6S: P2P Web index collecting and sharing application

Computer Science Department, Indiana University, Bloomington
Lindley Hall 215, 150 S. Woodlawn Ave. Bloomington, IN 47405-7104

http://homer.informatics.indiana.edu/~nan/6S/

Contact: Le-Shin Wu, Ruj Akavipat, Filippo Menczer ({lewu,rakavipa,fil}@indiana.edu)

Domain: Applications, Distributed Systems, Information Search and Retrieval

Keywords: Peer collaborative search, small world networks, adaptive query routing

Objectives

6S is a collaborative peer network application, which aims to address the scalability and context limitations of centralized search engines and also provides a complementary way for Web search.

Functional description of the application

6S uses the idea of modeling neighbor nodes by their content but without assuming the presence of special directory hubs. As shown in Fig. 1, each peer is both a (limited) directory hub and a content provider; it has its own topical crawler guided by its user's information content and local search engine. Peers communication is built on JXTA platform (www.jxta.org). When a user submits a query, it is first matched against the local engine, and then routed to neighbor peers to obtain more results. Ideally, the peer network should lead to the emergence of a clustered topology by intelligent collaboration between the peers. While traditional search engines such as Google and Yahoo provide access to very large document collections, the 6S P2P Web search application provides a complementary way for users to actively and collaboratively share their own document collections. However, the 6S framework allows traditional search engines to naturally be included as peers; such peers would quickly emerge as reliable, trustworthy, and general authority nodes.

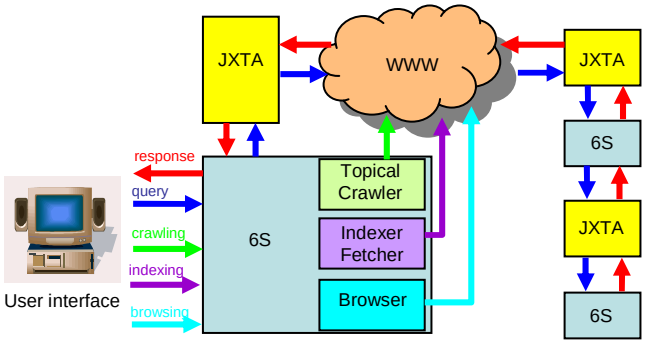

Figure 1: 6S architecture.

6S Features

6S is designed to make it easy and transparent for users to build and share a collection of Web pages, i.e. to build a micro search engine. The main features are Peer Search, Personal Web Index Management (auto & manual), and a Firefox Extension.

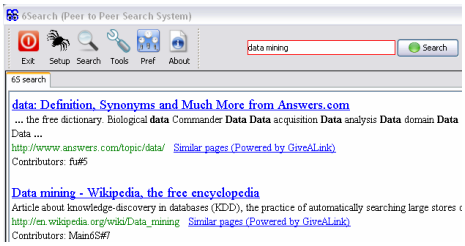

Figure 2: 6S peer search

Peer Search: When a user submits a query as shown in Fig. 2, in addition to getting results from the local engine, 6S automatatically determines which peers are best suited to answer the query based on its previous query-response experience, and forwards the query to those peers. The query and

answers received from each peer are analyzed and used to update local knowledge about what that peer has in its search engine index. This knowledge is then used to select peers for future queries.

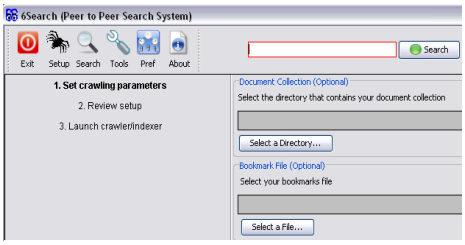

Figure 3: Creating a personal Web index

Personal Web Index Management: 6S automatically creates a Web index based on the user's bookmark file and Web search history during installation. Afterward, 6S will update the index by preodically adding new bookmarks or search history. Users can also manualy create or add to the personal index, as shown in Fig. 3, by running a ***best-N-first*** topical crawler [3], which crawls the Web in a more focused way guided by a provided topic. The crawling results will then be indexed for keyword searching.

The 6S index management tool displays all the documents indexed by the local engine (as showing in Fig. 4). For each indexed document, the user can assign tags, which are searchable by the local engine or modify existing tags. Users can also delete/undelete any document entries or remove the entire index. In addition, 6S integrates Luke (http://www.getopt.org/luke/) to provide more advanced index management options.

Figure 4: Index management

Firefox Extension: 6S FireFox allows access to 6S without leaving the Web browser. To search, users select the local or peer search options at the top right search box of the Firefox Web browser (as shown in Fig. 5) and type in a query. To export bookmarks to 6S, users simply click "Bookmarks" in a drop-down menu then select "Index All Bookmarks in 6S and Share" option. To index a single Web page, users can just right click on the page and select the "Bookmark in 6S and Share" option in the pop up menu. In order to use the Firefox extension, 6S application must be running as a background process.

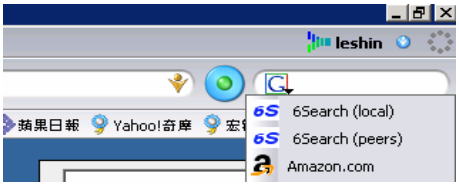

Figure 5: Firefox extension

Bibliographical References

[1] L.-S. Wu, R. Akavipat, and F. Menczer. 6S: Distributing crawling and searching across Web peers. In Proceedings of the IASTED International Conference on Web technologies, Applications, and Services, Calgary, Canada, July 2005.

[2] R. Akavipat, L.-S. Wu, F. Menczer, A. G. Maguitman: Emerging Semantic Communities in Peer Web Search. Proc. P2PIR workshop of CIKM 2006.

[3] G. Pant, F. Menczer: MySpiders: Evolve your own intelligent Web crawlers. Autonomous Agents and Multi-Agent Systems 5(2): 221-229, 2002

Current state of the application: beta version

Mode of distribution: Download at http://homer.informatics.indiana.edu/~nan/6S

Name of the main distributor:

Financial conditions (price, freeware, copyright): open source; ©2007 Le-Shin Wu, Ruj Akavipat, and the Trustees of Indiana University.

Obligations of the user (citation, copyright): citation

User and technical documentation: http://homer.informatics.indiana.edu/~nan/6S/document.html