

Educational Information in the Web: Discussing the Metadata Requirements for a Web Service Guiding Citizens' Education

Soile Hirvasniemi

Kai Öörni

Department of Information Studies

University of Oulu

soile.hirvasniemi@oulu.fi, kai.oorni@oulu.fi

Abstract

The paper discusses the metadata needs of a national web-based Opintoluotsi ("Study adviser", "Study pilot") service, which is being developed by the Finnish Ministry of Education. This service aims to provide information about education and training for citizens. The service is discussed here from the viewpoint of metadata and its role in information seeking especially in the World Wide Web. The results of a study made in the first phase of the Opintoluotsi project about educational institutions' web sites and their use of metadata show that metadata is used rarely. From this point of view, it is important to inform the educational institutions about the importance of metadata for information retrieval and visibility in the Internet.

Keywords: Educational information, Metadata, Dublin Core, Information seeking, Public WWW services.

1 Introduction

Currently, the use of metadata seems to be very rare in Finnish web-based public services, though there do exist instructions issued by the Finnish Public Administration for improving the accessibility of web pages. By accessibility, these instructions mean ease of finding and using the web-based service. In addition to other factors affecting accessibility, the guidelines also discuss metadata and the use of keywords. The instructions mentioned above refer to the Dublin Core metadata standards as an accepted scheme for applying metadata [1]. In spite of the official promotion of content descriptions, the actual content and quality of metadata on web pages are not yet sufficient.

Information seeking in the web using free text searches involves the typical problems of large electronic information seeking environments. It is difficult to distinguish the important information from the unimportant, even with the use of precise keywords, because of the diverse contexts of the words and phrases being used, and because of the

lack of fields that describe the document's content. The use of structured metadata can eliminate at least some of the problems with the help of pre-defined sets of fields and labels, which define the amount, type and form of meta-information. Separation of the content description explicitly from the text body aids in targeting the search terms more precisely. However, it should be borne in mind that an ability to make full use of metadata also requires metadata-oriented search engines. This paper discusses the role of metadata in educational information retrieval and introduces a Finnish educational information web service *Opintoluotsi*.

2 Background of the *Opintoluotsi* web service

The *Opintoluotsi* ("Study adviser", "Study pilot") service aims to provide various forms of support to help people find educational and training opportunities and to encourage them to use the different services available. The project is estimated to last from the year 2000 to 2006 and is funded by the European Social Fund.

The citizens' web service will be publicly available in April 2002 (<http://www.opintoluotsi.fi>). The service will include a broad variety of educational information and search facilities. Additionally, the concept will have two support services, one public service for educational institutions and education providers to help them to enhance their presence in the web and another, more limited service for counselling and guiding professionals to promote professional development and cooperation.

The service's search facility will rely strongly on hypertext documents published by the education providers. The *Opintoluotsi* service will use its own search engine and indexing robot, which can be tuned to utilize metadata more efficiently than the general WWW search engines do now (e.g. Google and Altavista). During the development of the service, a database of information provided by the Finnish educational institutions has been collected, including

information about each institution's homepage in the Internet. This means that the search can be targeted specifically to the institutions' and schools' homepages, which will naturally reduce the number of irrelevant documents in the retrieved set. As mentioned earlier, the future versions of the search engine will probably make use of metadata by stressing the impact of metadata fields' content when the results are ranked. For this reason, it was important to find out how educational institutions nowadays use metadata fields on their web sites. The study concerning metadata use on educational web sites carried out by the *Opintoluotsi* project team was performed during the spring 2001, and it covered 198 Finnish educational institutes of 19 different types¹.

3 Metadata is used rarely on Finnish educational institutions' web sites

The study evaluated the main web sites of 198 different educational institutions in order to find out how well they presented information about their institution and courses on their web sites. In this evaluation, metadata fields were also examined and collected for further inspection. The examination of metadata was limited to the "Keywords" and "Description" metadata fields.

It turned out that only 41 educational institutions (21%) out of the total of 198 used metadata. This percentage includes all the educational institutions that had any metadata on their web site, including those with a single word in their metadata "Description" field. All sites used the "Keywords/Description" pair instead of a metadata scheme.

