A Latvian Company Aims at Google's Territory

Gints Ernestsons, CIO, Lursoft IT, Latvia

Common wisdom holds it that the Google search engine is based on superior technologies. Today, however, one can compare Google's results to those that are achieved with SIETS software – a search technology that has been developed by the Latvian company Lursoft IT.

The SIETS technology is being used on the Latvian Internet search portal www.siets.lv, on the largest full text archive of Latvian newspaper articles (www.news.lv), and on one of Latvia's leading general purpose Internet portals (www.apollo.lv), which is run by the national telecoms operator Lattelekom. The system is also used on the Internet portal of Latvia's central bank, www.bank.lv. Lursoft IT began to distribute the SIETS software as a commercial technology in September 2004.

SIETS – A NEXT-GENERATION CONTENT-BASED SEARCH ENGINE

"Siets" means "sieve" in the Latvian language. For thousands of years, people have been using the sieve as a simple and useful tool for retrieving food grains from raw crops. With information, the situation is very similar – you can sift through a lot of data to find useful grains of information. The SIETS software was developed so as better to address this task. It's a high performance search and indexing technology for business applications. The innovative software can be described as a next-generation content-based search engine – a combination of three emerging mainstream software technologies into a single product – XML document storage, full text search engine and clustering.

Almost every company and organization today finds itself collecting more and more data in the form of electronic documents, spreadsheets, database records, E-mails and various other corporate files. This is a process which will not disappear in the future. It is becoming more and more time consuming to search for information among these vast and growing volumes of data. Despite the availability of powerful computers, user demand for fast and relevant search results in corporate systems has increased much more quickly than existing technologies can support. The demand has been stimulated by popular Internet search portals through which users can quickly obtain links to important Web pages just by entering basic keywords as search queries. There is no doubt that the corporate search experience should match and exceed the leading Internet search services.

Many legacy data processing systems were not developed for full text search and indexing. Slow and inefficient search algorithms often prevail in these legacy systems. The architecture of the most common file systems and databases also serves to limit search speeds. Data are stored in various proprietary file formats and encoding. In many data centres, the true information processing potential of hardware is not fully utilised during search operations. Sometimes the capacity is expended unnecessarily by scanning through gigabytes of accumulated data.

HOW DOES THE SIETS SERVER WORK?

The SIETS Server can store and search data objects when they are represented as a collection of simple XML documents. Web pages, Word documents, SQL data records, E-mail messages and other files can be wrapped in XML and stored in the SIETS Server database. The SIETS Server automatically creates an inverted index for all of
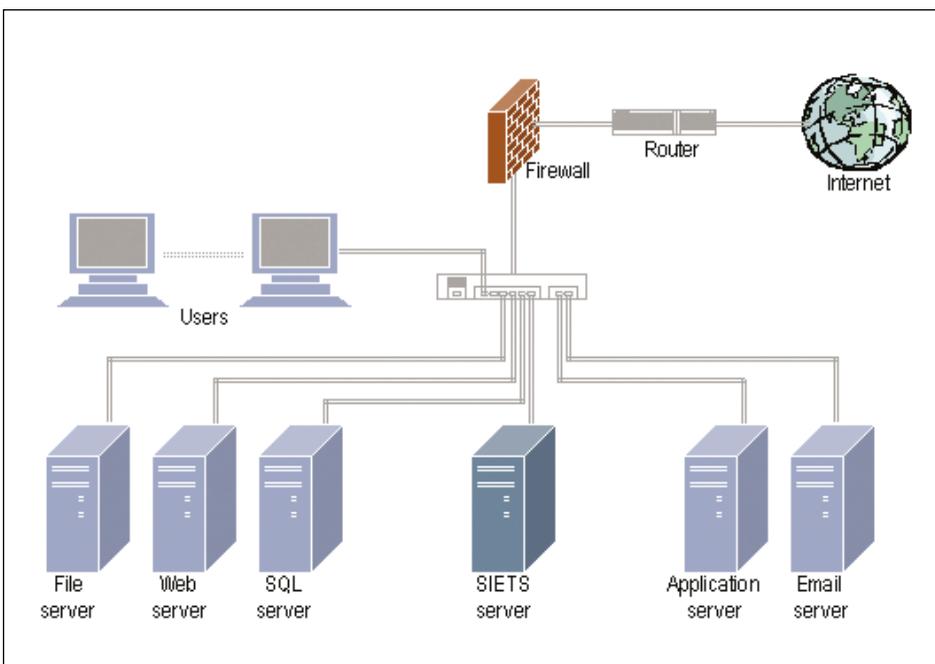

Figure 1. The Siets Server can operate in a corporate network as a search tool

the stored documents so that very rapid full text searches become possible.

The SIETS Server uses a client-server architecture that is based on simple XML messaging (SIETS API) and on a network-friendly http protocol. Search results are returned as XML-formatted replies. The use of open standards allows for the installation and use of the SIETS Server in any corporate data centre. It can run as an enterprise search appliance or as a software server. Clients can access the SIETS Server directly through any favourite programming language of environment (C, Java, Perl, Php, Microsoft NET languages, etc.).

