MEDIATING THE VIEWS OF DATABASES AND DATABASE USERS

Gary G. Hendrix
SRI International
333 Ravenswood Avenue
Menlo Park, Ca. 94025

The Natural-Language/Deduction group at SRI International has undertaken several large projects integrating knowledge representation, the modeling and use of distributed conventional databases, logical deduction, and natural-language processing.* One of the largest projects, LADDER [2], involved accessing data distributed over a computer network by using queries expressed in English. Work with LADDER (and several similar systems) has revealed that:

(1) Users wish to talk about data in terms of the enterprise in which the data are to be used. Users do not confine their questions to concepts and terminology covered by the database per se.

(2) Users are seldom satisfied with access only to the data in a database. They need to know the KIND of data available (i.e., they want to ask questions about the DB schema), and they expect systems to include information that can be computed from "common knowledge" and information stored explicitly in the database (e.g., if a database records where two ships are, users expect the system to know the distance between them).

(3) Users are not satisfied with access to an existing database. They want to tell the system new facts. Some of these are not suitable for storage in conventional databases (e.g., statements involving quantification), and some involve counter factuals (e.g., "Suppose the ship were 100 miles south of its current location...").

(4) Given natural-language access to a DBMS, users expect to interact in natural language with other types of software, too. Moreover, they expect the various underlying software packages to understand one another's results (e.g., User: "Who is the commander of the ship?" System: "Admiral Brown." User: "Send him a copy of Smith's memo." The mailer is expected to understand the output from the database).

These problems are manifestations of the generic problem of modeling the resources available in software packages, modeling the concepts and terminology of the domains to which users wish to apply the resources, and modeling the relationships between the two.

To explore these problems, we are currently undertaking a major new project called KLAUS. The core concept of KLAUS is that of an interactive system preprogrammed with essential skills for learning the concepts and vocabulary of new domains, and with expertise for applying acquired knowledge in fact-retrieval and problem-solving situations. KLAUS is tutored about new domains in English. While being taught, KLAUS actively looks for gaps and inconsistencies in its knowledge, asking its tutor pointed clarification questions. In this manner, KLAUS aids the user in formalizing, organizing, and clarifying his ideas. Subsequent to tutoring, KLAUS can aid its tutor and others in performing tasks that require combining knowledge of the new domain with knowledge of how to use various computer systems, most notably DBMSs. In essence, a tutored KLAUS acts as a broker between the user's needs, as expressed in the user's terms, and a computer system's resources.

KLAUS is being built on a base of knowledge acquisition, knowledge representation, deduction, and natural-language abilities. (A pilot KLAUS system, called NANOKLAUS, has already been constructed and is described in [1].) Because the purpose of conventional DBMSs and KLAUS is to file data and to sort, selectively recall, and display data in various formats, KLAUS may be viewed as a sophisticated DBMS. But KLAUS supplements its conventional DBMS with mechanisms and representation schemas that provide the full representational power of first-order logic, and therefore much greater power than conventional DBMSs, which are limited to manipulating collections of ground literals.

---

* This research was supported by the Defense Advance Research Projects Agency under contract N00039-79-C-0118 with the Naval Electronic Sy stems Command. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advance Research Projects Agency of the United States Government.

Natural language plays a key role in KLAUS, being used both for data query and for acquiring new facts and concepts. Because KLAUS acquires information (including linguistic information) through interactions in natural language, it is well positioned to acquire the terminology and perspective on a domain that users are likely to bring to a problem. KLAUS is intended to acquire not only the data that would go into a conventional database, but also a wider (but shallower) body of general knowledge about the broader enterprise in which the data are to be used.

An interesting possibility is to tell KLAUS how information about a domain is recorded in an existing database. By telling KLAUS how the database is interconnected logically with various aspects of the real world that have been communicated to KLAUS previously, KLAUS may relate the utility of the database to users who know only about the real application and who do not understand the underlying DBMS technology and encoding schemes and conventions. (Although space prevents elaboration, we note that relationships between user concepts and relevant database structures may be very complex logically.)

Key questions for ongoing research include how to:

(1) Acquire a user's perspective on the enterprise in which data are used (e.g., What seed concepts can be used to initiate acquisition of new concepts?).

(2) Formulate the relationship between user concepts and computer-based resources.

(3) Interrelate the sort types of natural language (e.g., natural-kind terms) to the types used by formal systems.

(4) Understand and respond intelligently to queries and requests that cannot be processed successfully by the underlying software packages due to gaps in their knowledge and coverage.

REFERENCES

1. Haas, N. and Hendrix, G. G. "An Approach to Acquiring and Applying Knowledge." Proceedings of the First Annual National Conference on Artificial Intelligence. Stanford, California, August 18-21, 1980, pages 235-239.

2. Hendrix, G. G., Sacerdoti, E. D., Sagalowicz, D., and Slocum, J. "Developing a Natural Language Interface to Complex Data." Transactions on Database Systems, Vol. 3, No. 2, June 1978, pages 105-147.

3. Epstein, M. N., and Walker, D. E. "Natural Language Access to a Melanoma Data Base." Proceedings of the Second Annual Symposium on Computer Applications in Medical Care. Washington, D. C., 5-8 November 1978.