LIS 415
Assignment 3
Paul Hoffman <hoffman@simmons.edu>

Part B

In this paper, I compare the types of metadata used in three subject-based Web resource catalogs. I determined the metadata fields available in the three catalogs by examining several resource records in each. The most noteworthy differences lie in the nature and presentation of classification and subject analysis of resources, so I take particular note of these areas.

EEVL

The ominously named EEVL (http://www.eevl.ac.uk/catalogue.htm) bills itself as 'the Internet guide to engineering, mathematics, and computing.'

Of the three resource guides, only EEVL provides an alternative or added title field. Alternative titles are common in EEVL; for example, the record for *Association for Computing Machinery Special Interest Group on Software Engineering* includes the alternate title *ACM SIGSOFT*.

The field labeled 'Keywords' includes postcoordinate subject descriptors; another labeled 'Resources' contains a mix of descriptors denoting resource format and other non-topical subject terms.

Each record also contains a field holding one or more classification strings. The label for this field indicates which of EEVL's three disciplines the resources falls under; for example, a resource relating to software testing would have a field labeled 'Computing Classification' whose contents might be 'Software – Software Engineering – Testing and Debugging.' It's conceivable that a single record might fall under more than one of the three disciplines, but I couldn't find any such records.

These classification strings do double duty: they function as precoordinate topical headings while also indicating the path(s) by which one may arrive at the record when browsing EEVL. The paths generally follow a topic-subtopic-subsubtopic (and so on) pattern in which the first element is taken from the list of 33 main subjects. In some cases, however, the strings seem poorly chosen ('General – General Mathematics Resources') or misformulated (e.g., 'Mechanical Engineering and Related Industries – Control Engineering – Electrical, Electronic and Computer Engineering – Control Engineering').

Each element in a classification string appears as a hyperlink. Clicking on the last element takes the user to a list of all records with the same classification string; clicking on any other element leads only to a search form in which that element has been preselected in a pull-down menu used to specify the scope of the search.

Humbul

The goal of Humbul Humanities Hub (http://www.humbul.ac.uk/) is to be 'UK higher and further education's first choice for accessing online humanities resources.'

The Humbul home page features a prominent list of 18 topical subjects and a link to a full list containing these and an additional 15 subjects. The significance of this division is not explained.

Each resource in Humbul is categorized using a two-level, browsable hierarchical classification; for example, 'American Studies / Projects and Organizations.' The first level consists of one of the 33 subjects, and the choice for the second level is taken from one of three axes. The first axis is labeled 'Type' and includes a wide variety of descriptors, for example 'Projects and Organisations,' 'Research Related,' or 'Primary Sources'. The second axis is labeled 'Period' and consists of 12 time periods ranging from a single century to a span of 1,500 years. The third axis denotes the intended audience of the resource: 'Postgraduate/Faculty,'

'Undergraduate,' or 'General Public.' A resource may fall under more than top-level subject; the field 'Humbul Subject' lists each of these in the form of a clickable hyperlink.

When browsing, each record's location within this hierarchy is indicated at the top of the record in the form of a breadcrumb string. For example, the record for *Abraham Lincoln papers at the Library of Congress* contains a header '<u>Home</u> / <u>Museums, Libraries, Archives</u> / <u>Primary source</u> / Abraham Lincoln papers at the Library of Congress << You are here.' The underlined elements are hyperlinks; <u>Museums, Libraries, Archives</u> leads to a list of second-level nodes under that subject, while <u>Primary Source</u> leads to a list of records for primary sources in that subject.

This breadcrumb, however, is present only when the user has arrived at a record by browsing; it does not appear in records retrieved via a search. What's more, even if the record falls under more than one of the 33 top-level subjects, only one breadcrumb string is given – the path by which the user arrived at the record. Thus, Humbul fails to take advantage of the full potential of its own classificational system.

Humbul also features a set of metadata fields that link to other resources; the name of each of these fields indicates the nature of the relationship between the two resources. For example, the record for *Abraham Lincoln papers at the Library of Congress* includes a field labeled 'Is Part Of' containing the title *American memory: historical collections for the national digital library* and a clickable hyperlink to the latter's web site. However, Humbul again fails to fully exploit this potentially useful feature: it doesn't supply a hyperlink to the *record* for the related resource.

