in the ranking algorithm involves the location and frequency of keywords on a web page … Pages with the search terms appearing in the HTML title tag are often assumed to be more relevant than others to the topic".[8]

The results of a study by Craven[9] clearly validate the decision to assign the highest weight to the title tag as extraction cues for web page description construction, although it is noted that the extent to which the title matches description wording is quite variable.

Zhao[10] assesses the affect of keyword positions in the title of the homepage content (keywords inserted between the first-level *heading* i.e. <H1> and </H1>), in the HTML title, and in the uniform resource locator (URL) on ranking. The relation between the page ranking and the keywords found in the title of the homepage content is studied. This study examines the content between the title tags in the HTML of each homepage. The examination reveals that the web designers do not always put the same words in the content title (first-level heading <H1>) and in the HTML title tag. She concludes that (1) the presence of keywords in the HTML title seems to have little influence on ranking in Google retrievals; and (2) the keyword in the URL affects the ranking. The less the keyword appears in the URL, the lower the ranking. This examination of keywords in the content title, HTML title, and URL shows that none of the three has a strong effect on overall rank of a web site by Google, although a decline in URL keyword strength does coincide somewhat with overall rank.

Spencer[11] argues that the tried-and-true optimization tactics, such as keyword-rich title tags, appear to work quite well. The text within a page title is given more weight by the search engines than any other text on the page; keywords at the beginning of the title tag are given the most weight.

Raisinghani[12] and Dahm[13] suggest that web site designers should take a little time to fine-tune title tags to increase traffic to their sites. Improving the title tag is one technique that applies to almost all search engines. The appearance of keywords within the page title is one of the major factors determining a web site's score in many engines. Changing page titles to

include some of the site's keywords can greatly increase the chance that a page will appear with a strong ranking in a query for those keywords.

Noruzi[14] examines a sample comprising twenty-five countries' top-level domains (TLDs) from five continents (i.e. five countries from each continent). It is shown that more than twenty-six million web pages from these countries are untitled. It is also shown that more than thirty million web pages from generic top-level domains (i.e. com, org, net, edu, gov) are untitled.

Baeza-Yates, Castillo and López[15] analyzed the title of pages of Spain and found that over 9 percent of the pages have no title, and 3 percent have a default title such as the Spanish equivalents of "*Untitled document*" or "*New document 1*". It is unlikely that a title that is shared by many pages is a good description of the contents of a page. The titles of pages are repeated several times across pages, mainly because authors use just the name of the site as the title of many unrelated pages. They argue that many search engines give more importance to keywords appearing in the title of the page than to the same keywords appearing in the main text of the page, so without a title or without good keywords in the title there are less possibilities of appearing in the top results of a search engine […] as titles are omitted or repeated across web pages, less than 10 percent of the pages in the Web of Spain have titles that can be used by search engines.

## Materials and Methods

The approach used in this study includes the following steps:

First, we conducted a search on two search engines (Google and MSN) to compare the amount of untitled web pages on the entire Web. Other searches were conducted on Google to find web pages created in HTML, PHP, ASP, PDF, etc., not having a title (see Tables 1 and 2).

Second, we considered another sample comprising twenty-five countries' second-level domains (SLDs) for universities from five continents (i.e. five countries from each continent) in order to compare the amount of untitled web pages. Academic top-level domains were chosen as the sample base for the dataset because academic domains include a large part of the entire Web in most countries of the world. Note that in this case the sample consists of countries which have used a second-level domain, such as 'edu' or 'ac' for academic web sites (see Table 3).

Third, we tried to compile a list of English meaningless expressions used in title tags throughout the Web in order to show the frequency of these expressions on the Web. Note that the title of a homepage should contain at least the name of the organization, company, association, or institute, and should describe in a few words its activities, and should avoid using generic titles like "Our Home Page", "Welcome to Our Web Site", etc. There is no point in using a generic title (see Table 4).

## Results

When a user searches on the Web, the *Untitled* mention often appears in the search results of all search engines. We conducted a search on both Google and MSN search engines to find *Untitled* pages (see Table 1). A considerable number of web pages do not have a title. However, there are a significant number of pages using the *Untitled* word in their titles. This is because some web authoring softwares use the *Untitled* word as a default setting and it is also because many web pages use this word in their titles. In the later case, the pages are not untitled. For example, "*Untitled Oliver Stone*"

**Table 1. Untitled pages on Google and MSN** (December 3, 2006)

| Search engines | Search Command | No. of Untitled Pages |
|---|---|---|
| Google | allintitle: "Untitled Document" | 28,700,000 |
| MSN | intitle:"Untitled Document" | 35,664,681 |

