

However, in many casts the query will not map directly onto the database.

Instead of simply telling the user that his question is not "the question to ask", we will try to transform his request into something meaningful. In the particular database and semantically close to the original request - this part of the system is based on G.P. Zarri's RESEDA system (Zarn. 1983, 1984a, 1984b, 1984c). We define "question1" to be semantically close to "question2" iff the answer to "question1" is implied by the answer to "question2".

III. THE KNOWLEDGE BASE

To achieve this, we need of course the knowledge of what is "semantically close".

In the example presented above the system was able to transform the original question, because it knew that "if a person is born in a particular place then that person must have been in that place at least for a short period of time".

The second alternative of the transformed query is produced because the system knows that "in order to obtain a degree, a person has to attend courses at a particular university or do some research work at that university" and that "in both cases the person had to be physically present there at one time or another" (to simplify the presented example we deliberately ignore all the cases of exception to this rule).

It follows from the example that the system manipulates rather general common sense knowledge.

We decided to take the domain of staff management as our test-bed. However we still need to represent some more general, extra-domain knowledge.

From a technical point of view, our overall knowledge base is divided into a number of smaller knowledge bases, as illustrated in Figure 2.

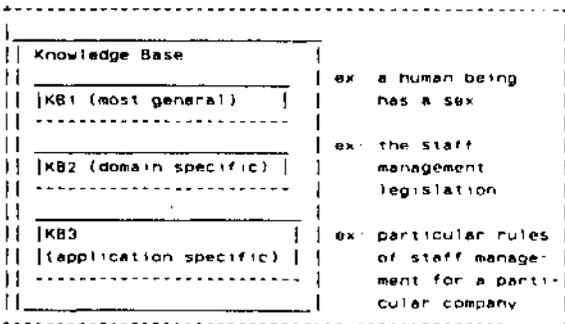


Figure 2

KB1 contains bits of knowledge that are usable in any application (of course, it will never be complete).

KB2 contains general knowledge for a particular domain. We hope to develop a certain number of KB2s (with a common KB1 nucleus) for each real world domain that our system will have to deal with.

Each KB1 ♦ KB2 couple is transportable "as is", as long as we do not switch to a completely different domain.

Note that KB3 is not specific of a particular database, the system can use the same KB3 in order to access different personnel databases of the same company.

The model of the database itself (the most specific level of knowledge) is not included here since it does not take part directly in the transformation process.

IV. TYPES OF KNOWLEDGE

Any one of the KBs mentioned above may contain two types of knowledge: factual knowledge and rules (represented in a formalism based on RESEDA).

A. Factual knowledge

A fact such that "a university is in a city...", which is a piece of factual knowledge, is represented by the highlighted part of Figure 3.

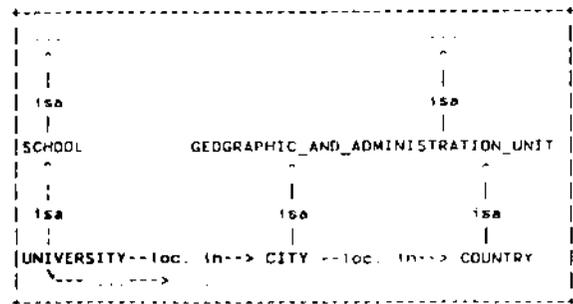


Figure 3

This frame-like network describes "specific-generic" type of relations ("isa" links, we allow polyhierarchies) used as matching criteria when applying rules (cf below). Other specific relations (used in the rules, for example topological ones) are also present in the network.

This representation does not have the full power of frame systems since presently there is no procedural knowledge represented in the network.

B. Rules

Currently we have two types of rules:

1. Transformation rules
2. Standardization rules

A fact such as "if a person (x) has received ("BE-AFFECTED-BY") a degree in a particular place (y) then that person has been ("BE-PRESENT") in that place", which is a transformation rule, is represented as in Figure 4.

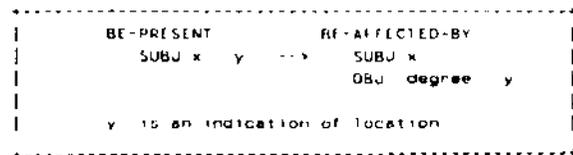


Figure 4

Note that:

- a. This is a deep case representation with a limited number of predicates ("deep verbs").
- b. The direction of the arrow is following RESEDA's convention. From a practical point of view the arrow indicates that the right-hand side schema is allowed to replace something that matches the left-hand side schema. Conceptually it means that the information possibly retrieved using the right-hand side schema implies the information searched for by the (original) left hand side schema.

