

Emily Brown

LIBR 247

Assignment #1

10/6/12

This report will examine the differences and similarities between the index terms I found using ERIC's thesaurus and the abstract I wrote on an article, to ERIC's index terms and abstract of the article, and OIR's keywords and abstract of the same article. The experience of actually choosing the indexing terms for an article was very enjoyable and interesting. While I had written an abstract before this time I will be comparing it to two other abstracts. I hope to dissect the process and explain the choices I made and what I feel the most successful strategy is for both indexing and abstracting.

Comparing the indexing terms that I chose to ERIC's I noticed that we had a number of them in common. Five nearly perfect matches. The few differences I noticed were I chose Internet instead of World Wide Web. I selected Internet because I find the term World Wide Web to be severely outdated. I used Subject Access instead of Subject Index Terms. I see a correlation between the two, but I feel that Subject Access is more accurate since it was used repeatedly in the article itself. I used Authority Control instead of Authority Control (Information), which is a more narrow term that I missed. I used Library Research instead of Research Methodology, which I must admit is a much better term than mine.

Data Processing and Online Searching are the only two other terms that didn't show up at all in ERIC's list. I felt that those were two minor points that were worth mentioning. The Identifiers are also very different, I used Semantic Web and Dublin Core since they were both mentioned quite a bit in the article, and ERIC used XML and Interoperability. I recall searching specifically for

Interoperability, not finding it and electing to remove the term all together. I hadn't even considered adding it as an identifier. XML was mentioned very sparingly in the article so I'm not sure if I agree with it's use. Terms I now wish I had used now that I see both lists would be Indexing and Cataloging. While both of those terms were not mentioned by name specifically in the article, the concepts were mentioned quite a bit. One of those between the lines readings that I didn't see at the time.

Overall, I found the guidelines to be easy enough to work with. I wrote down a number of terms to look up in ERIC's thesaurus and as I was going through them I was a bit surprised to see that some of them weren't included. For the most part, I simply moved on to a similar word or skipped it all together. Most interestingly I ran out of terms by the time I got to the four minor identifiers so I tried the browse by category option which I found somewhat useful. Of course, the selections I made that was were terms not at all similar to ERIC's list of terms. I must concede that ERIC had much better identifier terms overall than I did.

I would also like to compare the keyword terms that OIR assigned to the article with my identifier terms. Classification is the only term that we had in common. The terms that OIR used seemed far more general than I would have thought appropriate due to the target audience of library information professionals. That particular nugget of information is in the second paragraph of the introduction. The terms Web Sites and Design Optimization seem far too broad for the specific ideas elaborated on in the article.

Comparing my abstract to ERIC's I saw a few similar terms and turns of phrase. However, I feel that the differences outweigh the similarities. ERIC's abstract seemed to pull a number of things directly from the article itself, or more specifically the headings. It then combined the terms together in a coherent and insightful manner using roughly half the number of words that I did in mine. I chose my phrasing to explain points about the article that I felt were important. While I did use some terms

and phrasing from the article itself, much of what is in the abstract is in my own voice. I also mentioned the article's target audience, which ERIC did not. In fact, in rereading it, I notice that I repeat the same idea in the first two sentences. I've noticed that I found the Dublin Core discussion to be far more important than ERIC did, both in indexing terms and the abstract.

Similarities between ERIC's abstract and my own include some phrases. Subject access on the web from ERIC, versus subject access on the internet. Interoperability and authority control are two subjects mentioned. That is where the similarities end. ERIC's abstract starts discussing the topics mentioned in the article throughout the abstract. Mine stops a few in and then discusses the results of each. I believe this difference in strategy of how the information is presented is quite telling. Despite having a few things in common, ERIC's abstract is very different, and superior, compared to mine.

Examining both ERIC and my abstracts with OIR, I think that OIR is far more similar to my abstract. OIR's abstract mentions Dublin Core, and elaborates on ideas in the article in a nearly narrative manner. This isn't an entirely successful technique. I feel that OIR's abstract misses some of the points implied in the article. It also over explained some concepts. The information that OIR chose to convey in it's abstract is different in the turns of phrases used, as well as details. For example, OIR's abstract mentions Library of Congress Subject Headings in a non-abbreviated way. This seems to completely ignore the target audience for this article, the library information professionals. If nothing else, you would expect it to be abbreviated. It was such a minor part of the article that I didn't mention it at all in mine, nor did ERIC's. Also worth mentioning is the alternative spelling of "Cataloguing" which OIR uses and is I believe an English or Canadian variant of the word. This alternate spelling is used in both the abstract and the keyword, which may create problems in information retrieval for the document in the future. In regard to abstract length, OIR's abstract is 174 words, compared to mine at 91 and ERIC's at 48. OIR seemed to convey less about the article's contents in more words.