Table 1 shows that a significant number of web pages on the Web are untitled. However, this is partly due to their file types. Other searches were conducted to find web pages created in HTML (i.e. their file types are HTML, HTM or XHTML), as a search engine-friendly file type. Table 2 shows the number of pages (e.g., HTML, PHP, ASP, PDF, etc.) not having a title. The following search was determined: the number of *Untitled* pages created in HTML, for example:

allintitle:Untitled filetype:HTML

**Table 2. Untitled pages on Google** (December 3, 2006)

| File type | No. of Untitled Pages |
|---|---|
| PDF | 13,200,000 |
| HTM | 10,100,000 |
| HTML | 5,260,000 |
| ASP | 845,000 |
| PHP | 836,000 |
| SHTML | 352,000 |
| EPS | 208, 000 |
| ASPX | 176,000 |
| CFM | 136,000 |
| PS | 126,000 |
| PHP3 | 36,600 |
| XML | 14,300 |
| DOC | 13,800 |
| TXT | 13,400 |
| SHTM | 10,400 |
| PHTML | 10,300 |
| SWF | 1,080 |
| PPT | 759 |
| XLS | 618 |
| RTF | 315 |
| XHTML | 209 |
| **Total** | **31,340,781** |

Table 2 shows that a significant number of HTML and HTM pages do not have a title; although Google may not always show the accurate number of untitled pages. The main objective of this research is to demonstrate the importance of the TITLE tag.

Another objective of this study is to examine how universities use the title tag in their web pages. We searched for the number of *Untitled* academic pages appear in our sample of twenty-five countries, using a combination command for each country on Google (see Table 3). The following searches were determined:

- the number of *Untitled* academic pages, for example:

allintitle:Untitled site:ac.jp/

- the total number of academic pages from each country, for example:

site:ac.jp/

**Table 3. Untitled academic pages found on Google**

| Country | Academic Domain | No. of Untitled Academic Pages | No. of Academic Pages in each country | Percentage of Untitled Academic Pages |
|---|---|---|---|---|
| India | ac.in/ | 475,000 | 641,000 | 74.10 |
| China | edu.cn/ | 352,000 | 44,800,000 | 0.79 |
| Japan | ac.jp/ | 221,000 | 28,500,000 | 0.78 |
| UK | ac.uk/ | 169,000 | 37,300,000 | 0.45 |
| Australia | edu.au/ | 55,800 | 16,700,000 | 0.33 |
| Colombia | edu.co/ | 29,500 | 3,840,000 | 0.77 |
| Austria | ac.at/ | 22,900 | 6,960,000 | 0.33 |
| Poland | edu.pl/ | 11,600 | 4,930,000 | 0.24 |
| Argentina | edu.ar/ | 10,300 | 2,130,000 | 0.48 |
| Belgium | ac.be/ | 6,960 | 5,220,000 | 0.13 |
| Iran | ac.ir/ | 5,830 | 701,000 | 0.83 |
| South Africa | ac.za/ | 5,670 | 1,810,000 | 0.31 |
| Yugoslavia | ac.yu/ | 5,350 | 462,000 | 1.16 |
| Peru | edu.pe/ | 4,770 | 773,000 | 0.62 |
| Brazil | edu.br/ | 4,340 | 674,000 | 0.64 |
| New Zealand | ac.nz/ | 2,800 | 1,330,000 | 0.21 |
| Saudi Arabia | edu.sa/ | 1,260 | 287,000 | 0.44 |
| Egypt | edu.eg/ | 848 | 140,000 | 0.61 |
| Morocco | ac.ma/ | 780 | 63,700 | 1.22 |
| Panama | ac.pa/ | 261 | 17,800 | 1.47 |
| Fiji | ac.fj/ | 232 | 49,400 | 0.47 |
| Papua New Guinea | ac.pg/ | 137 | 10,900 | 1.26 |
| Guam | edu.gu/ | 105 | 3,120 | 3.37 |
| Nigeria | edu.ng/ | 69 | 27,800 | 0.25 |
| Algeria | edu.dz/ | 33 | 40,700 | 0.08 |
| **Total** | | **1,386,545** | **157,411,420** | **0.88** |

We found that more than one million of academic web pages from these countries do not use the title tag (see Table 3) and many academic pages have incorrect or misleading titles. This means that the use of title tags needs to be considerably reviewed and improved by academic webmasters. We also conducted several searches on Google to choose a sample of meaningless expressions used as the title tag on English-language web pages throughout the Web (see Table 4).

