Addressing the Retrieval Problem in Large Knowledge Bases

Ipke Wachsmuth and Josef Meyer-Fujara

Universität Bielefeld
Technische Fakultät
P. O. Box 8640
D-4800 Bielefeld 1

COMPUTATIONAL INTELLIGENCE 90
Research in Progress Session
Milano, Italy, 26 September 1990

INTRODUCTION AND OVERVIEW

Research currently in progress at the University of Bielefeld is concerned with conceptualizing a methodological framework which allows large-scale knowledge bases to be constructed, utilized and maintained in a modular fashion. Challenges to be met span from the modular conception of domain knowledge in a given application area on the one hand to the adaption of principles for modular system development to knowledge-based systems on the other. Important aspects of this work are to make expert system knowledge bases more easily comprehensible and maintainable, in particular, when several developers are involved or when long-term maintenance of a knowledge base is planned to be carried out by prospective users.

Current work in the field of expert systems has heavily addressed the notion of task-specific problem-solving methods and architectures, e.g., generic tasks (Chandrasekaran, 1987), and of high-level specifications of knowledge-based systems, e.g., KADS (Wielinga & Schreiber, 1990); see also (Steels, 1990). However, little attention has been given so far to the problem of retrieving relevant facts in large-scale domain-specific knowledge bases. With knowledge bases accumulating thousands and perhaps ten-thousands of elements of domain knowledge at the horizon, this problem can hardly be ignored for it involves the need of constraining search in possibly exponentially complex search spaces. Moreover, very large knowledge bases can only be compiled by teams. Therefore a structure must be found that allows individual developers to work on parts of the base independently from one another to the extent possible *(modularization)*.

As one of the particular issues involved in this work, this paper addresses the problem how large, and possibly heterogeneous, domain knowledge bases should be organized so as to allow for a context-driven access to task-relevant domain knowledge. That is, it deals with the problem that in a large knowledge base the particular piece of domain knowledge needed for solving a certain problem or subproblem may actually be very hard to find. A core question, then, is whether memory search can be guided by contextual clues as can be observed with human experts. Three issues that are particularly relevant and place special difficulty on system development are discussed in detail: *Largeness* (size), *granularity*, and *diversity* of domain knowledge bases.

To cope with the problem of largeness of domain knowledge bases, the approach taken at Bielefeld is to identify packets of knowledge that are self-contained in the sense that they can account for special cases in the general application area and can be left unaccessed when other packets are found relevant. A set of nine principles have been formulated that serve to guide organization of and retrieval in structured knowledge bases (Wachsmuth, 1987). We emphasize that these principles have been derived from empirical studies of human knowledge organization (Wachsmuth, 1989). The general aim of these principles is to serve as a knowledge-level specification for a structured knowledge base. The principles of domain-oriented knowledge structuring are presented and motivated in the second section.

Granularity is an aspect of particular importance since modelling at various levels of detail is necessary for problem solving in complex domains. While human problem solvers readily identify the adequate level of precision and resolve discrepancies among domain models of different granularity, principal questions have yet to be solved with respect to problem-solving systems. The core problem to be attacked is how to provide a system with views of a knowledge base of task-adequate granularity and what methods are necessary to mediate a smooth change among views. A further question discussed is how granularity can aid in a semantic structuring of knowledge and in retrieving relevant parts of it. The different aspects of granularity are addressed in the third section.

Beyond largeness and granularity, the diversity of domain-specific knowledge held in wide-focused future knowledge-based systems needs attention. A human expert's ability includes rapid access of highly specific knowledge in diverse subdomains without actually attempting to apply irrelevant facts. For an artificial system to exhibit such a selective retrieval, it needs to be capable of locating elements of its knowledge with respect to their relevance for a given task. The topic of retrieval in large knowledge bases is addressed in the fourth section. New techniques emerging from neural networks are discussed as to how they may help to organize symbolically represented semantic knowledge and guide memory search especially when diverse knowledge is involved.

The issues discussed in this paper are illustrated by examples drawn from medicine which is a chosen application area for the expert systems group at the University of Bielefeld. In preparation for research aimed at building a medical knowledge base concerned with the genesis of chronic diseases, the field of blood hypertension was inspected in more detail. An expert system for consultation about blood hypertension requires diverse knowledge about heart, kidneys, arteries, blood pressure regulation processes, normal levels of potassium and calcium, heredity, effects of pregnancy, nutrition and medication. In contrast to acute illness, reaching a diagnosis for a chronic, multi-facetted disease is a subtle recognition process that involves a vast body of domain-specific heuristic and principled knowledge (Heller, Kauffmann, & Wetter, 1989b).

