

TopicalNet, Inc (formerly Continuum Software, Inc.)

## Building a Database to Predict Customer Needs

*Since the early 1990s, organizations have used data warehouses and data-mining tools for business planning and decision-making. For example, more than 95 percent of the 250 companies surveyed by the Meta Group in 1995 said they planned to build a data warehouse, up from only 15 percent in 1994. Data warehouses and data-mining tools are able to provide a retrospective look at customer behavior through online analytical processing and other common statistical tools, but businesses want to forecast production changes and anticipate customer needs. Continuum Software, Inc., decided to construct a predictive modeling tool that the average business user could use easily. The technology was seen as too high risk, however, to attract private sources of capital, and so, in 1997, Continuum applied for and received an Advanced Technology Program (ATP) award. By mid-2000, Continuum's research in predictive models had led to the development of patented technology that classified tens of millions of web pages solely on the basis of predicted user interest. This technology also aided in the development of a classification tool that automatically classifies content from any text-based electronic document and assigns it to a topic to aid in the organization of electronic data. During 2001, Continuum (which by then had changed its name to TopicalNet) acquired several companies. TopicalNet's ATP-supported technology and its subsequent acquisitions have allowed the company to provide its customers with traffic verification, analysis, and research solutions for the World Wide Web.*

### COMPOSITE PERFORMANCE SCORE

(based on a four star rating)

\* \* \*

Research and data for Status Report 97-01-0087 were collected during April - June 2002.

### Traditional Data-Mining Tools Are Too Complex for Nonexperts

Many companies store information in data warehouses in an attempt to derive value from the massive amount of transactional data collected. Companies want answers to key business questions, such as: Who are our most valuable customers? How can we design and target promotions to increase sales? Which customers are most likely to leave in the future and how can these customers be retained? Corporations could get some answers to these questions by taking a retrospective look at already collected data. Relational databases, online analytical processing (OLAP), and statistical tools are used for historical data analysis, which assists users in answering questions such as: How much did sales increase in the eastern region over the past two

quarters? These tools have become invaluable because they have sophisticated graphical user interfaces that lead the person with the question directly to the answer, making historical information accessible to the average businessperson.

By 1996, several generic data-mining tools were available to predict future behavior, but, to be used effectively, they required scarce and expensive artificial intelligence and machine-learning expertise. Predictive models based on traditional data-mining tools pose two significant limitations: they require knowledge in machine-learning technology, a skill the average business user does not possess; and, for every query the business user poses, a specific model responsive to the query has to be built.

Building separate models in response to each query is time consuming and requires input from data-mining experts. Continuum recognized that the lengthy process of hiring experts and building specific models was too slow for this age of rapid data exchange. By the time a question has been asked and then simulated, the answer is often irrelevant, thereby thwarting the company's ability to gain a competitive advantage.

### **Continuum Intends to Predict Customer Behavior with Future Database**

Using machine-learning, statistical, and visualization techniques to discover and present knowledge in a form that could be accessed quickly and is easily comprehensible, however, required an innovative approach.

---

*Generic data-mining tools were available to predict future behavior, but they required scarce and expensive artificial intelligence and machine-learning expertise.*

---

Continuum planned to create a unique application to build simulators using data-driven, machine-learning technology. The simulators would be used to simulate what customer behavior would be like months into the future by extrapolating from the behavior of existing customers. That information would then be used to create a predictive database in the same format as a historical database.

The business user would be able to extract, display, and analyze information from this future database using the same familiar relational, OLAP, and statistical tools that are currently used for extracting information from historical databases. Continuum's simulator also would be able to respond to questions that involve dynamic situations, unlike conventional data-mining models, which are only capable of analyzing problems involving static situations. The future database could answer a question such as: What would happen if a company lowered its prices? To answer the question, a company would start its projection with the original customer base and would project results both with and without the hypothetical event of lowering prices. Then the

company could compare the two projected futures, identify the customers who would respond to the lower prices, and determine what total revenues and costs would be with and without the action.

Continuum planned to investigate extending current technology to handle huge amounts of customer data found in data warehouses to enable the construction of simulators for creating a future database of those customer records. The future database would contain future records projected from current records using projective visualization (PV), a new and unique application of machine learning that is able to analyze huge amounts of warehoused data, even when the records in the warehouse are incomplete. Unlike conventional machine-learning techniques, which only provide an answer to the specific question for which they are created, PV could be used to answer many questions, just like historical databases.

Prior to the ATP project, Continuum had already received the enthusiastic support of companies intrigued with the potential impact the future database could have on their businesses, but the nature of the research was still too high risk to attract funding. To demonstrate the applicability of the future database technology, Continuum planned to work with Switchboard.com to provide training sets of data and test domains for the application of the future database. Because Links2Go was the most promising test domain, Continuum decided to work with it rather than with Switchboard's. However, Continuum's relationship with Switchboard ended amicably.

