

Terminology services

Making knowledge organization schemes more accessible to people and computers

BY DIANE VIZINE-GOETZ, Consulting Research Scientist, OCLC Research

Which of the following most closely defines the term 'vog'?

- a. the latest in Japanese street fashion
- b. not expressing one's thoughts clearly
- c. volcanic smog
- d. a young fjord horse

According to the 2004 version of the Library of Congress Subject Headings (LCSH), the correct answer is c. Vog is volcanic smog. Concepts like this are constantly being added to knowledge organization schemes, such as thesauri, subject heading systems and classification schemes.

The goal of OCLC's terminology services project is to make the concepts in knowledge organization schemes and the relationships within and between schemes more accessible to people and computer applications. For example, if a hypothetical Web service provided access to the equivalent and related terms for concepts in LC Subject Heading records, it would be possible for software developers to create tools to improve Web searching. To test this hypothesis, go to your favorite search engine and search for the word 'vog.' Then modify your search to include the words 'vog volcanic smog volcanic gases.' The latter search, which includes variant and related terms from LCSH, will likely produce higher quality search results for materials about volcanic smog.

Before a Web service can be developed for a given knowledge organization scheme, it's often necessary to preprocess the concept data. For some schemes, it's necessary to convert the data from word processing documents or html pages to structured data formats, such as the MARC 21 formats for authority or classification data, or the SKOS core, an RDF schema for thesauri and related knowledge organization schemes. Once a scheme is in a structured format, it can be enhanced in several ways. Typical enhancements include mappings to other schemes, the addition of persistent identifiers, and the addition of coding to track the origin of records and the sources of changes. The end products of these processes are XML files that can be used as the basis for terminology Web services.

Terminology services are Web services that involve various types of knowledge organization resources, including authority files, subject heading systems, thesauri, Web taxonomies and classification schemes. OCLC researchers have prototyped several experimental terminology services. One Web service that uses the Dewey Decimal Classification (DDC) provides access to the DDC summaries. The service returns captions, in four languages, for DDC numbers at the top three levels of the classification. For example, when DDC class number 798 is submitted to this service, the service returns the following information:

```
<skos:Concept rdf:ID="S22.798">
<skos:inScheme rdf:resource="#S22"/>
<skos:prefLabel>798</skos:prefLabel>
<skos:altLabel xml:lang="de">Reitsport,
Tierrennen</skos:altLabel>
<skos:altLabel xml:lang="en">Equestrian sports
& amp; animal racing</skos:altLabel>
<skos:altLabel xml:lang="es">Deportes ecuestres y
carreras de animales</skos:altLabel>
<skos:altLabel xml:lang="fr">Sports équestres et
courses d'animaux</skos:altLabel>
</skos:Concept>
```

Although this response might not satisfy most human users, it will be quite acceptable to machines. If the results are intended for human eyes, it is up to computer applications to format them appropriately.

Another prototype terminology service, with a human interface component, uses the Microsoft Office 2003 Research services pane to search a database of genre terms for fiction. Without ever leaving the Office application, a user can issue a search and paste results from the search directly into a document. For example, if a college student wishes to categorize a reading list of fiction titles based on genre, he could copy the titles into a Microsoft Excel 2003 workbook, open the Research services pane, send a search to the OCLC Research GSAFD vocabulary service, and then place the results into his document.

For more information on Terminology Services, see: www.oclc.org/research/projects/termservices/.