

Inverse Inference Engine for High-Precision Web Search

Enhanced Search Engine Provides Better Intelligence Faster

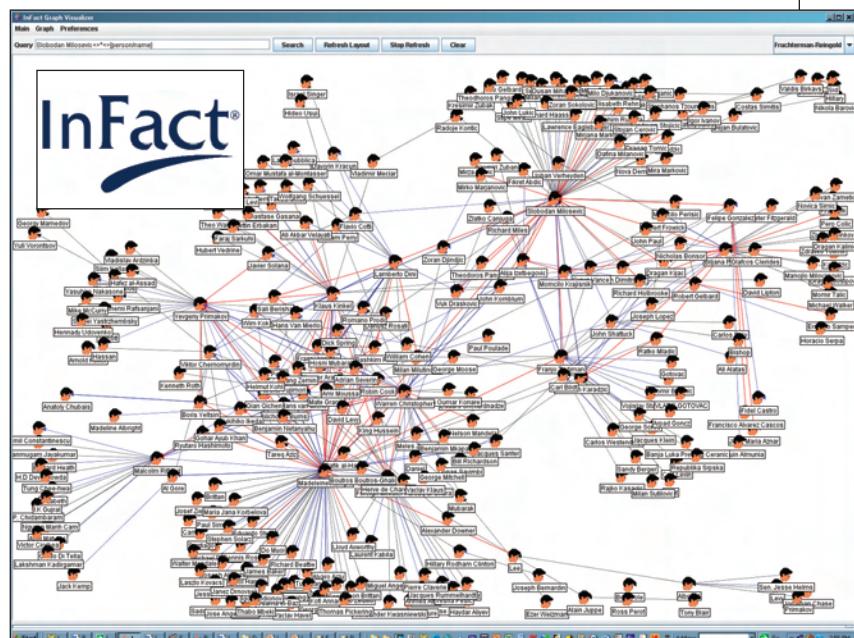


Technology and Innovation

The technology developed by Insightful Corporation (Insightful) under DARPA's SBIR helps analysts work through documentation more quickly and efficiently than traditional technologies, such as keyword search or simple entity extraction software. The program comprises two distinct parts: (1) a new algorithm for latent semantic analysis, and (2) deep parsing technology for relationship extraction from unstructured information.

Latent semantic analysis is a promising technique with applications to enhanced keyword retrieval and cross-lingual retrieval. A number of government laboratories have built prototypes based on the '90s-era Bell Labs Latent Semantic Indexing (LSI) algorithm. Insightful's development of a new algorithm, labeled Latent Semantic Regression (LSR), is scalable to large data sets and forms the basis of the company's InFact® product line. InFact® is a production quality, end-to-end deployable software system that performs relationship extraction, text mining, and search.

According to Giovanni Marchisio, Vice President of Engineering-Text Analysis, LSR represents a major step forward for the company. "The creative leap involved a deep understanding of computational techniques. We realized that the LSI algorithm, which everyone else uses to perform latent semantic analysis, is only a special solution of a

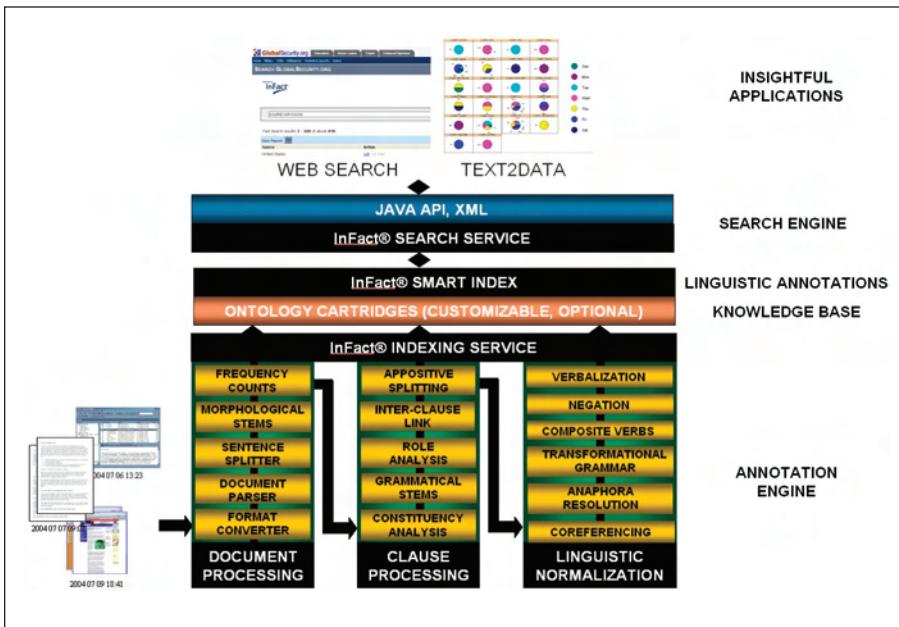


The results of an InFact® query showing relationships between individuals

more general computational problem. Our prototype replaced the core algorithm for LSI with the best of many computational alternatives that are available to solve the more general problem, resulting in a substantial increase in scalability and in an improvement in speed of several orders of magnitude."

The deep parsing effort initiated under DARPA's SBIR program produced the first prototype of a search engine that can understand and retrieve relationships facts and trends. Intelligence analysts and pharmaceutical researchers have requirements for search functionality that goes well beyond simple keyword search and involves an understanding of relationships and facts.