Two other notable metadata fields denote the geographical coverage of each resource and its latitude and longitude. (The anticipated use of the latter field is not indicated.)

INFOMINE

INFOMINE (http://infomine.ucr.edu/) provides access to 'scholarly Internet resource collections.' Its home page lists nine categories; most of these are subjects (e.g., 'Business and economics') but some denote a format ('Ejournals') or a combination of subject and format ('Maps and GIS'). Due to a lack of space – a large image of a many-faceted gem takes up much of the available space – several category names are abbreviated; for example, 'PhysSci, Engr, CS & Math' is used for 'Physical sciences, engineering, computer science and mathematics.'

Besides the field in which a resource's category is listed, INFOMINE records include a field for LC subject headings and another for LC classification. A field labeled 'Resource Types' contains such values as 'Abstracts and Indexes,' 'Companies,' and 'Electronic Journals.' Other fields indicate the nature of access (free or not free), the resource's audience level ('Academic' was the only value I was able to find), some basic statistics on clickthroughs to the resource, and an indication of whether the resource's URL is still 'live' (and when it was last checked).

The contents of all these fields are hyperlinks; clicking on a link leads one to a list of all the records containing the same value in the same field. The user can also browse records by the contents of most of these fields, or by author or title.

Comparison

The differences among the three resource catalogs are much fewer, but far more interesting, than the similarities. Each catalog provides basic metadata fields such as title, author, URL, and description, and each divides its resources into broad subject areas that appear as searchable fields in the resource records and function as points of access for browsing. Each catalog further provides additional subject access using postcoordinated descriptors, variously referred to as 'keywords' or 'subjects.' INFOMINE goes one step further by including precoordinated LC subject headings and LC classification..

EEVL models its collection of resources as a network of interrelated nodes; this takes the form of a graph, in which each node may have multiple incoming, and multiple outgoing, links. Each of INFOMINE's broad subject divisions is a flat space, but alternative browsable spaces are

provided for resource type and postcoordinate subject descriptors ('keywords'). Humbul, on the other hand, provides an intermediate solution, a two-level tree structure – or, rather, 33 two-level trees, one for each of its 33 subjects.

EEVL and Humbul display 'you are here' or 'breadcrumb' information indicating each record's location within the catalog's classificational system; Humbul uses this information only as a navigational header, while EEVL explicitly lists it in a classification field within the record.

Two of the catalogs include a field for the language of the resource, and two include 'metametadata' indicating when the metadata record was created and when it was last modified.

The fields I found are summarized in Table 1 (non-subject fields) and Table 2 (subject fields). All three catalogs provide an indication of the resource type, though this is sometimes conflated with subject information. Table 3 shows some of these type descriptors and tentatively divides them into five broad categories.

Field	**EEVL**	**Humbul**	**INFOMINE**
Title	+	+	+
Alternate title(s)	+	–	–
URL	+	+	+
URL check (resource availability and date last checked)	–	–	+
Author/creator	+	+	+
Publisher	+	–	+
Description	+	+	+
Resource type/format	+	+	+
Language	+	+	–
Country of origin	+	–	–
Intended audience	–	+	+
Regions covered	–	+	–
Periods covered	–	+	–
Access (free or not)	–	–	+
Record creation and modification metadata	–	+	+
Statistics	–	–	+

Table 1. Summary of non-subject metadata fields.
A plus sign (+) indicates that a field is available; a minus sign (–) indicates that it is not. Some less significant fields are omitted.

Feature	EEVL	Humbul	INFOMINE
Broad postcoordinate topical descriptors	–	+	+
Keywords (narrow topical descriptors)	+	–	+
Precoordinate subject headings	–	–	+ (LCSH)
Location of the resource within a hierarchical arrangement	+ (messy)	+	–
Coded classifications	–	–	+ (LCC)
Number of top-level topics or categories	3	33	9

Table 2. Summary of subject- and classification-related features.

Resource type category	Sample descriptors
Broad class of content type	Database Full-text documents Journal – Abstract/Contents
Broad class of intellectual content	Conference reports/Papers Electronic Journals Electronic Texts and Books Mailing List Bibliographic source [i.e., bibliography]
Intended use	Primary source Research-related Teaching-related
Entity type	Project Website Society/Institution Organization Place
Format	Image Audio Software [i.e., executable programs]

Table 3. Categories of resource type descriptors.