ON VERBOSITY LEVELS IN COGNITIVE PROBLEM SOLVERS

P. Jirků and T. Havránek

Center of Biomathematics, Czechoslovak Academy of Sciences, 142 20 Prague 4, Vídeňská 1083, Czechoslovakia

The aim of the paper is to discuss several issues that usually occur when computational linguistics comes into interactions with so rapidly growing areas of artificial intelligence as it can be seen e.g. in designing expert and consulting systems or in the area of automated programming of knowledge--based problem solving systems. We will mention here problems of communicating knowledge between machine and researcher (a user of a system) which is not an expert in programming technique. Since natural language is a "natural" form for expressing knowledge (and most extramathematical knowledge actually exists in this form) it could be seemed that natural language would be also the best support for communicating knowledge through a cognitive process performed on a computer. Is it really so? We want to bring arguments for a rather opposite claim by pointed out several formats for expressing and communicating scientific knowledge which differs from usual natural language ones.

What kinds of data structures for expressing knowledge and for representing it in a computer memory we need? For answering this question we have to distinguish at least two roles of a user in the process of communication with machine. First of them can be called a speaker. It characterizes the situation in which the user loads knowledge into a machine. For such a situation there is important that knowledge entering a computer should be precise and exact as much as possible. So the data structures should enable disambiguation of information. Such a requirement leads naturally to hard constraints of input formats. On the other hand, not only for the user convenience, it should be desirable to allow highly free format of input information to avoid any apriori limitations. Thus, the system should be equipped by an effective interpreter transforming various kinds of input information into internal representation suitable for potential reasoning processes. The second role of the user in communication with computer can be called <u>a listener.</u> This situation is rather different from the first one. Here the user has to understand results of computations and (especially in consulting systems) also various explanations resulting from reasoning processes. These two roles can, of course, infiltrate one into another. Analogous roles can be recognized for a machine, too.

As a testbed for our considerations an AI problem-oriented designed for an <u>automatic data analysis</u> (called GUHA-80) has been chosen. The task of GUHA-80 system is to generate and develop interesting views onto given empirical data (recognize interesting logical patterns). These views should represent relevant information contained in the data and be useful for formation of hypotheses.

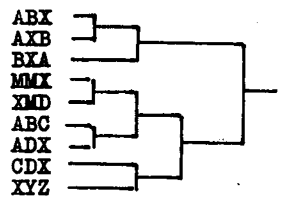
From the point of view of two above mentioned roles of user the following types of information in the GUHA-80 system could be distinguished:

1. Information coming into system could contain a) data (observations on objects), b) supporting knowledge (apriori knowledge about the problem in question, answers of questions).

2. Information coming from the system could contain a) trace of computations and reasoning activities, b) results of computations, c) explanations (why such and such operations have been performed), cf. MYCIN.

Let us give a hypothetical example of user´s communication with GUHA-80 system:

GUHA-80: by a single linkage method using euclidian distance on the set of objects the following dendrogram expressing the similarity between these objects was obtained:



Is the dendogram in accordance with your knowledge?

USER: No.

GUHA-80: May I suggest another pattern using different clustering techniques?

USER: Yes.

GUHA-80: Do you prefer some of the following techniques: (<u>a table</u> of relevant techniques follows)

It can be easily seen that in such a conversation different levels of understanding language are needed. NL level will be appropriate mainly for user´s answers, simple questions etc. But try to express the information contained in the dendrogram in NL form! Moreover, for communication process from GUHA-80 to the user it will be <u>typical a graphic representation</u> of information (which in many cases is more transparent than NL one).

Thus the language understanding take place mainly in the case sub 1b) i.e. when entering supporting knowledge. But for practical reasons it can be performed in a very simple level as e.g. in very high level programming languages. Example: INPUT FORMAT IS ´( )´. VARIABLES ARE 25. CASES ARE 102. VARIABLE NAMES ARE ... . PRINT CORRELATIONS.

In each case, such an understanding must lead to

1. to the elimination of redundant information and in such a way to the core of a statement;

2. to the possibility to work only with minimal cores of statements.

The reason for a second requirement is that a user experienced with the system tends to replace syntactic sugar by an appropriate slang to minimize his effort paid e.g. to punching or typing statements.

Conclusion. We have distinguished different types of communication of scientific knowledge through a mechanized cognitive process. It leads first to the claim that not only different levels of understanding language but also different levels of verbosity are needed. Moreover, in some cases the use of verbal information can be undesirable or even impossible. Moreover, in many cases when understanding language is needed it would be enough to understand only a small relevant fragment of it.

References:

Hájek P., Havránek T., GUHA-80 An Application of Artificial Intelligence to Data Analysis, in Počítače a umelá inteligencia, 1982, Vol. 1, No.2

Chytil M., Mathematical Methods as Cognitive Problem Solvers, in Kybernetes 1980, Vol. 9, 197-205

Jirků P., Logical and Linguistic Aspects of Computer-Based Inference Processes, in The Prague Bull. of Math. Linguistics 35, 1981, 41-53

Skuce D.R., Toward Communicating Qualitative Knowledge between Scientists and Machines, PhD Thesis, Dept. of EE, McGill Univ., Montreal, May 1977, 269 p.

Swartout W.R., Producing Explanations and Justifications of Expert Consulting Programs, MIT/LCS/TR-251, Massachusetts, Jan. 1981, 117 p.