It was also considered useful to examine the quality of the metadata that had been used by the educational institutions. Table 1 presents the information found in the metadata fields. It should be noted that the information describing the fields of study was, in most cases, given using uncontrolled vocabulary. Only one institution used the Finnish National Board of Education's official descriptions concerning the fields of study available in the institution. Fourteen institutions listed course names, study subjects or attainable degrees in the metadata. It was not studied how accurate this information was in relation to the school's actual supply of courses. Description about the institution in table 1 means a short text describing the institution and its purpose. Eight institutions provided this information. Only one

educational institution indicated the level of education – though in most cases the level can also be seen from the name of the institution.

It seems that there is a need for guidance about how to use keywords and descriptions. It can be questioned if keywords have much practical use in information retrieval if words with very broad or specific meaning are used, especially in such a form that they are not likely to be used in conjunction with other search terms. By this we mean such words and phrases in the keyword field as "additional services", "departments", "world of physical exercise", "target-oriented", "apply education", "beginning education" and "studying". In two cases, the metadata was cluttered with improper use of HTML or the character set. The description field was confused with the keywords field in one case.

Table 1. Contents of the metadata fields examined.

Type of information found	% of pages containing metadata	% of total set of pages	N
Name of the institution	58.5 %	12.1 %	24
Type of the institution (refer to footnote 1)	56 %	11.7 %	23
Fields of study	41.5 %	8.5 %	17
Course names, study subjects or degrees	34.1 %	7.1 %	14
Description about the institution	19.5 %	4.0 %	8

Many sites used the same words as metadata that they also used in the page's content. In case the search engine's support for metadata is limited, there is little use for that kind of metadata, except that it directs the user of the search engine to the page containing those keywords.

4 Dublin Core metadata and the *Opintoluotsi* web service

Dublin Core is one of the several metadata schemes. Another metadata scheme that could be useful in the case of educational information is Learning Object Metadata (LOM), which has been developed by the IEEE Learning Technology Standards Committee (<http://ltsc.ieee.org/wg12/>). LOM builds on the work done by now already completed ARIADNE project (<http://ariadne.unil.ch/Metadata/>). For this reason, both LOM and ARIADNE metadata have structurally a lot in common. An example of a practical development project, which developed its own set of metadata elements, is the European CUBER project

¹ The types of educational institutions whose web pages were examined: Adult educational centres, advisory organizations' learning centres, continuing education centres, evening institutes, folk high schools, music schools, open polytechnics, open universities, physical education centres, polytechnics, senior high schools, senior secondary schools for adults, specialized schools, special schools, study centres, universities, university summer courses, vocational education institutions, police and military academies.

(<http://www.cuber.net>). The aim of CUBER is to build an information system for the information technology courses of European universities. CUBER uses LOM as a base of its own scheme. Lamminaho presents a review of LOM and other metadata schemes [3].

Dublin Core Metadata has several advantages compared to the other metadata schemes, which influenced the decision that the *Opintoluotsi* service will make use of the Dublin Core Metadata Standard.

The most important reason for using Dublin Core Metadata is that it has several national formats, including Finnish [2]. The availability of a Dublin Core platform in the native language makes its use easier and causes fewer mistakes in coding. The Finnish Dublin Core platform site is hosted by University of Helsinki (<http://www.lib.helsinki.fi/cgi-bin/dc.pl>). It can also be said that Dublin Core is easy to use even for people with no previous cataloguing experience [2]. This has also been proven by a questionnaire study performed by The Nordic Metadata Project in 1998. The Dublin Core elements were very well understood. Although there were only 21 people who completed the questionnaire on the Dublin Core user guide web site, it could be said that Dublin Core seemed to be simple enough even for novice users to understand and use [2]. Another advantage is the small number of elements in the Dublin Core basic element set (15 elements) compared to LOM (47 elements) and ARIADNE metadata structure (38 elements). The use of the Dublin Core set for educational resources (DCEd) will, however, increase the number of elements. Perhaps, in the future, there will not be marked differences between the DCEd, LOM and ARIADNE metadata schemes in this sense.

