A Novel Knowledge Organization Scheme for the Web: Superlinks with Semantic Roles

Abstract
We discuss the needs of a knowledge organization scheme for supporting Web-based software applications. We show how it differs from traditional knowledge organization schemes due to the virtual, dynamic, ad-hoc, user-specific and application-specific nature of Web-based knowledge. The sheer size of Web resources also adds to the complexity of organizing knowledge on the Web. As such, a standard, global scheme such as a single ontology for classifying and organizing all Web-based content is unrealistic. There is nevertheless a strong and immediate need for effective knowledge organization schemes to improve the efficiency and effectiveness of Web-based applications. In this context, we propose a novel knowledge organization scheme wherein concepts in the ontology of a domain are semantically interlinked with specific pieces of Web-based content using a rich hyper-linking structure known as Superlinks with well-defined semantic roles. We illustrate how such a knowledge organization scheme improves the efficiency and effectiveness of a Web-based e-commerce retail store.

Introduction
The Web is a vast, highly dynamic, multilingual, multicultural social knowledge repository. It is forcing rapid changes in the ways by which knowledge is created, organized, shared and found. For the most part, the Web has grown both in size and in the variety of its uses without waiting for underlying theories, structures, standards and schemes of organization to catch up. Significant portions of the content on the Web are not well-edited or well-organized. Yet, its social and democratic nature coupled with omnipresent functionality such as search and navigation has created a vast level playing field for mankind to manage its knowledge in previously unthinkable ways. It is impractical, at this point of the growth of the Web, to propose a universal knowledge organization scheme for the entire Web.

In the present work, we focus on Web-based software applications such as electronic-commerce stores on the Web and investigate how knowledge is organized in those Web sites. Finding several deficiencies therein, we propose a novel knowledge organization (KO) scheme for such Web-based applications. The primary objective of a Web-based store (say for books or electronic gadgets) is to ensure that customers who visit their Web sites find (and purchase) what they are looking for. Additionally, it is important that customers do not place orders for incorrect or incompatible products given that they do not have the benefits of either examining the products physically or consulting a human salesperson before placing their orders. How should the store organize all of the knowledge about their products so that these requirements are met?

Currently, most Web-based stores are using ad-hoc taxonomies to classify their products. To overcome any inefficiency in navigating through the taxonomical categories, they provide a keyword search facility, and, sometimes even links to related categories and products. Links are standard hyperlinks, that is, clicking the link takes the user to a single approximate location on the Web. When a user has a more specific need, for example, to select a battery that is compatible with the particular model of a cell phone, even where such knowledge is available, it is usually embedded deeply in a document on the site that outlines all the specifications of the cell phone or the battery in unstructured content. The KO scheme is typically not rich enough to provide a direct link clicking on which takes the user to all compatible accessories for the product.
Additional reasons for requiring a good KO scheme for the Web include improving search quality, ease of browsing, enabling effective social knowledge management, better social networking and enabling newer semantic applications on the Web.

**Six Principles of Knowledge Organization for the Web**

In this section, we attempt to summarize our observations about the divergent needs of knowledge organization for traditional and Web-based applications in the form of six principles. It may be noted that the following principles are significantly different from those proposed earlier for organizing portals on the Web for information access by particular communities (Rowley, 2010).

**Principle 1: Virtuality:** *Knowledge Organization for the Web is virtual, not physical.*

In Web-based applications such as e-commerce, there is no physical organization of knowledge assets that could serve as a guiding structure for KO. At the same time, there are no constraints imposed by physical structure on KO. For example, the same item can be present simultaneously in multiple places in the virtual showcases of an e-commerce store which is impossible in a physical organization scheme.

Further, there is no constraint imposed by the structure of data or metadata. In traditional computerized applications, for example, the structure of data structures, tables and databases often constrains what information can be added. The inherently unstructured nature of Web-based applications with their flexible structures such as non-relational data stores and XML-based semi-structured schemas pose no such problems.

**Corollary:** *KO for the Web is inherently different from KO for primarily physical collections.*

In addition to virtuality, Web content is distributed, highly dynamic and large scale (also known as “Web-scale”). It often uses an uncontrolled vocabulary with no single acceptable hierarchy of well-defined concepts. The richness of meta-data and the resulting fine granularity of classification are also distinguishing features of KO schemes for the Web.

**Principle 2: Multiplicity:** *KO for the Web needs multiple classifications.*

Although a product may belong to only one category from an ontological point of view, single categories are not very effective in ensuring that customers can find the product easily. Nor is it possible to design a single set of facets that effectively classify all the products in a Web store.