It would be inefficient and unnatural to search and match all knowledge potentially relevant as a uniform unstructured heap in order to answer an arbitrary question. For example, there is no point in considering pregnancy-related factors of hypertension for patients that are not pregnant. In preceding discussions with a research group at IBM Scientific Center, Heidelberg, their experiences with constructing a prototype hypertension consultation system were evaluated (Heller, Kauffmann, & Wetter, 1989a). These evaluations served as a basis for concretizing ideas of modular system development and for taking steps toward improving the design and extending the scope of domain knowledge bases in both extent and depth. At the onset of formulating and proposing concrete research goals, the particular intent of this paper is to address these ideas to a larger audience, and to receive their early critiques.

PRINCIPLES OF DOMAIN-ORIENTED KNOWLEDGE STRUCTURING

In this section, the principles adopted for guiding the organizational structuring of large knowledge bases are briefly discussed. There are principles reflecting *static* aspects, i.e. aspects guiding the definition of so-called knowledge packets and the restrictions of their access. Other principles concern *dynamic* aspects of changing access conditions and of guiding retrieval. By the term "knowledge elements" we refer to any fact, rule, or structured object that expresses a domain-related piece of knowledge. We implicitly assume a domain-independent strategy that controls interpretation and use of domain knowledge. Some knowledge elements may be included to control the invocation of elementary problem-solving actions. Three principles describing the *organization of knowledge* are presented subsequently followed by a brief motivation.

Packing knowledge elements. Collections of knowledge elements that pertain to a specific domain of knowledge are comprised in a *packet*. We say the packet *owns* these knowledge elements. A packet may properly contain further packets of knowledge elements that constitute identifiable subbodies of more specific knowledge within the outer packet.

Competitive knowledge. Collections of knowledge elements that concern alternative methods or views in a given domain of knowledge are packed separately within the surrounding packet. Such packets are referred to as *competitive*.

Local consistency. The collection of knowledge elements in one packet must not permit conclusions that are contradictory (or actions that are incompatible). A packet P may only contain contradictory (or incompatible) knowledge elements if they are packed separately within P. A collection of knowledge elements satisfying this principle is called *locally consistent.*

We give a brief motivation of these notions. The knowledge concerning the criteria for hypertension such as thresholds of diastolic and systolic pressure may form a packet P. Inside this packet, the knowledge about rules how many different measurements at which spots of the body and at which intervals must be taken in order to determine the pressure, may form a further packet Q strictly contained in P.

Next, competitive knowledge: As in many areas of human expertise, different views about cause-effect relationships exist among medical experts, and this is also true for hypertension. Until recently it was assumed that its origin is with the heart and that kidney problems are consequences. Hypertension is seen now by some researchers as being originally caused by kidney problems which later lead to heart problems (Rettig, 1989; de Wardener, 1990). Knowledge reflecting these alternative views and their respective implications may be understood as competitive; the corresponding knowledge elements should thus be packed separately.

Local consistency: A situation may arise where the two aforementioned views lead to incompatible therapy suggestions. With an unstructured, global knowledge base this could entail absurd if not harmful prescription mixing. Packing the knowledge corresponding to the different views into separate packets, together with the principles of access described shortly, allows the prevention of such mishaps. Nevertheless, both theories could be represented and be called upon when adequate. The question has been raised if for broad knowledge bases the requirement of global consistency can be weakened such that consistency checks are restricted to portions of a knowledge base that potentially interact during a problem-solving process (Lenat & Feigenbaum, 1987). The principle of local consistency was introduced to account for this issue; it requires both theories to be internally consistent but permits conflicting theories to be integrated in the larger system.

On the development side, these organizational principles aid in subdividing work among different team members as each developer's contribution can be checked independently of others or at least with less cross-consulting. Cross-checking may still be necessary, though, when more general knowl-edge is represented in surrounding packets which are still requested to be consistent with any of their subpackets.

Principles describing *static access conditions* are presented next.

Eligibility of knowledge elements. The knowledge elements owned by a packet P are conjointly eligible for use by the knowledge-based system when their packet P, or a packet within P, is tagged ACCESSED, but only as far as they are not also owned by a packet contained within the one tagged ACCESSED. We say a knowledge element (or a set of knowledge elements) eligible for use is VISIBLE. All knowledge elements packed separately from the packet tagged ACCESSED are not eligible.