After reading all three abstracts, I must say that I prefer ERIC's for the most part. It's clear, concise and elegant. I felt very humbled, frankly, after comparing my abstract with ERIC's. I was abashed to see that the unsuccessful OIR's abstract resembled mine so much. I do stubbornly think that Dublin Core should be mentioned. The second best is my abstract. I feel that ERIC's abstract conveys the aboutness of the article much more clearly than either of the other two abstracts. It also does it in far less words, which is especially useful. Changes that I would make to my own abstract after reading ERIC's would probably include shortening the abstract. There is a bit of superfluous wording that isn't necessary. I would also try to stick to the topics mentioned instead of going into results.

In conclusion, I would like to say that I can see why ERIC exists. It does its job of indexing and abstracting very well. Using the thesaurus experience was a pleasant one and I was pleased with my results. The abstracting process was also interesting, more so when comparing the three abstracts. The actual selection of index terms was at once easier and quite a bit more difficult than I would have imagined. For me, picking the terms I wanted to look up in ERIC's thesaurus was very simple. The difficult part was checking with the thesaurus and deciding what I wanted to do with the terms that weren't included. Obviously I could have put them all in the identifier section, but I felt that would be a poor choice. At least one of those, perhaps interoperability, would have been a good idea to include. Overall, this was a very interesting assignment.

### Bibliography

Franklin, R. (2003). Re-Inventing subject access for the semantic web. *Online Information Review*, v27 n2, pg94-101.

### My Identifiers and Descriptors:

ID: Dublin Core, Semantic Web

Descriptor Major (6): Authority Control, Internet, Metadata, Subject Access, Thesauri, Classification

Descriptor Minor (4): Data Processing, Information Retrieval, Online Searching, Library Research

### My Informative Abstract:

Article discusses need for subject access on the Internet, with library information professionals as the audience. Current state of the Web is discussed with specific needs to create subject access. Authority control, interoperability of classification schemes, metadata schemes, and thesauri are required. Dublin Core is cited as one possible scheme. Long term growth and scalability are main issues to overcome in regards to thesaurus design. Faceted classification schemes are a likely answer to the problem of creating subject access to documents on the Web that are too complex to normally classify.

### ERIC Record:

#### Click on any of the links below to perform a new search

ERIC #: [EJ675099](#)

Title: [Re-Inventing Subject Access for the Semantic Web.](#)

Authors: [Franklin, Rosemary Aud](#)

Descriptors: [Authority Control \(Information\)](#); [Cataloging](#); [Classification](#); [Indexing](#); [Information Retrieval](#); [Metadata](#); [Research Methodology](#); [Subject Index Terms](#); [Thesauri](#); [World Wide Web](#)

Source: [Online Information Review, v27 n2 p94-101 2003](#)

Peer Reviewed: Yes

Publisher: N/A

Publication Date: 2003-00-00

Pages: N/A

Pub Types: [Journal Articles](#); [Reports - Descriptive](#)

**Abstract:** Examines the developing state of subject access on the Web. Topics include new scholarly research methods; authority control; cataloging and metadata; interoperability and thesauri development; extensibility, including XML and RDF (resource description framework); indexing and information retrieval; flexibility in subject classification; and current Web subject classification projects. (LRW)

**Abstractor:** N/A

**Reference  
Count:** 0

**Note:** N/A

**Identifiers:** Interoperability; XML

**Record  
Type:** Journal

**Level:** N/A

**Institutions:** N/A

**Sponsors:** N/A

**ISBN:** N/A

**ISSN:** ISSN-1468-4527

**Audiences:** N/A

**Languages:** English

**Education  
Level:** N/A

**Direct Link:**

OIR's

### **Abstract**

First generation scholarly research on the Web lacked a firm system of authority control. Second generation Web research is beginning to model subject access with library science principles of bibliographic control and cataloguing. Harnessing the Web and organising the intellectual content with standards and controlled vocabulary provides precise search and retrieval capability, increasing relevance and efficient use of technology. Dublin Core metadata standards permit a full evaluation and cataloguing of Web resources appropriate to highly specific research needs and discovery. Current research points to a type of structure based on a system of faceted classification. This system allows the semantic and syntactic relationships to be defined. Controlled vocabulary, such as the Library of

Congress Subject Headings, can be assigned, not in a hierarchical structure, but rather as descriptive facets of relating concepts. Web design features such as this are adding value to discovery and filtering out data that lack authority. The system design allows for scalability and extensibility, two technical features that are integral to future development of the digital library and resource discovery.

Article Type: General review

Keyword(s): Design optimization; Classification; Digital libraries; Web sites; Cataloguing.

Journal: Online Information Review

Volume: 27

Number: 2

Year: 2003

Pp: 94-101

Copyright © MCB UP Ltd

ISSN: 1468-4527