Rules can place restrictions on variables (as is the case in RESEDA). For instance, the rule shown above could have required that "x" be a human being. This rule would then apply for "x ■ y", where "y" is known to be a teacher. The match (between x and y) is made possible, because we have stored in the factual knowledge representation the fact that "teacher '1s_a' human_be1ng".

Standardization rules are used to translate the parser's output (containing surface verbs) into a more canonical form, i.e. usable by transformation rules. All of the standardization rules belong to the KB1 level, as they are absolutely general.

V. THE INFERENCE ENGINE

We are in the course of developing (December 84) a prototype version with an "exhaustive" engine, meaning that all the possible transformations are executed (with simple destructive chronological backtracking) until a representation is found. This approach is very close to the RESEDA inference engine (see Zarri, 1984c, for a description of the later one).

This simplistic approach is justified since our current knowledge base contains only "KB3" types of knowledge.** We are planning to move onto more sophisticated approaches (ex. choosing among several representations in the database of the same question, a.s.o...) after the prototype has been well tested.

VI. CONCLUSIONS

The presented NLI system is designed to tackle issues such as transportability and "helpful understanding" of a naive user.

Transportability, the first issue, has been examined in many research systems, however all of these seem to lack a clear distinction between different levels of transportability of the knowledge represented (HAMANS, Hoepfner et al., 1983 seems to be an exception).

In addition SAPHIR+RESEDA is using general common sense knowledge (reference Hemphill and Rhyne, 1978, describes a project where the use of Schank's general formalism in DB queries has been explored, but its goals seem to be more limited than ours).

The second issue is somewhat close to "cooperation" as defined in the COOP system (Kaplan, 1982) Kaplan

** Our rules (temporarily) have a very specific content, i.e. they have not been "factored out". This is due to the fact that only a limited number of queries have been tested on a single database (we are beginning the implementation phase, following a one year study period).

however seems to consider only the case where the DB produces a null answer. We believe his and our approach to be complementary.

Our system uses rules to produce a DB query, this is similar to a deductive DBMS approach. However our rules are not formal and we never access the database in the course of transformation (because in most real world cases it is prohibitively expensive).

REFERENCES

- <1> Gross, M. Methodes en syntaxe. Paris. Hermann. 1975.
- <2> Hoepfner, W., T. Christaller, H. Marburger, K. Morik, B. Nebel, M. O'Leary and W. Wahlster. "Beyond Domain-Independence". In Proc. of IJCAI-83, Karlsruhe. W. Germany, 1983.
- <3> Hemphill L. and J. Rhyne, A Model for Knowledge Representation in Natural Language Query Systems, Research Report Laboratory RJ 2304(31046), IBM Research Laboratory, San Jose. September 1978.
- <4> Kaplan, S. J. "Cooperative Responses from a Portable Natural Language Query System" Artificial Intelligence, vol 19 (1982) 165-187.
- <5> Normier, B. SAPHIR - Presentation generale, ERLI. Janvier 1984.
- <6> Normier, B., M.E. Aubert, G. Clemencin, B. Euzenat, S. Lapep, N. Lellouche, A. Ogonowski and J. Vega, Extension de la gestion du dialogue, Rapport final (Marche DRET 83/234). ERLI. Decembre 1984.
- <7> Euzenat, B., B. Normier, A. Ogonowski and G. P. Zarri, Les mecanismes d'inference dans l'interrogation de bases de donnees relationnelles en langage naturel, Convention ADI-84-068 rapport intermediaire N.1. ERLI. Juin 1984a
- <8> Euzenat, B., B. Normier, A. Ogonowski and G. P. Zarri, Les mecanismes d'inference dans l'interrogation de bases de donnees relationnelles en langage naturel, Convention ADI-84-068 rapport final. ERLI. Decembre 1984b
- <9> Zarri, G. P. "An Outline of the Representation and Use of Temporal Data in the RESEDA System" Information Technology/ Research and Development, vol 2 (1983) 89-108
- <10> Zarri, G. P. "Expert Systems and Information Retrieval: an Experiment in the Domain of Biographical Data Management", in Developments in Expert Systems, Courts M.J ed. London. Academic Press. 1984a
- <11> Zarri, G. P. "Constructing and Utilizing a Large Fact Database Using AI Techniques". in Proc. of the First International Workshop on Expert Database Systems, Univ. of South Carolina. College of Business Administration. Kerschberg L. ed. 1984b.
- <12> Zarri, G. P. "Intelligent Information Retrieval on Interesting Application Area for the New Generation Computer Systems", in Proc. of the International Conference on Fifth Generation Computer Systems, Tokyo. Japan. 1984c.