Single access to packed knowledge. Only one packet at any given time may be tagged ACCESSED.

Reachability of knowledge. When a knowledge packet P tagged ACCESSED owns knowledge elements that are also owned by a packet Q within P, then the set of knowledge elements in Q (or likewise, the packet Q) is REACHABLE. A collection of knowledge elements packed separately from the one tagged ACCESSED is NOT REACHABLE.

Restricting eligibility pertains to the above problem of preventing the search-and-match process from becoming inefficient since it restricts the problem solver's interpretative processes to VISIBLE knowledge elements. Given the packing criterion of specificity, this amounts to considering the more general knowledge first and not to inspect too specific knowledge without necessity. The principle of single access has packets of specific knowledge considered one at a time. In tagging a packet ACCESSED, more specific knowledge becomes REACHABLE although not immediately VISIBLE.

In the domain of hypertension, the situation might be as follows: As long as a consultation is concerned with anamnesis, a packet "anamnesis knowledge" is ACCESSED, and its more general contents (say the fact that heredity factors play a role in hypertension) are VISIBLE. Other knowledge, for instance, about kidney morphology and normal electrolyte concentration, can be ignored by the interpretative procedure if it is packed separately from "anamnesis knowledge." Very specific knowledge about the hereditary character of particular forms of hypertension can be ignored as well if it is further packed as a subpacket of "anamnesis knowledge". However, in contrast to the kidney morphology knowledge, it is REACHABLE and may become VISIBLE when the topic of heredity is triggered, e.g., when the patient mentions her mother having suffered from hypertension during pregnancy. See Figure 1 for an illustration.

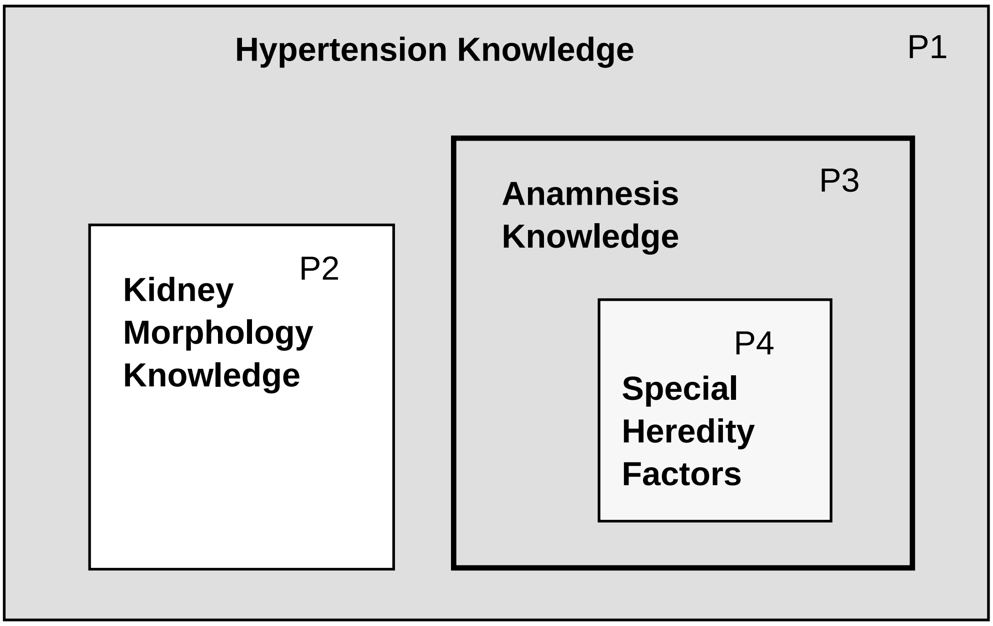

Figure 1. Illustration of static aspects of KB structuring principles. Anamnesis knowledge packet (tagged ACCESSED) and general Hypertension knowledge VISIBLE, and knowledge about special heredity factors REACHABLE. Kidney morphology knowledge neither VISIBLE nor REACHABLE in the access condition shown.

For the sake of completeness, the principles describing *dynamic access conditions* are given here. As they are particularly related to the issue of retrieval, discussion of these principles is postponed to the fourth section. We just note here that by changing access conditions a dynamic partitioning of the total knowledge base in VISIBLE and invisible knowledge results.

