Untangle: A New Ontology for Card Catalog Systems

Christopher Welty and Jessica Jenkins

Vassar College Computer Science Dept.
Poughkeepsie, NY 12604-0462
weltyc@cs.vassar.edu
http://untangle.cs.vassar.edu/

Abstract

The ontology used by most card catalog and bibliographic systems is based on a now outdated assumption that users of the systems would be looking for *books on shelves,* and therefore only books were first-class objects, with people, organizations, etc. as simple attributes. This limited the ability of a user to browse. A new ontology for card catalog systems is proposed that suggests that persons, organizations, conferences, etc., should be first-class objects with attributes and relations of their own, creating a rich space of background information that helps users find what they are looking for. This new ontology has been implemented in a knowledge-based system called Untangle, which demonstrates two key advantages of this rich information space: it enables automatic augmentation of the data through reasoning, and it enables a new paradigm for search that combines querying and browsing.

Introduction

Library card-catalog systems are quite old, with evidence of recorded indexes dating back at least as far as the 12th century. Through the centuries the underlying assumption of these systems – implicit or explicit – was simple: the only object of any search is a book. This assumption led to an *ontology* that still exists today: the only type of object in a card-catalog system is a book, which has attributes such as author, title, publisher, date, etc. While we understand that these attributes correspond to people, organizations, etc., in implemented systems they are nothing more than strings.

As the size of, and access to, libraries increase dramatically with modern communication, this ontology no longer serves, and a new ontology, that supports a new paradigm in library search, is called for. The new ontology supports the notion that persons, places, events, organizations, etc., are first-class objects, and that a user *browsing* through the information may submit individual queries whose results may be these objects, *not books*. For the web, the consequence of this is the same, the result of an individual query does not always have to be a web page. A search may be for information about a person, place, organization, etc.

This implies that search must be, in this age of information, an iterative process that alternates between *browsing* the information space, and *querying* it. The results of queries will be new starting places for browsing, and informa-
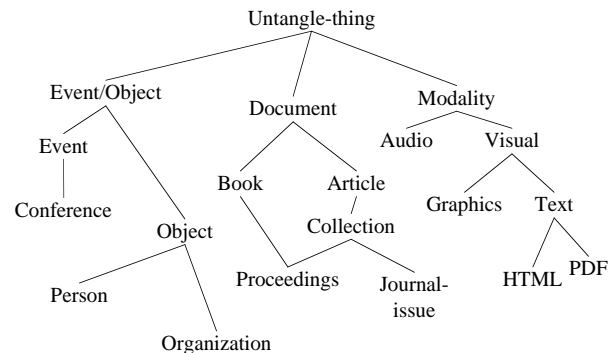
Figure 1: Taxonomy of Untangle concepts.

tion discovered during browsing may suggest new queries, since clearly users may not have perfect knowledge of what they are looking for.

We present here the ontology of the Untangle system, and then describe the system itself, which is available on the web.

The Ontology

The Untangle Ontology contains two types of objects: things related to topics (TOPICAL-THING) and things related to entities in the real world (UNTANGLE-THING). A more complete description of topic related issues can be found in (Welty and Jenkins, 1999), and will not be discussed here.

Basic Types

The domain of real entities for Untangle is split into three main types: Modalities, Documents, and Event/Objects (Welty, 1996). Some of the taxonomy is shown in Figure 1, and in the demo you can select "list the main objects in the ontology" to get a textual view of the ontology taxonomy.

Documents are the central objects in the system. The background information is there mainly to support finding these objects. Although the result of any particular query may be any type of object, it is assumed that the result of any sustained interaction will be a document. The document hierarchy is fairly straightforward, and was derived from Bibtex with a few extras added for web support.

Modalities are an object type which are used to represent the physical manifestations of a document in the real world (Welty, 1998). A single document, e.g. a book, may have a

published hard-copy version, an on-line PDF version, an on-line HTML version, and an audio-cassette version. Each of these real things are *the same book*, they share the author, title, subject, they contain the same information. They each also have attributes of their own; on-line versions have a URL and format, hardcopy books have a shelving code, etc. In addition to making more sense, breaking these attributes out of the document object makes it possible to represent the fact that several web pages may be the same thing, and therefore returning a single "hit" for a query instead of one per web page helps prune and narrow search results.