### **Predictive Simulation Tools Promise Powerful, New Data-Analysis Capabilities**

When Continuum approached ATP for funding to continue its research and development in 1997, ATP recognized that the company's machine-learning techniques would advance the slow and costly practice of building simulation models to analyze customer behavior. Continuum could tap into the explosive use of the Internet, and the increasing volume of data being generated and warehoused in numerous web sites, to address new opportunities that were impossible with the data-analysis methods then available.

The market for data-analysis tools was growing rapidly, and the commercialization of Continuum's predictive data models also had potential for market spillover.

---

*Continuum organized and classified tens of millions of web pages solely on predicted user interest.*

---

### Continuum Validates Projective Visualization

The principal technical goal of this project was to develop a software tool that would enable the creation of databases that could predict future behavior. The following requirements had to be fulfilled to validate the use of PV:

1. *Increase the quantity of training data. Results:* The company made significant progress in its ability to load and analyze large training sets of data in a reasonable time. At the beginning of the project, Continuum was able to analyze 600,000 training instances on a machine with 1 gigabyte (GB) of random access memory (RAM). By the end of the project, it was able to analyze 7 million instances per GB of RAM and had built systems with as many as 40 million training instances (almost a hundredfold increase).
2. *Increase the quality of training data. Results:* The company developed numerous automated and semi-automated techniques for enriching data sources and for fusing additional data sources, allowing systems with greater accuracy to be built for fixed training sets.
3. *Identify the most effective machine-learning techniques for specific problem types. Results:* Continuum examined a large number of alternative machine-learning techniques and integrated the best ones into a unified tool set that can be applied to a wide variety of problems.
4. *Develop effective user interfaces for complex information. Results:* The company developed and deployed interfaces for visualization and

manipulation of the system during the course of development.

5. *Validate results on selected test domains. Results:* Continuum monitored software performance by applying the predictive technology to selected test domains to examine its applicability and to receive feedback on the system's functionality.

Continuum achieved these technical requirements and developed a software tool that uses PV to project the effects of an action on the basis of prior behavior. Not only was Continuum able to develop the necessary tools for a future database, it also organized and classified tens of millions of web pages solely on predicted user interest.

### Continuum Refocuses on Content Classification

In one of the initial test domains, referred to as Links2Go, the results of the company's work far exceeded any of the other test domains. Continuum realized that its technology also could be used to project user interest in sets of topics and pages in the context of an online research tool, which would allow the classification of related web data. Continuum was so impressed with the capabilities of the research tool that it set out to further develop and commercialize the technology for the classification of web content.

The Links2Go test site was well received because it had the ability to automatically organize web content and to provide end users with highly relevant links pertaining to the topic of their query. Typical search directories rely heavily on human editors; these directories have the ability to manually classify several million pages and are organized on the basis of the editor's subjectivity. In contrast, the Links2Go directory automatically organized 70 million web pages by topic, thus providing a significantly higher number of relevant retrievals. Furthermore, competing search engines typically refresh their web pages every 30 days; during this lag time, they could potentially send web users to expired uniform resource locators (URLs). The Links2Go directory was refreshed overnight, thus ensuring that users were viewing the most current versions of the web pages in the directory. With no marketing or public relations effort, the Links2Go web

site experienced more than one million distinct visitors per month.

Continuum was gaining a tremendous amount of knowledge about various topics through the organization of these web pages. Subsequently, the company was able to use the technology from this project to develop a classifier software tool, which could automatically classify arbitrary, unstructured documents.

---

*TopicalNet used the predictive technology developed as a result of the ATP project to provide businesses with a software solution to classify massive amounts of related, electronically stored data into easily accessible topics.*

---

The tool was able to read a document and assign it to a topic on the basis of knowledge gained from the topical directory. The tool also could build a taxonomy and decide where the document should fit within that categorization scheme.

### **Test Domain Leads to Compelling Business Opportunities**

By the midpoint of the project in 1999, Continuum executives decided that the commercial consequences of the technology emerging from the ATP project were so significant that they formed a new company. In August 1999, Continuum Software, Inc., became Links2Go.com, Inc. The new company decided to focus on the Links2Go web site after it closed a \$4 million investment deal with the venture capital firm, Bertelsmann, Inc. This investment financing enabled Links2Go to begin to aggressively commercialize the initial outcomes of its research. Links2Go not only provided end users with a powerful topical search and directory tool, it also offered businesses a variety of unique options such as targeted web advertising.

