An open source search engine for electronic commerce

Utilizing Real-World Knowledge
to
Reduce Friction in Digital Commerce

Caboodle Networks, Inc.
April 2006

Visit: http://semantical.org
Contact: info@caboodlenetworks.com

Properly designed, the Semantic Web can assist the evolution of human knowledge as a whole.
– Tim Berners-Lee[1]

Introduction

Search and commerce are closely related. As users become more dependent on search, what they search for has extended beyond traditional text, to include images, games, music, video, and other non-traditional content types. Relying on text-based keyword search to find relevant non-traditional content is meeting with diminishing returns; multiple searches and filtering through thousands of products has become standard practice to find an exact item in one of these content types. Complicating this problem, is the fact that there does not exist a single definition or standard "tag" representing a non-traditional content type that any two machines anywhere on the Internet can agree upon.

The practical use of Semantic Web standards presents an opportunity to overcome these obstacles. The result of practical applications of semantic web standards in search and tagging is to provide a better user experience, and the potential for highly intelligent, relevant search results and recommendations. This technology is also useful to address the problem of "The Long Tail"[2] (desirable content that is not syndicated or mass-marketed) where resources are spent to build limitless shelf space on Internet and mobile portal.

The Semantic Web and Ontologies

The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework[3] (RDF) and Web Ontology Language[4] (OWL), which integrate a variety of applications using XML for syntax and URIs for naming. To remain competitive and achieve repeat business, user queries must return either exact matches or highly relevant recommendations to provide a great user experience. Customers will simply switch from whom they make their purchases if the vendor search results contain nothing of interest.

An ontology is an explicit formal specification of how to represent the objects, concepts, and other entities in some area of interest and their relationships. The vast amount of information created and placed for public consumption must be organized in an intelligent way to make it machine-readable, thus automating the mapping and pinpointing of relevant items for each query.

---

[1] "The Semantic Web," Tim Berners-Lee, James Hendler and Ora Lassila, Scientific American, May 2001
http://www.scientificamerican.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21
[2] "The Long Tail," Chris Anderson, *Wired Magazine, October 2004*.
http://www.wired.com/wired/archive/12.10/tail_pr.html
[3] RDF: http://www.w3.org/RDF/
[4] OWL: http://www.w3.org/TR/owl-features/

These technologies will be invaluable as we move forward in building digital storefronts accessible both from the mobile phone and the computer.

Caboodle Networks – Open source solution available now

Caboodle Networks provides a standards-based, comprehensive search and tagging solution for Internet and mobile portals to integrate a Semantic Web search solution which can find and recommend monetizable links or non-traditional content types, thus reducing the friction in digital commerce. Our patent-pending technology allows us to utilize software agents, ontologies, and aspects of the artificial intelligence to build a more powerful and customized solution specifically designed to address search on Internet portals and mobile devices. The advantage comes from an understanding of the data and the original query and not just matching keywords in a relational database. Our technology also can establish relationships between concepts and helps recommend related content and services – ultimately creating a "soft-landing" for subscribers instead of an empty search result. This is particularly helpful when catering to the shifting trends of what is hot or most popular, but not currently available in the catalog. Moreover, this feature allows the portal to track trends more dynamically and immediately, catering to the needs to maximize revenue and enhance the subscriber experience.

Powerful text-based search engines are showing signs of diminishing returns particularly as the number of items on a commerce portal increase:

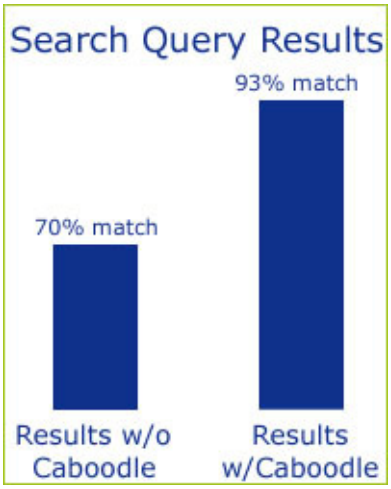
> "88% of sites return no results for simple queries when the content is actually on the site. When studying users…shopping for clothes, they abandoned their shopping 55% of the time queries returned no results. For shoppers of Books, CDs, and DVDs, they abandoned 100% of the time they encountered a no results screen. In almost all cases, what they were shopping for *was* on the site, so the sites lost the sales because their search didn't work effectively."[5]