Event/Objects are the background information that enhances the catalog and makes browsing possible. Events are things like conferences, workshops, etc., that often result in people getting together exchanging information and publishing papers. It is not unusual, for example, for a person to search for a paper they remember *hearing* at a conference, without remembering the author or title. Objects are things like people and organizations.

Relationships

Each type of object has a set of attributes and relationships that help define it. To explore the ontology and see the relationships defined for each type of entity, in the demo you may click on "list the main objects in the ontology" and then click on any concept. Another good starting place is "explore the Untangle topic space" select some topics, and then look at the entities classified under that topic. Each entity has links for all the entities it is related to.

The complete set of relationships for each object can be accessed through the demo, we provide here the basic set of relationships for navigating through the space.

Articles have an *author*, and are *published-in* some kind of *COLLECTION*, such as a *JOURNAL-ISSUE* or a proceedings. Collections may have editors, and are published-by an organization. A journal-issues may be an issue-of a journal, and a proceedings is the proceedings-of a conference. A conference has participants, a location, and is sponsored by an organization. A person may have an affiliation, and may be a student-of some school. We also support former-employee-of and former-student-of.

Reasoning

In addition to supporting combined querying and browsing, the new ontology supports reasoning to enrich the information space. The main point of using a KR system instead of a database is that reasoning can be used to enhance the data that can be automatically mined from existing bibliographic databases. The bulk of the reasoning is used to support the spatial representation of topics used in the system, and to support data consistency. We describe here a few examples of the reasoning over the types described above that are used to augment the data:

- A publisher is any organization that publishes something.
- Two people are collaborators if they are authors of the same document.
- A person participates in a conference if they author a

paper published in the conference proceedings.
- A person is interested in a topic if they write a paper on that subject.
- If a person is an employee-of a department, they are an employee-of the organization the department is a part of.

Demo Tour

Click on the small blue "home" icon at the bottom of the screen to get to the untangle home page, or go to the URL: http://untangle.cs.vassar.edu/. The first three links in the list you see are the main links for exploring the demo.

Let's say you're looking for articles on ontologies for information systems. You remember hearing a talk about this at a conference in Florida some years ago.

1. Hit the "query" link on the Untangle home page.
2. Select conference from the taxonomy (under event).
3. Type "Florida" into the text box labelled "location", and click "do it." The result is a list of conferences in Florida.
4. Click on "FLAIRS-96" (you can select it with the radio button, but in this case clicking on it is a shortcut). The result is a dynamically generated page that describes all the information the system has about the conference. At this point you may recognize someone in the list of participants. Let's say you don't.
5. Click on the proceedings. This results in a list of articles published in the proceedings (and in the knowledge base). You notice one, "Intelligent Assistance for Navigating the Web." which sounds familiar.
6. Click on it. You get information on this article. Aha! That's the one. Now you see in the list of topics "formal ontologies in information systems."
7. Click on that topic, and you get a list of people, events, publications, and organizations that have been classified under this topic.

Continue to explore. The icons on the bottom of every page go to the home page, to the query page, and to the FAQ page.

References

Welty, Chris and Jenkins, Jessica. An Ontology for Subject. *J. Data and Knowledge Engineering.* **31**(2):155-182. September, 1999. Elsevier.

Welty, Chris and Ide, Nancy. Using the Right Tools: Enhancing Retrieval of Marked-Up Documents. *J. Computers in the Humanities.* Summer, 1999. **33**(10):59-84. Kluwer.

Welty, Chris. Towards an Ontology for Library Modalities. In S. Ali, ed. *Proceedings of the AAAI-98 Workshop on Representations for Multi-Modal Human-Computer Interaction.* AAAI Press. July, 1998.

Welty, Chris. Intelligent Assistance for Web Navigation. *Proceedings of the 1996 Florida AI Research Symposium.* May, 1996.

Welty, Chris. A Knowledge-Based Email Distribution System. *Proceedings of the 1994 Florida AI Research Symposium.* Journal of AI Press. May, 1994.