

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 Data/Text 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. There are also other subsets of data mining, such as web mining, which focus on a specific set of data.

Data mining is only going to grow in the future. Implementations of this technology are foreseen in every imaginable field. Members in the auto, banking, medical and retail industries could all benefit from data mining. Data mining is starting to show up in personal computers as well, giving everyone the chance to benefit from the technology.

The information extracted from data mining allows companies to increase revenue and decrease cost. The information can also be used to detect fraud; analyze defects; and create a hiring profile. This technology can be used to determine the credit risk of individuals by analyzing loan records. It can also help out retailers by recognizing shopping patterns of individuals, types of goods purchased, and at what prices. Credit card companies could use data mining to help identify credit fraud, and thus protect themselves as well as their clients.

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, KnowledgeExcelerator, 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 data mining experts. 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.

Lucent Technologies should continue its pursuit of data 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. Continued venture into data mining will only help this company grow. If we do not continue growth in data mining, we will ultimately fall to our competitors.