The Web provides multiple ways of finding a resource. For example, if we perform a search on an e-commerce site like Amazon.com for a book titled “Diary of a Wimpy Kid,” we find that Amazon places this book under multiple categories such as *Children’s Books, Humour, Social Situations* and *Friendship*. Such a strategy of placing an item under multiple categories for better access is followed on almost all e-commerce sites. It is the “findability” of the book that matters to a user when it comes to retrieval.

**Principle 3: Ad-hoc-ness:** *Impure categories are necessary.*

Ontological purity in the form of set-theoretic subsumption relations among categories and sub-categories is less important than usability of the categories. For the needs of Web applications, we often have to create ad-hoc categories that are ill-defined (e.g., “hot-selling products,” “most-emailed stories,” etc.). Categories must be intuitive, natural, using common terms, readable, and easy to remember and recall. Standardization is an unrealistic goal in the rapidly changing world of Web commerce. What matters is how a category is going to be used in practice.
**Principle 4: Customization:** *User-specific and localized categories are necessary.*

Users need to be provided categories that are to their liking and are most appropriate to their specific local context. This is not the same as the community principle (Rowley, 2010). In an e-commerce application, for example, there may be no community, just individual customers who are often anonymous. In some situations, there may be extensive knowledge about customers from which one could recognize various demographic patterns. Continuing with the example of “Diary of a Wimpy Kid” on Amazon.com, hundreds of users have tagged the book under *Books for Boys, Boys Books, So Funny, Comics, Cartoon, Comedy* and *Jeff Kinney* (its author).

**Principle 5: Ontology:** *KO for the Web needs ontological structure.*

E-commerce on the Web requires organization of complex relationships and constraints among products. For example, cell phones, their hardware features, accessories and software applications have complex relationships of compatibility. Such interdependencies not only need to be represented in well-structured units in an ontology, they also need to be presented to the user through effective linking structures so that the user is guided efficiently in selecting the most appropriate products.

In addition, manual classification may be ruled out in other applications. If classification has to be automatic, it must be based on an ontology. Categories must not only be well defined, their discriminating features must be made explicit. This will enable the user to exploit the semantics of the domain to navigate efficiently to the desired products.

**Principle 6: Pragmatism:** *KO for the Web is determined by the needs of the Web-based application.*

Finally, all other aspects of the KO scheme must be determined with pragmatism as the key. Should a category be present? Should it have an alternative name? What kind of links should be there among knowledge elements? All such questions must ultimately be decided based on the needs of the particular Web-based application.

**Knowledge Organization for the Web: Superlinks with Semantic Roles**

Although the above principles highlight the importance of ad-hoc and user-specific categories, it does not mean that we can simply do away with knowledge organization and allow social tagging and folksonomies (Lambe, 2007) to take their place. While social tagging and democratic review and rating processes are useful and must be included in the overall scheme for KO, social anarchy cannot be the model for organizing knowledge. There is no guarantee that such a loose scheme can actually deliver acceptable results. At the same time, the variety of KO schemes that have been proposed for traditional as well as Web-based applications (Garshol, 2004), such as decimal and faceted classification, Dublin Core meta-data, topic maps, concepts maps, RDF, and so on, also do not fully meet the requirements of a Web-based application.

Elsewhere, we have demonstrated that conventional hyper-linking on the Web is inadequate for semantics and KO purposes and proposed a new kind of semantic, multi-way linking known as Superlinks (Mahesh and Karanth, 2012; Karanth and Mahesh, 2012). We show here that Superlinks are ideal as a KO scheme for Web-based applications. A Superlink is n-ary in nature in that we can link one source to multiple target locations while also capturing the context in which link is relevant by specifying the semantic role for each target. The source and destinations of the link can be at any level of granularity ranging from a single word or phrase to a paragraph, section or the entire document itself.

Consider a Web-based e-commerce site for selling cellular phones along with their accessories and software applications. The site has to make it easy for users to match specifications of cell phones with those of accessories to determine compatibility between
an accessory and a particular phone model. There is no ready-made set of subject headings or a standard set of facets for a domain like this. Nor is the content made up of well-edited, self-contained or properly interlinked documents. What we have is a mix of product-related documents (e.g., specification sheets), data in the form of tables and charts, and much unstructured content in the form of customer reviews, questions and comments. Nevertheless, the richness of the domain as well as the context of a user’s interaction with the site must be captured by the KO scheme applied to organize all of this content in a manner most suitable for the user.

