J. Norwood Crout
Artificial Intelligence Corporation

The INTELLECT natural language database query system, a product of Artificial Intelligence Corporation, is the only commercially available system with true English query capability. Based on experience with INTELLECT in the areas of quality assurance and customer support, a number of issues in evaluating a natural language database query system, particularly the INTELLECT system, will be discussed.

A. I. Corporation offers licenses for customers to use the INTELLECT software on their computers, to access their databases. We now have a number of customer installations, plus reports from companies that are marketing INTELLECT under agreements with us, so that we can begin to discuss user reactions as possible criteria for evaluating our system.

INTELLECT's basic function is to translate typed English queries into retrieval commands for a database management system, then present the retrieved data, or answers based on it, to the terminal user. It is a general software tool, which can be easily applied to a wide variety of databases and user environments. For each database, a Lexicon, or dictionary, must be prepared. The Lexicon describes the words and phrases relevant to the data and how they relate to the data items. The system maintains a log of all queries, for analysis of its performance.

Artificial Intelligence Corporation was founded about five years ago, for the specific purpose of developing and marketing an English language database query product. INTELLECT was the creation of Dr. Larry Harris, who presently supervises its on-going development. The company has been successful in developing a marketable product and now looks forward to significant expansion of both its customer base and its product line. Versions of the product presently exist for interfacing with ADABAS, VSAM, Multics Relational Data Store, and A. I. Corporation's own Derived File Access Method. Additional interfaces, including one to Cullinane's Integrated Database Management System, are nearing completion.

A. I. Corporation's quality assurance program tests the ability of the system to perform all of its intended retrieval, processing, and data presentation functions. We also test its fluency: its ability to understand, retrieve, and process requests that are expressed in a wide variety of English phrasings. Part of this fluency testing consists of free-wheeling queries, but a major component of it is conducted in a formalized way: a number of phrases (between 20 and 50) are chosen, each of which represents either selection of records, specification of the data items or expressions to be retrieved, or the formatting and processing to be performed. A query generator program then selects different combinations of these phrases and, for each set of phrases, generates queries by arranging the phrases in different permutations, with and without connecting prepositions, conjunctions, and articles. The file of queries is then processed by the INTELLECT system in a batch mode, and the resulting transcript of queries and responses is scanned to look for instances of improper interpretation. Such a file of queries will contain, in addition to reasonable English sentences, both sentence fragments and unnatural phrasings. This kind of test is desirable, since users who are familiar with the system will frequently enter only those words and phrases that are necessary to express their needs, with little regard for English syntax, in order to minimize the number of key-strokes. The system in fact performs quite well with such terse queries, and users appreciate this capability. Query statistics from this kind of testing are not meaningful as a measure of system fluency since many of the queries were deliberately phrased in an un-English way.

In addition to our testing program, information on INTELLECT's performance comes from the experiences of our customers. Customer evaluations of its fluency are uniformly good; there is a lot of enthusiasm for this technical achievement and its usefulness. Statistics on several hundred queries from two customer sites are presented. They show a high rate of successful processing of queries. The main conclusion to be drawn from this is that the users are able to communicate effectively with INTELLECT in their environment.

INTELLECT's basic capability is data retrieval. Within the language domain defined by the retrieval semantics of the particular DBMS and the vocabulary of the particular database, INTELLECT's understanding is fluent.

INTELLECT's capabilities go beyond simple retrieval, however. It can refer back to previous queries, do arithmetic calculations with numeric fields, calculate basic functions such as maximum and total, sort and break down records in categories, and vary its output format. Through this augmentation of its retrieval capability, INTELLECT has become more useful in a business environment, but the expanded language domain is not so easily characterized, or described, to naive users.

A big advantage of English language query systems is the absence of training as a requirement for its use; this permits people to access data who are unwilling or unable to learn how to use a structured query system. All that is required is that a person know enough about the data to be able to pose a meaningful question and be able to type on a terminal keyboard. INTELLECT is a very attractive system for such casual or technically unsophisticated users. Such people, however, often do not have a clear concept of the data model being used and cannot distinguish between the data retrieval, summarization, or categorization of retrieved data which INTELLECT can do, and more complex processing. They may ask for things that are outside the system's functional capabilities and, hence, its domain of language comprehension.

In summary, we feel that INTELLECT has effectively solved the man-machine communication problem for database retrieval, within its realm of applicability. We are now addressing the question of what business environments are best served by English-language database retrieval while at the same time continuing our development by significantly expanding INTELLECT's semantic, and hence its linguistic, domain.