

Metadata and Metatags

There are many cases in which you need to include metadata (data describing other data) on a web page. For example, you want to record the authors and the summary of a web page, and the most natural place to store the information is the page itself. However, it is often inappropriate to display this information on the web browser for viewers to see (e.g., viewers are not interested in who is the author of a product information page but you want to record it for administrative purpose).

Metadata is specified with the `<meta>` HTML tag within the header of a web page. Therefore, this information will not be displayed on the browser but you can examine the HTML source of the page to find out the metatags defined in the web page.

There are several common metatags used in web pages. The most common ones are `title`, `keywords`, and `description` tags. The less common but useful tags are `author` and `date` tags. The following is an example showing the use of metatags. Notice that most web pages use the `<title></title>` tag in place of the `title` metatag.

```
<head>
<meta name="title"
      content="Best Chinese Search Engine">
<meta name="author"
      content="Suntek Computer Systems Ltd.">
<meta name="description"
      content="The best Chinese search engine in the world was
launched ...">
</head>
```

You may reference the [standard element set](#) defined in the [Dublin Core Metadata Initiative](#) (commonly known as the Dublin Core), which is an open forum for the development of metadata standards for a broad range of applications.

The Dublin Core recommends the use of the `description` metatag to contain the table of contents or an abstract of the web page. However, most web pages do not follow this

recommendation. Perhaps it is because the metatags are not displayed by the browser anyway, there does not seem to be any point to put in an abstract which is not going to be displayed. However, this point is invalid since metatags are there to help applications, not human users, to manipulate the pages in a meaningful way. For example, you have written a number of articles and you want to display a list of the articles and a summary for each article on the list. Since there is no automatic way till now to produce a good summary for an article, you have to type it in by yourself and put it in the `description` metatag. Then, you can easily write a CGI script to extract the summary from each article and display them in the list. It certainly requires some typing, but the convenience to the users resulting from the summaries far outweighs the costs.

Metadata for Search Engines

Search engines rely on an indexer to extract useful keywords from web pages and record them in their index databases. Search queries are conducted on the keywords and statistical means are used to judge the relevance (or usefulness) of a web page against the query. The relevance judgement is based on keywords contained in the web pages, where and how often they appear on the web pages and across the entire collection, and, of course, the number of links pointing to the pages.

No matter how great search engines are, there is no magical way (yet!) for search engines to determine the content (or semantics) of web pages and use them for judging the relevance of the web pages. Likewise, there is no easy way to produce precise summaries for web pages automatically. Many search engines just extract the first few lines from the beginning of a web page as the summary. Some search engines such as Google simply report the lines containing the search keywords. Neither of them are satisfactory methods for summarizing the content.

It is too easy to put the blame on search engines and proclaim how stupid they are. Granted that search engines make mistakes, but, as content developers, you can help the search engines, and yourselves by the way, to write your web pages in such a way that the search engines can easily capture the important content of your web pages. In fact, the whole industry, call search engine optimization, is built to help users to write their web pages in the right way and include metatags that would make their web pages rank high on search engines.

You can find out more about the importance of metatags in articles from [search engine positioning companies](#), [search engine registration companies](#), and [companies building affiliate programs](#).

Suntek Indexes and Searches Metatags

Suntek's search engine indexes not only the body of web pages but also *any* metatags that appear in the HTML header section. With a configuration file, you can tell the search engine specifically which metatags to index (common tags are title, keywords and description). For a life example employing Suntek's search engine, take a look at the [Chinese interface](#) or [English interface](#) of the Hong Kong Government's search engine, one of the largest search portals in Hong Kong. You may notice that there is a selection menu called *Search Scope*, with which you can specify a search to be applied to the *full text* or one of the metatags of the web pages.

Another example can be found at the [City University's CityUToday](#) publication. Try a few queries and you will notice that the results contain the titles, dates, summaries of each matched article. Examining the HTML source of a matched article reveals that metatags are used to define these attributes and that all metatags are indexed and searchable.

You may observe from this life example that summaries are very useful in helping the user to determine if he or she wants to go into the page and browse it. It also makes the result page look a lot nicer. Without the `description` tag, most search engine, including Suntek, would simply extract the words from the beginning of a page to serve as the summary. While better than nothing, the first few words don't always represent the summary of an article. Suntek's search engine allows the search engine administrator to specify that a metatag (in this case the description metatag) be used to produce the summary of a web page. For the [HTML segment](#) above, the following will be displayed in the result:

Best Chinese Search Engine

The best Chinese search engine in the world was launched ... Of course, Suntek allows you to include in the result page other useful information, such as last modification date, URL of the page, size of the page, the category that the page belongs to, etc. With a template-driven approach, you can format the result page in any way you want. You can try out the various web sites built on Suntek search engine and note the diversified designs employed.

The "date" Metatag

While Suntek allows you to index and search any metatag created in your web pages, the "date" metatag receives a special treatment. This is because while other metatag values are simply treated as text strings and indexed accordingly, the "date" value has to follow a specific format, which is defined by W3C as <DATE>T<TIME>+<TIMEZONE>. The following is a more specific example:

```
<meta name="date" content="CCYY-MM-DDThh:mm:ss+mm:ss">
```

For example, the following date corresponds to the Hong Kong/Taiwan time 01:02:30 of June 1, 2000.

```
<meta name="date" content="2000-06-01T01:02:30+08:00">
```

Why do we need to input a date metatag? If you want to search or display the date of the web page (whether it is the creation date or the last edit date), we recommend you to use the date metatag to specify it. This is because not all web servers return the last modification date of a page (not to mention that the last modification date is not necessarily the same as the date when the document was created or modified).

Other References

1. [An article on metatags, including the HTTP-EQUIV tags](#)