The Links2Go directory gave businesses the opportunity to advertise and receive competitive intelligence. Businesses were able to target their web advertisements to all links and keyword searches on topics related to their particular business areas. The businesses would receive substantial access to their specific markets by targeting web users with an interest

in topics pertaining to the companies' scope of business. Links2Go also offered statistics on a company's web site traffic and how the site compared with competitors' sites.

Links2Go approached several companies in the WWW market segment that could benefit from its technology. Links2Go proposed to allow these companies to use Links2Go's vertical technology for six months at no cost, with payment to begin after the trial period. After the six months was up, however, Links2Go had difficulty collecting payment for its tool. It continued to market its services, with limited success. When new Chief Executive Officer Ray Kingman joined the organization in October 2000, the company began to change its focus to further expand the classifier technology developed during the ATP-funded project. At this time, the company was renamed TopicalNet, Inc., to commemorate a new step in its path to commercialization.

TopicalNet used the predictive technology developed as a result of the ATP project to provide businesses with a software solution to classify massive amounts of related, electronically stored data into easily accessible topics. The technology can classify information from the Internet, corporate intranets, and extranets. Content classification is an ever-increasing need among businesses and end users; in fact, it is estimated that 80 percent of the content within an enterprise is unstructured. As information continues to be created and stored, content classification is essential in quickly obtaining accurate, relevant information. By 2004, the content classification and web analytics industry is predicted to grow to an estimated \$2.2 billion. TopicalNet's technology has the potential to significantly impact this new and expanding market.

### **Conclusion**

The ATP grant offered Continuum (later renamed Links2Go and then TopicalNet) the means to explore the possibilities of a predictive modeling tool, which led to the development and commercialization of the company's content classification technology. ATP's support permitted the company to examine the benefits of this technology without assuming the total risk. The project's successful completion led to additional funding and commercial viability in the content application and web analytics industry. TopicalNet's future is bright as it makes inroads in this growing market.

## PROJECT HIGHLIGHTS

### TopicalNet (formerly Continuum Software)

**Project Title:** Building a Database to Predict Customer Needs (Building a Future Database)

**Project:** To develop software tools that apply new techniques in machine learning and data mining to give business managers powerful, predictive data models represented as common business databases that can be queried with familiar database tools.

**Duration:** 10/01/1997-9/30/2000

**ATP Number:** 97-01-0087

#### Funding (in thousands):

ATP Final Cost	\$ 1,164	70%
Participant Final Cost	<u>500</u>	30%
Total	\$ 1,664	

**Accomplishments:** This project successfully developed the patented technology used to create projective software tools. The funding enabled TopicalNet, Inc., to create a search and directory service with technology that automatically created large topic directories from web page analysis and automatically extracted and classified text in documents on the Internet. TopicalNet's technology automatically and accurately classifies electronic documents more quickly and reliably than do human editors, and it increases the relevancy of the results from search inquiries.

The following patent was awarded as a result of this ATP project:

- o "World wide web link referral system and method for generating and providing related links for links identified in web pages" (No. 5,999,929: filed September 29, 1997, granted December 7, 1999)

**Commercialization Status:** Since completion of the ATP-funded project, TopicalNet has not proceeded with any further commercialization of the Links2Go directory; however, the company is aggressively pursuing enhancements to the classification technology. TopicalNet continues to make advancements in the web analytics market as the company positions itself to acquire other companies with capabilities that further enhance the classifier technology. During 2001, TopicalNet acquired the following companies: I/PRO, Inc. (technology used for site measurement, advanced customer analytics, and online audit services);

TeraLytics, Inc. (applications that analyze large volumes of customer information to help plan successful sales strategies); and Collectively Sharper, Inc. (experts in the content integration field). TopicalNet's ATP-supported technology, combined with its 2001 acquisitions, has allowed the company to provide customers with advanced software solutions to help them manage their own proprietary data as well as the data they may package and sell to customers. The company also provides customers with web site traffic verification, analysis, and research solutions.

TopicalNet customers include Fortune 100 companies and well-known Web-based businesses. Marketed as an automated, out-of-the-box solution, TopicalNet's classification solution can be up and running in about an hour. Its robust taxonomy of more than one million topics helps customers save considerable time, energy, and resources traditionally required to learn and use other classification methods and applications.

**Outlook:** The outlook for TopicalNet is strong. The company continues to grow within the content application and web analytics industry. The technology developed as a result of the ATP-supported project has enhanced TopicalNet's stature and helped it to attract venture capital opportunities. As the market for content application and web analytics grows, TopicalNet continues to prove it has the technological innovation and business confidence to successfully compete in this area.

**Composite Performance Score:** \* \* \*

**Number of Employees:** Five employees at project start, 70 as of June 2002.

#### Company:

TopicalNet, Inc.  
800 West Cummings Park, Suite 2900  
Woburn, MA 01801

**Contact:** John Mucci

**Phone:** (781) 932-8400