The SIETS Server can index and search millions of documents on a single computer. There are more than 50 enterprise search and indexing options that are available for developers – word and phrase search, multi-level Boolean search, the use of wildcards, highlighting of hits in search results, Web-friendly navigation of results, support for more than 160 languages, real-time indexing, metadata support, automatic vocabulary building, spell-checking, sorting of results by relevance and rating, return of text snippets around matches, proximity and content similarity searches, indexing of different document formats, server security and integrity controls, extensive logging, etc.

IMPROVE PRODUCTIVITY AND COST CONTROLS

The SIETS Server software runs on commodity PC hardware and Linux. Growing performance needs can be addressed by clustering many low-cost computers which run parts of the same database. The software can run as an enterprise search appliance or as a farm of clustered servers. The SIETS Enterprise Manager tool is used for centralised Web-based administration, configuration and monitoring of all SIETS Servers in the corporate network. The SIETS Crawler utility is used for collecting and indexing of documents from corporate Web sites and file systems. The software can achieve search speed and reliability in corporate application environments which competes successfully with the leading Internet search portals.

For more information about the SIETS technology, please see the SIETS Server product site on the Internet – www.siets.net.
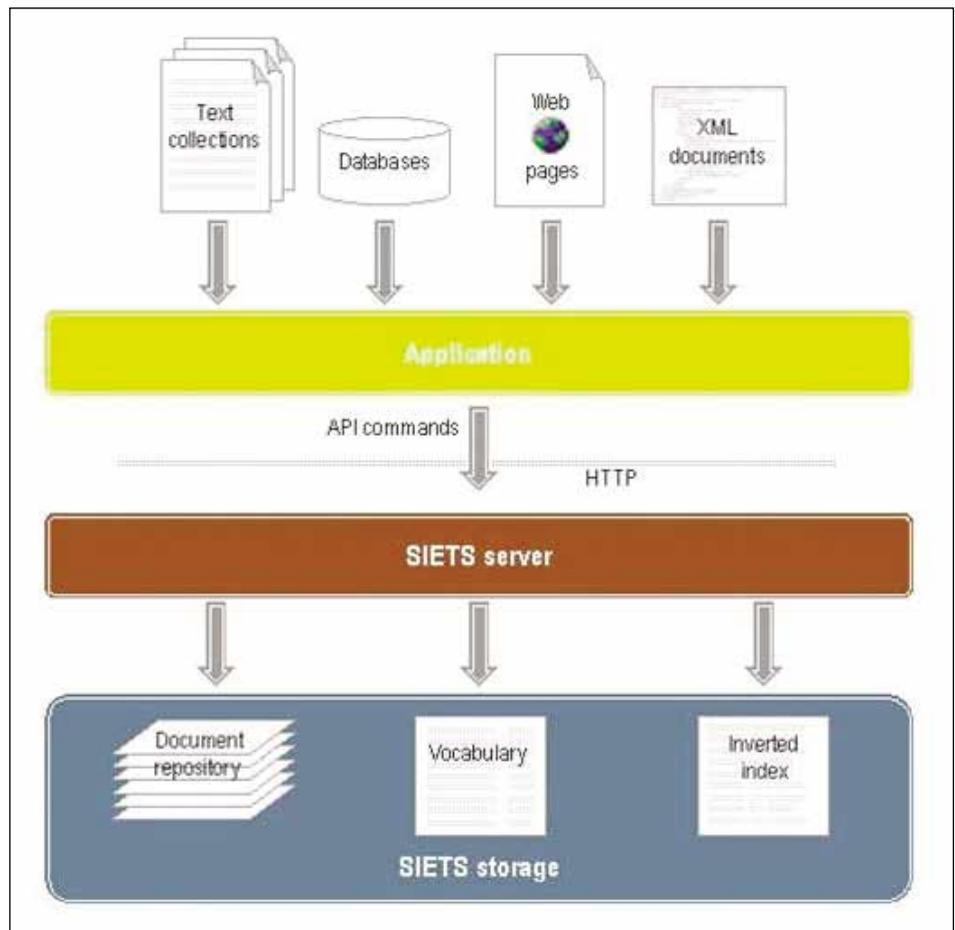
Figure 2. The Siets Server automatically builds an inverted index for all documents

MARKET OPPORTUNITIES

Lursoft foresees good market growth opportunities for corporate search engines over the next three to five years. A search appliance is becoming the norm in many corporate data centres. As the SIETS Server core software is 100% owned by Lursoft, the software can be licensed and used as a versatile platform for the building up of different high-value application products for corporate use, e.g., document archives, data backup solutions, mail appliances, global or national Internet portals, text retrieval libraries, or other products which require instant access to data. The Lursoft SIETS technology costs a fraction of what other leading search technologies charge for their products. Companies can achieve a better return on investment by using the SIETS Server. It provides search speeds similar to Google's, as well as results that are at the high level of quality which end users today expect from everyone in the Internet age.

LURSOFT

Lursoft and its sister company, Lursoft IT, are information technology companies with more than 12 years of experience in information processing and software development. Since 2001, Lursoft IT has regularly been listed in the Fast50 and Fast500 nominations of the most rapidly growing technology companies in Central Europe, as compiled each year by Deloitte & Touche. In 2003, Lursoft It was also included on an "all star" list of technology companies by the same firm. Lursoft is a member of the Latvian Information Technology and Telecommunications Association. ❏