InFact® understands far more than keyword searches—it understands facts and actions. In a search for 'blackhawk,' a type of military helicopter, InFact® generates an aggregated list of all combat situations where the Blackhawk helicopter was deployed in the recent past, including all locations and military units that employ it, along with information about acquisition and maintenance costs and much more.



Insightful's technology is finding its way into a number of military applications, including use as an advanced search engine and as an information extraction tool. In information extraction, InFact® can process millions of heterogeneous document sources and automatically populate databases with information it has understood and extracted across all documents it has read.

The company currently has a contract for information extraction with the U.S. Army Corps of Engineers. The effort focuses on automating the labor-intensive process of tagging geographic information in text files, interactively editing the tagged files, and displaying the tagged text and associated map. The U.S. Army estimates that 80 percent of its data contains spatial reference information in the form of addresses, place names, or coordinates. In order to support spatial analysis of text sources, our approach requires structured text files with place name tags containing or referencing coordinate information. While tools exist for automatically identifying place names in text, our efforts are focused on significantly improving current

capabilities within an end-to-end system. By reliably automating these processes, InFact® has the potential to save the Corps of Engineers and other government agencies millions of dollars per year.

The company's technology is also being used by:

- The Air Force Research Labs, where researchers are developing a system for rapid diagnosis and improved screening of personnel that have been exposed to toxic agents in the field.
- Major pharmaceutical companies to generate search and retrieve relationship-based summary reports of millions of biomedical journal articles and reports.
- A major food manufacturer to extract facts from consumer reports.

Joint Collaborations

The work Insightful has done on its LSR algorithm and InFact® product has spawned a number of SBIRs, including:

- U.S. Air Force
- National Institutes of Health
- U.S. Army Corps of Engineers
- Office of the Secretary of Defense/Army Research Institute

Additionally, Insightful is collaborating with several major systems integrators, including Science Applications International Corporation (SAIC), Mantech and Booz Allen Hamilton.

The ingestion engine's map for understanding semantic and syntactic relationships

Insightful Corporation

Insightful seeks out new military applications by making the upper echelons of the Department of Defense aware of the technology's practical and strategic potential.

Lessons Learned

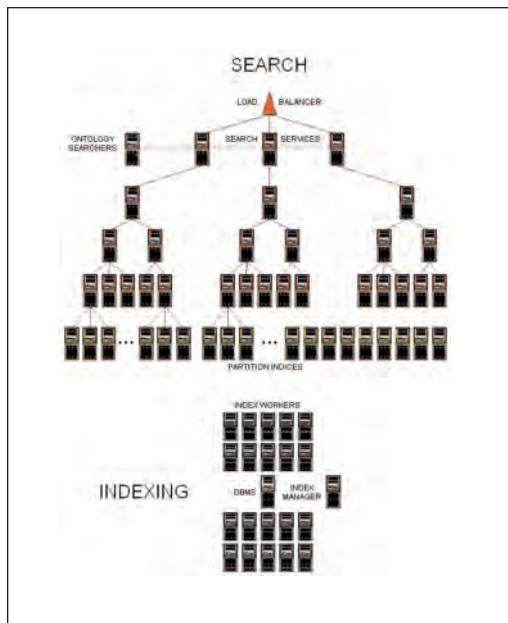
The company offers the following advice gleaned from its own SBIR experiences:

- Do market research and validation to ensure that there is a real, commercial outlet for the company's innovative technology, once all the technical hurdles have been overcome.
- Be able to deploy to a large community of end users and gather their feedback from the very beginning. Validation with the end user must drive all product development.
- Communicate changes in direction to the DARPA project manager promptly. Changes in direction often are motivated by lessons learned from end-user feedback. Changes can be straightforward to justify to the program manager, but still require advance notice.

Economic Impact

The funding received by Insightful under this DARPA SBIR has had a significant and positive impact on the company's research efforts. Besides being Insightful's initial key contract in this research area, the DARPA SBIR offset approximately 25 percent of the company's development costs during the period 2000-2002. The InFact® product line has contributed to the company's revenues—approximately \$900,000 in 2004, and \$1.5 million in 2005.

After the successful debut of InFact®, Insightful created a separate business unit for text analysis and search, and hired additional business development and sales personnel. The DARPA SBIR also led to three U.S. Patents: 6,510,406; 6,757,646; and 6,862,710.



The architecture of the InFact® search engine

About the Company

Insightful Corporation—based in Seattle, Washington—has 120 employees and annual revenues of \$22.3 million, representing an 18 percent increase over revenues for the previous year. Insightful also has offices in New York City, North Carolina, France, Switzerland, and the United Kingdom, with distributors around the world.

Insightful develops and delivers software and solutions for predictive analytics that have enabled thousands of companies to discern the patterns, trends, and relationships hidden in the data they collect. Insightful solutions are used by companies and organizations where analytics are critical to success, including financial services, pharmaceuticals, biotechnology, telecommunications, energy, and manufacturing, as well as research institutions and military and non-military government organizations. ■

Company Information

Insightful Corporation 1700 Westlake Avenue N., Suite 500 Seattle, WA 9810-3044 Phone: 206-283-8802 Fax: 206-283-8691 www.insightful.com	Jeff Coombs, President Founded: 1984 Annual revenue: \$22.3 million (2005) Revenue growth: 18% (2004–2005) Number of employees: 120
--	---