REPRESENTATION AND STRUCTURE OF INFORMATION THOUGH SEMANTIC WEB

Deepak Ahuja

Ph.D Research Scholar

ABSTRACT

The Semantic web play an important role for interactive and different transaction that takes place on the internet This paper will give an overview how the information retrieval and representation of information will take place in world wide web.

Keywords: Ontology, multimedia, world wide web, semantic web.

I INTRODUCTION

World wide web has millions of pages that are dynamically generated from content stored in databases. This not only makes managing a large site easier, but is necessary for fully functioning ecommerce and other large, interactive websites.

These local databases, in one sense, are not full participants in the web. Though they present normal looking HTML pages, the databases themselves are not interconnected in any way. Organization A has basically no way of using or understanding Organization B's data. If these two want to share or merge information, the database integration would be a fairly significant undertaking. It would also be a one time solution. If Organization C entered the picture, a new merging effort would have to be undertaken. As the web stands, this has not been a significant problem. By design, the web has been a vehicle for conveying information in a human readable form – computers had no need to understand the content. As dynamic sources of information have become omnipresent on the web, the World Wide Web Consortium has undertaken efforts to make information machine readable. This technology, collectively called the Semantic Web, allows computers to understand and communicate with one another. For site designers, this means data from other sites can be accessed and presented on your own website, and your own public data can be made easily accessible to anyone. It follows that just as web pages are currently hyperlinked, data can also be linked to form a second web behind the scenes, allowing full across-the-web integration of data. Dynamically generated pages driven by databases are becoming commonplace for most large websites, and even for medium and small ones. This trend, combined with the proliferation of non-text media files as one of the primary forms of content, poses several problems to the current web architecture. Search engines have difficulty indexing database driven pages, and cannot directly index the raw data used in the back end. Media searches, such as image or MP3 searches, are notoriously bad, because there is no text from which to extract keywords that could be used to index the media. The nature of the web, with interconnected information, does not extend backend databases or media either. It is usually not possible for a web designer to use

information from an external database to drive their own site. The database are not publicly accessible for queries, nor is the underlying organization of the database apparent.

II SEMANTIC WEB

The Semantic Web is a vision for the future of the World Wide Web that will give meaning to all of this data, as well as making it publicly accessible to anyone who is interested. While some web sites and designers will want to keep their backend data proprietary, many will find it in their interest, for design and public interest, to use semantic encodings.

Semantic web, explain how to organize content for use on the semantic web, and show several examples of how it can be used. Throughout the discussion, we will describe how the technologies affect the human factors in web design and use. What is the Semantic Web?

The World Wide Web today can be thought of as a collection of distributed, interlinked documents, encoded using (primarily) HTML. Any person can create their own HTML document, put it online, and point to other pages on the Web. Since the content of these pages is written in natural language, computers do not “know” about what is in the page, just how it looks. The Semantic Web makes it possible for machine-readable annotations to be added, linked to each other, and used for organizing and accessing Web content. Thus, the Semantic Web offers new capabilities, made possible by the addition of documents that encode the "knowledge" about a web page, photo, or database, in a publicly accessible, machine readable form. Driving the Semantic Web is the organization of content into specialized vocabularies, called ontologies, which can be used by Web tools to provide new capabilities. In this section, we present the basic ideas underlying ontologies, what they can be used for, and how they are encoded on the Web.

III ONTOLOGIES AND VOCABULARIES

Ontology is a collection of terms used to describe a particular domain. Some ontology is broad, covering a wide range of topics, while others are limited with very precise specifications about a given area. The general elements that make up ontology are the following: Classes – general categories of things in the domain of interest Properties – attributes that instances of those classes may have The relationships that can exist between classes and between properties Ontology are usually expressed in logic-based languages. This allows people to use reasons to analyze the anthologies and the relationships within them. XML (extensible Mark up Language) exists to add metadata to applications, but there is no way to connect or infer information from these statements. By encoding these relationships in a logic based language, these and other more interesting inferences can be made. Reasoners are the tools that use the logical statements to make inferences about classes, properties, instances, and their relationships to each other.. These reasoners can be used in advanced applications, such as semantic portals or intelligent web agents. Making data on the web accessible for these types of services and applications, ontologies, and languages for developing them, is a focus in the emerging Semantic Web. Motivations The “formal” models of the domain enabled by the ontologies provide a number of new capabilities, but also require extra work with respect to entering the metadata appropriately, developing the vocabularies, etc. To justify the significant added effort required for good encoding of the semantics behind a given application, users should

understand some of the benefits that will be available, and doors that are opened. There are many places where the semantic web can improve the way things are done on the web now, and add new capabilities beyond what is available now on the web. The following sections enumerate some of the visions for the Semantic Web as put forth by the World Wide Web Consortium's Web Ontology Working Group in a document outlining use cases and requirements for ontologies on the Web .

