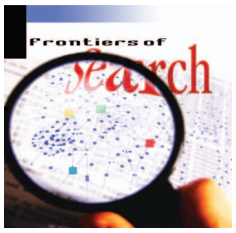
Search on the Semantic Web



To help human users and software agents find relevant knowledge on the Semantic Web, the Swoogle search engine discovers, indexes, and analyzes the ontologies and facts that are encoded in Semantic Web documents.

Li Ding
Tim Finin
Anupam Joshi
Yun Peng
Rong Pan
Pavan Reddivari
University of Maryland, Baltimore County

Search engines have assumed a central role in the World Wide Web's infrastructure as its scale and impact have increased. In the Web's earliest days, people found pages of interest by navigating (quickly dubbed *surfing*) from pages whose locations they remembered or bookmarked. Rapid growth in the number of pages gave rise to Web directories like Yahoo that manually organized Web pages into a hierarchy of topics.

As the growth continued, these directories were augmented by search engines such as Lycos, HotBot, and AltaVista, which automatically discovered new and modified Web pages, added them to databases and indexed them by their keywords and features. Today, search engines such as Google and Yahoo dominate the Web's infrastructure and largely define our Web experience.

Most knowledge on the Web is presented as natural-language text with occasional pictures and graphics. This is convenient for human users to read and view but difficult for computers to understand. It also limits the indexing capabilities of state-of-the-art search engines, since they cannot infer meaning—for example, does an occurrence of the word "raven" refer to the bird or to Baltimore's football team?

Thus, users share a significant burden in terms of constructing search queries intelligently. Even with increased use of XML-encoded information, computers still must use application-dependent semantics to process the tags and literal symbols.

SEMANTIC WEB SEARCH

The Semantic Web offers an approach in which computers can use symbols with well-defined, machine-interpretable semantics to share knowledge.[1] Search on the Semantic Web differs from conventional Web search for several reasons.

First, Semantic Web knowledge content is intended for publication by machines for machines—tools, Web services, software agents, information systems, and so forth. Although Semantic Web annotations and markup can help users find human-readable documents, there will likely be an "agent layer" between human users and Semantic Web search engines.

Second, knowledge encoded in Semantic Web languages such as the Resource Description Framework (RDF)[2] differs from both the largely unstructured free text found on most Web pages and the highly structured information found in databases. Such semistructured information requires using a combination of techniques for effective indexing and retrieval. RDF, RDF Schema (RDFS),[3] and the Web Ontology Language (OWL)[4] introduce semantic features beyond those used in ordinary XML, allowing users to define terms (for example, classes and properties), express relationships among them, and assert constraints and axioms that hold for well-formed data.

Third, even within a single document, Semantic Web documents (SWDs) can be a mixture of concrete facts, class and property definitions, logic constraints, and metadata. Fully understanding the

 Published by the IEEE Computer Society 0018-9162/05/$20.00 © 2005 IEEE