**Table 4. Meaningless expressions found on Google** (December 3, 2006)

| Search command | No. of Hits |
|---|---|
| allintitle: "Welcome to my website" | **209,000** |
| allintitle: "Welcome to my homepage" | 181,000 |
| allintitle: "Welcome to our website" | 81,900 |
| allintitle: "Welcome to our company" | 81,800 |
| allintitle: "Welcome to my web site" | 52,400 |
| allintitle: "Welcome to our site" | 39,200 |
| allintitle: "Welcome to my home" | 34,400 |
| allintitle: "Welcome to my site" | 31,600 |

| | |
|---|---|
| allintitle: "Welcome to my web page" | 27,300 |
| allintitle: "Welcome to my home page" | 21,500 |
| allintitle: "Welcome to my page" | 18,700 |
| allintitle: "Welcome to our web site" | 14,700 |
| allintitle: "Welcome to our home" | 12,800 |
| allintitle: "Welcome to our homepage" | 12,100 |
| allintitle: "Welcome to our home page" | 740 |
| allintitle: "Welcome to my webpage" | 590 |
| allintitle: "Welcome to our page" | 225 |
| allintitle: "Welcome to our webpage" | 190 |
| allintitle: "Welcome to our web page" | 119 |
| allintitle: "Our family homepage" | 75 |
| allintitle: "Our company homepage" | 20 |
| **Total** | **820,359** |

All these title expressions are far from being descriptive and meaningful enough. Moreover, page titles are important when the author has tried to convey a single, impressive meaning wrapped up in a catchy phrase to arouse the reader's thinking processes. Thus, a title that starts with "Welcome to ..." is a wasted opportunity.

We also evaluated the first ten long titles retrieved in response to the query "*Citation Indexing*", whose end-words are truncated by Google. We found that Google shows fifty-four characters (no space) on average, and up to fifty-nine characters per long title (i.e. 9.7 words on average). However, long titles are retrievable in response to a title command. We conducted a search with the following command (allintitle: "Citation Indexing" Humanities) and found that the first retrieved long title, which is truncated by Google, is "*Citation Indexing - Its Theory and Application in Science* **...**". It does not have the keyword "Humanities", although it has the keyword in its original title. It can be concluded that truncating a title has no impact on its retrieval. In other words, keywords which are truncated from the title are searchable and retrievable on Google, although the user cannot see the keywords in a truncated title. We have conducted several searches to be sure that this is true. As a result, a good title is up to ten words long (not exceed sixty characters), with important keywords placed near or at the beginning of the title.

## Discussion

The title is first information that a web user reads when a search engine presents him/her one of the pages of a site in its results. The title should summarize the contents of the web page efficiently, because users generally visit the first ten results. The results of studies on search behavior show that from 1996 to 1999, for more than seventy percent of the time, a user only views the top ten results. Over fifty percent of users do not access results beyond the first page. Jansen, Spink and Saracevic[16] found that more than three in four users do not go beyond viewing two pages. By 2001, only roughly one-third of users look beyond the second page of web sites retrieved.[17] By 2003, in general users view about five web pages per query.[18,19]

In addition, keywords occurring in the title tag of web pages are more important than those in the body of the text, both as an indexing factor (i.e. used to index web pages) and as a ranking criterion. Search engines often use the title of a page to find the pages most relevant to a search query. The title of a page that includes keywords is assumed by search engines to be more relevant than those without such words in the title. The more keywords in the title, the more likely the page is to be found. Moreover, singular and plural forms of keywords may be important for users. Thus, each page should have a unique title, containing specific

keyword phrases and relevant to its content, although the title of homepages should be general.

Search engines can substitute the string (text) that appeared in the first-level *heading* (i.e. <H1>…</H1>) as the title of untitled pages because web pages usually have a *heading* in the visible portion of the text. <H1> is the HTML element for the top-level heading of a web page. Consequently, to help search engines, the title tag should match the first *heading*.

Many web sites use the same title tag throughout the site, repeating a single title on all pages. This is a wasted opportunity for attracting users and traffic to the site. A good title can therefore increase the traffic to the site, because the page title is the only way that many users will be able to identify and find a page. So the title serves to capture the interest of the user.

## Suggestions for improving title tags

To sum up, the results suggest using:

- a meaningful title that summarizes the content of each web page;
- a unique title that contains the important keywords related to the content of each web page, so that main keywords appear at the beginning of the title tag;
- an attractive and concise title that includes the most important keywords, keeping them descriptive;
- a short title that includes ten words maximum for each web page;
- an informative title not including the repeated keywords. The organization or company name should be used as the homepage title tag, and should not be repeated in all pages. The organization's name can be added to the end of the title of pages if necessary, like a book title with main title and subtitle.

## Conclusion

In summary, titles of web pages are important to web crawlers, as well as to web users. Title tags give webmasters an opportunity to make sure their pages will come up correctly on search engines. Unfortunately, there are no standard rules about the content of title tags on the Web. The title of web pages may be changed by webmasters at any time, while traditional library materials, such as books, may have a subtitle or alternative title as well as a main title, but these never change after publication.[20] A uniform title for each web page is necessary as an access point. However, the procedure for determining the uniform title for a web page is not always simple and the title of web pages may not always reflect the true content.

