

TO: Marco Alameda, Vice-President of Venture Capital

FROM: Michael J. Sepcot, Research Assistant

DATE: October 3, 2002

CC: Frank Medeve, Director of Research

SUBJECT: Recommendation on Investment in Text/Concept Mining Technology

Data mining is the process of looking for patterns in databases. Text mining is a subset of data mining in which patterns are looked for in natural language text only. Text mining uses natural language processing to find useful information out of chaotic text databases. BeingMETA describes concept mining as combining “linguistic analysis with existing background knowledge bases to semi-automatically generate domain-specific knowledge bases and taxonomies which will aid in search, browsing, and business process automation.”

Text and concept mining is only going to grow in the future. Implementations of this technology are foreseen in every imaginable field. Concept mining will allow for customized browsing and searching interfaces incorporating the different structures, practices, products, and markets of organizations.

AvaQuest, Inc. outlines the real-world implications of text mining. “In life sciences, text mining is used to discover causal relationships between chemicals and symptoms, gene expressions and diseases, and other domain specific concepts. In finance, text mining is used to discover patterns and trends that relate financial and non-financial news events with market performance.” Text mining can also probe documents to determine what language they are written in or automatically categorize documents based on predetermined criteria.

Since text/concept mining is a subset of data mining, most of the companies in pursuit of this technology focus on data mining as a whole. A few of the key companies already pursuing data mining technology include, but not limited to, Angoss Software Corp.,

IBM Corp., Magnify, Inc., Oracle Corporation, and SPSS Inc. Angoss develops data mining applications including: KnowledgeSTUDIO, KnowledgeSEEKER, KnowledgeExceclerator, KnowledgeAccess, and add-ins for Microsoft Excel and Microsoft Access. IBM Corp. created the DB2 Intelligent Miner for Data to search transactions and e-commerce applications. Magnify, Inc. is a Chicago based firm who uses the technology to benefit the financial, insurance, and e-commerce industries. Oracle has developed Darwin, a data mining suite, to help corporations uncover patterns within their own data. SPSS Inc. focuses the technology on market and scientific research as well as quality improvement.

Illinois Institute of Technology has many courses available to begin the training of concept and text mining. Since text mining is a subset of data mining, knowledge in this field is also needed. IIT does provide courses in both text/concept mining as well as data mining. The following courses and descriptions are taken from the CS departments' web page. CS331 begins the basic sorting and searching algorithms with emphasis on data extraction. CS425 introduces database architectures. CS480 implements data-driven programs and techniques of searches. CS520 teaches database organization and design. CS522 focuses directly on data mining itself. CS525 teaches optimization techniques in databases. CS529 teaches information retrieval techniques. CS581 increases the artificial intelligence programming. CS584 teaches pattern recognition. CS585 introduces human languages which are vital in text mining.

Lucent Technologies should pursue concept and text mining through the venture of smaller businesses. We are already researching network data, statistical methods for data, data visualization, computing with data, statistics for very large sets of data, and manufacturing data which focus on data mining. Ventures into the subsets of data mining, including text/concept mining will only help our company grow. If we do not continue growth in data mining by venturing into text and concept mining, we will ultimately fall to our competitors.

Bibliography:

AvaQuest, Inc. www.avaquest.com. October 3, 2002. Information gathered from www.avaquest.com/services-mine-aq.html page.

BeingMETA. www.beingmeta.com. October 3, 2002. Information gathered from www.beingmeta.com/gnosis.fdxml page.

Illinois Institute of Technology. www.iit.edu. October 2, 2002. Information gathered from www.cs.iit.edu/ and its sub-pages.