IV WEB PORTALS

A Web portal is a web site that provides information content on a topic. While the term has become common for full-web search engines such as Google, portals in the traditional sense are also domain specific pages that do not necessarily have a search feature. The goal is to provide users with a centralized place to find links, newsgroups, and resources on topic.For portals to work well, they need to be good sources of information to encourage the community to participate in maintaining and updating their content. To create a semantic web portal, where information is well annotated and maintained in a semantic web format, the same is true. Users need some motivation to do the mark up that makes the site work. The vision for semantic web portals is to not only make them available as online web pages, but also to integrate them into tools. On web pages, users can find resources based on their semantic mark up. To encourage users to create their own metadata, tool integration of portal features is key. For example, if a scientist authoring a paper or web page uses a particular term from an online ontology, the semantic web portal feature should return other sources with similar mark up. Results would most certainly return related web pages. They will also provide links to images, video, audio files, or datasets whose content is described by the same term. By these sorts of providing useful information and resources, which could not be found with a standard text-based keyword search, users will be encouraged to mark up their documents so that they make take advantage of the portal. What allows this system to work more fully is the integration of the mark up process with the portal. The portal provides the most advantage to users while they are creating their own semantic web documents. Thus, after providing information to the user, the portal itself is extended when the new mark up is published. This interactive cycle means that semantic web portals will reach out to incorporate external resources, as well as creating a dense web of semantically interlinked documents.

V MULTIMEDIA COLLECTIONS

Ontologies can be used to provide semantic annotations for collections of images, audio, or other non-textual objects. Though one may choose to argue for keywords and natural language processing for computers to understand text documents, extracting information about media is much more difficult. Though some file formats do contain information about the file and the media, there is no way for a machine to understand what is happening in a picture, or the significance of who is pictured. Ontologies to describe media and its content address this problem. Multimedia ontologies can be of two types: media-specific and content-specific. Media specific ontologies describe the format of files and related information. For an image, ontologic markup may include file format and file size, plus information about how the image was produced, such as the camera that took the photo or focal length. Content-specific ontologies allow an author to describe what the media is about. For a photo, this could include the date and time it was taken, where it was taken, who and what is in the picture, and what is happening. For other media, like sound, attributes like lyrics, chord progressions, or historical

information may also be relevant. Data about the contents can be related to detailed instances declared in other files,

VI WEB SITE MANAGEMENT

Websites for even small organizations can have large collections of documents which fall into many categories. These can include news releases and announcements, papers, forms, contact pages, and downloads. As the number of documents increases, finding them, without structure, becomes all the more difficult. Even a taxonomy with a strong hierarchical structure can be insufficient. This is clearly seen in web directories, such as Yahoo!, where finding a particular page is difficult, even in a subset of the hierarchy. An ontology-based web site allows users to search and navigate using specific, ontologically defined terms. This will make documents easier to find, and cross references easier to track down. Later on, this chapter will discuss one website using semantic mark up as its foundation.

VII DESIGN DOCUMENTATION

Documentation of systems is often very complex. Large sets of documents with overlapping scopes have several presentation challenges. Since documents are generally grouped thematically, it is not unusual for several sets of documentation to address different aspects of the same sub problem. For a client who is trying to find data on the sub problem alone, there is sometimes no choice but to navigate through several sets of complex documents. Even when the desired information is contained in one set, the level of detail can often be overwhelming.

VIII CONCLUSION

This paper described the basic fundamental of ontology technique used with reference to the world wide web. The information available on the web can be used for information representation and formation of the structure. Therefore we can say that the ontology based web information retrieval is used for the semantic web.

REFERENCES

[1] http://img.cs.man.ac.uk

[2] http://www.jan-winkler.de

[3] http://www.daml.org

[4] http://www.cs.umd.edu

[5] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the el envelope. Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05), 2005.

[6] Sean Bechhofer and Raphael Volz. Patching syntax in owl ontologies. Proceedings of the 3rd International International Semantic Web Conference, 2004.

[7] Thorsten Liebig and Olaf Noppens. OntoTrack: Combining browsing and editing with reasoning and explaining for OWL Lite ontologies. Proceedings of the 3rd International International Semantic Web Conference, 2004.

[8] A. Magkanaraki, S. Alexaki, V. Christophides, and D. Plexousakis. Benchmarking rdf schemas for the semantic web. Proceedings of the 1rd International International Semantic Web Conference, 2002.