We propose that a KO scheme for this application must have these features:

- A domain ontology that interlinks all of its key entities and concepts (e.g., cell phone, operating system, application, capacitive touch screen, battery life, etc.);
- A linking mechanism that captures the semantics of the links;
- A multi-way and precise linking mechanism to interconnect several related pieces of text at a fine level of granularity; and
- Means for enabling users to comment, review and rate any part of the KO.

We build an ontology in OWL language covering all cell phones present in the Web-based store, their attributes like type, operating system and battery life, and features such as camera, MP3 player, Bluetooth, Wi-Fi, GPS and USB as well as different accessories for the phones. Detailed specifications of various phone models and accessories are available as semi-structured Web pages and PDF documents.

How do we create a linking mechanism that permits complex, semantic linking between concepts in the ontology and precise locations in associated texts (i.e., Web pages) as well as among different places across Web pages? Clearly, the existing hyperlink structure is inadequate since it neither allows multi-way links nor any semantic attributes of the links. We propose a type of link called Superlink that is built using the XML family of technologies including XSD, XLink and XPath (Harold and Means, 2001).

The key to establishing the context in describing the knowledge resources on the Web site is to link relevant concepts in the ontology to their sources in textual documents. A “Superlink” captures the semantic relations between concepts and pieces of text by specifying the semantic role played by the link. An application-specific typology of semantic roles can be employed, for example, definition, illustration, example, compatibleWith, etc. Superlinks with such semantic roles permit ad-hoc, impure and user-specific ways of linking source and targets across the Web site.

Figure 1 gives a pictorial representation of how Superlinks are created between specifications of phones and accessories. It shows a Web page about a cell phone that is linked together to two locations in two different pages where compatibility of the phone with related accessories is mentioned. A particular cell phone’s specification can be linked to multiple accessory specifications based on the semantic role CompatibleWith. The example shows how “HTC Wildfire S” specification can be linked to its multiple accessories like “HTC Portable Bluetooth Conference” and “Stereo Headset” using a single Superlink with CompatibleWith as the semantic role between the source and the two targets.

Figure 2 shows an XML encoding of the Superlink that includes two locator tags along with attributes such as xlink:show and xlink:actuate that specify what actions the Web browser should take when the Superlink is clicked. In addition, xlink:role specifies the semantic role of the link between the source page and that particular target which is itself linked by xlink:href, the target URI.
Fig. 1: Three Web pages linked together with a Superlink

Semantic Role - "Compatible with"

HTC Wildfire S
Live it. Love it. Share it.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>6.0 x 3.5 x 0.7 inches</td>
</tr>
<tr>
<td>Display</td>
<td>3.2 inch, 320 x 480 pixels</td>
</tr>
<tr>
<td>CPU</td>
<td>Qualcomm Snapdragon 1.0 GHz</td>
</tr>
</tbody>
</table>

Description
Contact two people at a time and talk hands-free. The HTC BD-300 Bluetooth Speaker connects to your phone or other Bluetooth enabled devices making your mobile office a hands-free zone.

Technical Details
- 30 feet in diameter; height 30mm (1.18")
- Talk time: 10 hours
- Conference call time: 10 hours
- Playback time: 8 hours
- Standby time: 100 hours
- Auto-pairing

Also Compatible With
- HTC Desire X
- HTC Desire 300
- HTC Desire
- HTC One
- HTC EVO
- HTC Captivate
- HTC Hero
- HTC Tattoo
- HTC Magic
- HTC Tattoo 2
- HTC Desire S
- HTC Desire HD
- HTC Legend
- HTC Freestyle
- HTC Hero S
- HTC Hero Q

Fig. 2: Superlink with multiple locators for CompatibleWith relation

<Superlink xlink:type="extended" xlink:show="new" xlink:actuate="onLoad">
</Superlink>

Live it. Love it. Share it.
Size: 101.3mm x 59.4mm x 12.4mm 3.99" x 2.34" x 0.49" Weight: 105 grams (3.7 ounces) with batteryDisplay: 3.2-inch touch screenScreen: 81.3 mm (3.2")
Conclusion

In this paper, we have summarized the knowledge organization requirements of Web-based applications in a set of six principles. Further, we have proposed ontology-based semantic Superlinking as the KO solution for applications such as Web-based e-commerce. Further work is needed to solve the challenges of engineering large-scale solutions based on this framework. Attempts must also be made to integrate such solutions across multiple Web-based services so that a common ontology as well as cross-links can be developed across compatible services.

References