The results of this study suggest the need for further research, including assessing the influence of other factors, such as keywords in the URL, the domain name, the language of the site, the web site link popularity and keyword density, on the ranking of web sites. Future studies should focus on improving web pages' title tags, finding new systems for mining titles of file types (such as PDF, PS, EPS, DOC, etc) with which search engines have problems. We also suggest estimating the meaningless expressions used as the title tag of web pages in other languages apart from English.

## Acknowledgements

The author wishes to thank Mrs. Marjorie Sweetko for her helpful comments. The constructive comments of two anonymous reviewers are also acknowledged.

## References


[1]. W3 Consortium, “The global structure of an HTML document,” (N.D.). Available: http://www.w3.org/TR/REC-html40/struct/global.html (accessed December 2, 2006).

[2]. Ibid.

[3]. Anurag Acharya, Matt Cutts, Jeffrey Dean, Paul Haahr, Monika Henzinger; Urs Hoelzle, Steve Lawrence, Karl Pfleger, Olcan Sercinoglu & Simon Tong, "*Information retrieval based on historical data*," United States Patent Application 2005/0071741, Kind Code A 1.

[4]. Alan Dawson, "Creating metadata that work for digital libraries and Google," *Library Review* 53 (September 2004): 347-350.

[5]. Heather Tunender & Jane Ervin, "How to succeed in promoting your web site: The impact of search engine registration on retrieval of a World Wide Web site," *Information Technology and Libraries* 17 (September 1998): 173-179.

[6]. G. Christopher, "Effective web site marketing starts in the design phase," *Infotech Update* 9 (October 2000): 8-12.

[7]. Danny Sullivan, "Search engine placement tip," *Search Engine Watch* (2002, October 14). Available: http://searchenginewatch.com/webmasters/article.php/2168021 (accessed December 2, 2006).

[8]. Danny Sullivan, "How search engines rank web pages," *Search Engine Watch* (2003, July 31). Available: http://searchenginewatch.com/webmasters/article.php/2167961 (accessed December 2, 2006).

[9]. Timothy C. Craven, "HTML tags as extraction cues for web page description construction," *Informing Science Journal* 6 (2003): 1-12. Available: http://inform.nu/Articles/Vol6/v6p001-012.pdf (accessed December 2, 2006).

[10]. Lisa Zhao, "Jump higher: analyzing web site rank in Google," *Information Technology and Libraries* 23 (2004): 108-118.

[11]. Stephan Spencer, "Gunning for Google," *Catalog Age* 22 (February 2005): 15-16.

[12]. Mahesh S. Raisinghani, "Future trends in search engines," *Journal of Electronic Commerce in Organizations* 3 (2005): I-VII.

[13]. Tom Dahm, "Search engine strategies and optimization tips," *The Web Developer's Journal*, (May 2000). Available: http://www.webdevelopersjournal.com/articles/search_strategies_tips.html (accessed December 2, 2006).

[14]. Alireza Noruzi, "Editorial: The HTML title tag and its importance," *Webology* 2 (December 2005), Editorial 6. Available: http://www.webology.ir/2005/v2n4/editorial6.html (accessed December 2, 2006).

[15]. Ricardo Baeza-Yates, Carlos Castillo & Vicente López, "Characteristics of the Web of Spain," *Cybermetrics* 9 (2005), Paper 3. Available: http://www.cindoc.csic.es/cybermetrics/articles/v9i1p3.html (accessed December 2, 2006).

[16]. Bernard J. Jansen, Amanda Spink & Tefko Saracevic, "Real life, real users and real needs: A study and analysis of users' queries on the Web," *Information Processing and Management* 36 (March 2000): 207-227.

[17]. Amanda Spink, Bernard J. Jansen, Dietmar Wolfram & Tefko Saracevic, "From e-sex to e-commerce: web search changes," *IEEE Computer* 35 (March 2002): 133-135.

[18]. Bernard J. Jansen & Amanda Spink, "An analysis of web information seeking and use: Documents retrieved versus documents viewed," in *Proceedings of the 4th International Conference on Internet Computing*, Las Vegas, Nevada, (June 23-26, 2003): 65-69.

[19]. Amanda Spink & Bernard J. Jansen, "A study of web search trends," *Webology* 1 (December 2004), Article 4. Available: http://www.webology.ir/2004/v1n2/a4.html (accessed December 2, 2006).

[20]. Alan Dawson, "Creating metadata that work for digital libraries and Google,"



**Bibliographic information of this paper**:

Noruzi, A. (2007). A Study of HTML Title Tag Creation Behavior of Academic Web Sites. *Journal of Academic Librarianship*, 33 (4): 501-506.