After finding out that the use of metadata on web sites of educational institutions was rare and the quality of such data was quite poor, it was clear that the *Opintoluotsi* service could not rely exclusively on metadata when designing the search facilities. To achieve better-quality search results, the *Opintoluotsi* service will encourage educational institutions to use metadata. The *Opintoluotsi* support service for educational institutions informs institutional web designers about how to use metadata on educational sites. The service contains various kinds of advice about the use of metadata and web design generally and also links to the Dublin Core sites. Educational institutions are instructed how to use descriptive keywords in their metadata in order to achieve better visibility in the *Opintoluotsi* service, and the Internet. The *Opintoluotsi* service also offers individual help through e-mail concerning different problems in creating well-defined metadata.

5 Design questions related to metadata policies for educational information

An interview study made by the authors about the search behaviour of individuals seeking educational information showed that information about the fields of study, the levels and prerequisites of the study programs and the possible institutions (and their locations) offering education were considered relevant in an information seeking situation [4]. Occupations related to each study program were also important according to the interviewees. It should be investigated if these findings could support the instructions that guide the use of metadata. The question is what kind of metadata could aid the searcher or could be used otherwise by the information system and would simultaneously also be information on whose availability the search facility does not vitally depend. *Opintoluotsi* will use hypertext documents created by the educational institutions' web authors, which will make it more difficult to predict if a document contains information about its content, in contrast to controlled systems, such as the aforementioned CUBER course database.

Another question is how the hypertext document's structure affects the use of metadata: what kind of directions should be given on how to structure the hypertext document so that it will support use of content descriptions? A document (i.e. a hypertext node in the web environment), which discusses only a single topic is naturally the easiest to describe precisely, while a content description of a compound document containing many topics may be long and complex. However, creating many fragments of text might not be a practical way to create the hypertext documents for the institution distributing information on its web site, and a "golden mean" between these alternatives should be aimed at.

We do not know yet how willing educational institutions will be to use metadata. It is crucial that the process of entering metadata should be seen as a useful and non-mechanical task. Lamminaho points out an interesting notion about the user friendliness of the metadata scheme [3]. To put it simply, she states that the metadata scheme should be concise enough and should not have too many obligatory elements, to avoid excess work when entering metadata. Methods of user-centered design from the field of human-computer interaction (HCI), such as user testing, contextual inquiry² and questionnaires, can be applied during the development of the scheme to achieve a more usable and understandable set of elements and, first and foremost, a scheme that will be actually used by the people who create web content.

² Contextual inquiry is a discovery process using interviewing method, which is focused on how a product is used in its context.

6 Conclusion

Well-defined metadata can make a difference in information retrieval in the web environment. As an organizer and classifier of information, it can help people to find the information they need. In spite of that, metadata is quite rarely used, which may be due to the fact that Internet search engines do not make much use of metadata. Also, the knowledge concerning metadata is not so widely spread as it could or should be.

For the Finnish *Opintoluotsi* web service, metadata is also very important. It can help educational institutions to achieve better visibility and help people to find the educational information they need. For that reason, the *Opintoluotsi* service must work as an encouraging partner and help educational institutions to learn how to use metadata efficiently. It might even be an advantage that metadata is not widely used on Finnish educational websites. The *Opintoluotsi* service can be a pilot even in this sense, developing common set of rules for use of Finnish educational metadata.

Note: More information about the Opintoluotsi project coordinated by the Finnish Ministry of Education can be found in the web at <http://www.opintoluotsi.fi/english>.

References

- [1] Web page of Finnish Public Administration, which includes guidelines concerning the use of metadata on Finnish public service web pages (text in Finnish): <http://www.intermin.fi/juhta/suosituksset/jhs129.htm>
- [2] The Nordic metadata project: Final report. Nordinfo, Helsinki, 1998.
- [3] Virva Lamminaho. Metadata Specification: Forms, Menus for Description of Courses and all other Objects. CUBER Deliverable D3.1. University of Helsinki, 2000.
<http://www.cuber.net/web/publications/cuber-d3-1.pdf>
- [4] Soile Hirvasniemi and Kai Öörni. Information behaviour related to educational information. An unpublished report. University of Oulu, 2001.