Another study examined how users are evaluating search results and found that 41% of consumers changed engines or their search term if they did not find what they were searching for on the first page[6]. Lackluster performance is also occurring on the mobile Internet where users are looking for music, games, and images. A recent study[7] showed that 76% of mobile subscribers found it too complex to access and operate mobile content and services from their mobile device. Unlocking the value of the entire catalog and selling each and every item, and not just the Top 10 or Most Popular, will maximize the investment and increase ARPU[8].

---

[5] User Interface Engineering study on web on-site Search
[6] Jupiter Research and iProspect
[7] Wacom Components study, November 16, 2004
[8] Average Revenue Per User

A study was conducted to compare the existing text-based, keyword search with Caboodle Networks' Semantic Web-based search solution. A music catalog of over 5,000 entries from a wireless operator was chosen to provide an adequate number of ringtones to search. 4,654 queries were performed, first using the text-based keyword search, then Caboodle Networks' Semantic Web-based search. The results of the test overwhelmingly show a marked improvement. Caboodle had a 93% match vs. a 70% on a text-based keyword search. An even higher percentage match is achievable by augmenting the search with spellchecking.



By using Caboodle Networks' software, any type of content, such as games, photos, music, images, and video, can be described and more easily discovered. Not only can you find the content you want, but Caboodle Networks' software is designed to recommend intelligently related alternative content.

> **EXAMPLE:** On a clothing website, you put in the search term: "blue collared shirt" The site returns no results found. You perform another search on "collared shirt" This time several results are returned. Clicking on the first result, you notice that instead of a blue shirt, the color names are cyan and sapphire. Implementing Caboodle Networks' software, it interprets "blue, cyan, and sapphire" as equivalent colors that are closely matching; thus no empty results on the first attempt.

Caboodle Networks – Future-Proof

Caboodle Networks' extensible and flexible software architecture can quickly adapt to new services deployed and continue to provide highly relevant results to user queries. Our solution is based entirely upon semantic web standards. Additionally, as the mobile Internet evolves, the software can easily be integrated with a wireless portal or a wireless operator. Additional features such as our patent-pending Reverse Search™ can utilize user preferences, purchase history, and potential location-based information to provide a customized recommendation of content and services to the user.

The Open Source advantage

Caboodle Networks chose to release its software under the GPL license. By making ontologies open source, anyone, anywhere in the world, can create an expert knowledge model of a particular domain, and anyone, anywhere in the world, can then search this domain. Enabling a standard tagging approach for non-traditional content types enables anyone to create and register their own semantic web-standard tag. This offers two community benefits: First, it removes the dependency on a single company to create a proprietary scheme for finding information. Second, two

machines anywhere on the Internet can now agree on the same digital good or service and then transact (a machine-to-machine *lingua franca* for digital commerce, whether fixed or mobile).

Open source software presents many benefits to the developer and ultimately the user of the services. Studies have shown that Open Source software provides high performance, better functionality, higher stability as well as lower total cost of ownership. Caboodle Networks utilizes the proven open source software from MySQL and JBoss to create a viable solution that rivals much more expensive commercial software with similar performance and stability, but lower acquisition and ownership costs.

Benefits

The Caboodle Networks' software and the Semantic Web benefits three major groups of people:

Web (both wired and wireless) Portal Developers:

- A free, open source solution based upon Semantic Web standards, ideal for navigating multimedia content, such as games, music, video, and images that may be extended to the mobile subscriber, means that site operators no longer have to develop a search and recommendation solution for their content from scratch.
- Now the true value of the "long-tail" of content can be exploited by leveraging hidden relationships between and among content categories.

Wireless Operators:

- Significantly better user experience in finding mobile data creates increased usage and lower churn.
- Machine-to-machine digital commerce is enhanced by existence of open, standard content tags.

Users:

- The ability to quickly and easily find content and services, with more personal, relevant results and recommendations makes a better experience.