Structure-dependent access to knowledge. When dealing with a task situation on the basis of the knowledge currently VISIBLE turns out unsuccessful, the ACCESSED tag is moved to one of those knowledge packets REACHABLE next.

Keyword-dependent access to knowledge. A means to tag a packet of domain-specific knowledge ACCESSED is the finding of certain concept words (or combinations of concept words) directly associated with knowledge elements in this packet. We refer to such words as *keywords*.

Persistence of access conditions. Upon completion of a goal, the current ACCESSED tag persists as a start-off condition for the partitioning state of the knowledge base when the next goal is issued.

ASPECTS OF GRANULARITY

For many tasks, domain models of various levels of detail are necessary at different stages of problem solving. As is further substantiated below, the following aspects of detail appear to be important: *resolution*, *aperture*, *depth*, *stepsize*, *completeness*, *precision* and *correctness*. We call them aspects of *granularity*. For each aspect, different sorts of knowledge are relevant, and it may happen that even the representation of such knowledge needs to be adjusted to be adequate for efficient reasoning.

Resolution. In order to describe the mechanisms involved in blood pressure control, different time scales ranging from seconds to days, or even months, have to be considered. Depending on the time scale actually used, some mechanisms will appear to be instantaneous events (e.g., baroreceptor control), some to be extended processes causing gradual changes (e.g., capillary control), and still others to be stationary (e.g., aldosteron controlled sodium balance), cf. (Guyton, 1981) discussed in (Kuipers, 1987). As Kuipers (1987) observes, an expert system must be able to treat a relation among shared variables of a fast process as a functional relationship when reasoning about a slow process, and to treat an even slower one as being constant. Change of time detail (or *temporal resolution*) will thus lead to change of detail in object. Some objects may be ignored, some may change their representation type (event instead of process) or even be modeled by entirely different entities (functional relationship instead of process effect).

An example with respect to *spatial resolution:* at the microscopic scale, cell forms and tissue textures are relevant whereas an organ's outer form and its position in the body are not. This relation may be used in the other direction as well: when a physician describes cell forms to the system, the system may adjust the scale correspondingly and thus rule out considerable amounts of knowledge as irrelevant for further inferencing.

Aperture is an aspect of granularity which often is intimately related to resolution. By this we understand the size of the "visual field" of the system, i.e. of the region, time span or object class taken as domain of validity of propositions. The following example illustrates its role in retrieving relevant knowledge: Considering the total population of the world, a different spectrum of diseases must be highlighted as frequent and be made accessible than when reasoning about the population of Europe.

Depth in reasoning may differ in stages of problem solving as well as in explanation. Obviously, most often heuristic knowledge will suffice to solve a problem and should therefore be accessed first. Only when it turns out to be insufficient should deeper, e.g., causal knowledge be looked for.

Stepsize, completeness, precision, and correctness are not discussed in this paper since they mainly relate to issues other than retrieval. We note that examples for the role of these aspects of granularity in problem solving are easily found.

Granularity seems to us an important principle of structuring knowledge and of guiding access to the knowledge. Human problem solvers readily cope with the difficulties of different granularity levels, but cannot yet be mastered by artificial systems. A number of aspects as named above must be taken into account by systems for use in large-scale domains if they are to be sufficiently flexible and efficient. Incorporating them necessitates different models of the world depending on the level of detail required. Consequently, mechanisms are needed to identify the level adequate for the task at hand as well as to guarantee a smooth change between the associated models. Finally, description methods must be offered which enable users to comfortably express the level of detail they are interested in.

RETRIEVAL IN LARGE KNOWLEDGE BASES

One reason for expert systems to have large knowledge bases lies in the *diversity* of knowledge needed in a wide-focussed task domain. About hypertension consultation we noted that specific knowledge about heart, arteries, effects of pregnancy, etc. is required and that some of this knowledge needs only to be considered at given evidence that it is relevant. It is characteristic for a human expert's ability that rapid access to relevant knowledge in diverse situations is often achieved without actually attempting to apply irrelevant facts. For an artificial system to exhibit such a selective retrieval, it needs to be capable of locating elements of its knowledge with respect to their relevance for a given task.

By evidence from empirical findings presented in (Wachsmuth, 1989), the association of significant domain vocabulary with knowledge elements constitutes a primary means of access to specific domain knowledge. Rapid access to diverse knowledge is often triggered by wordings or technical terms in conveying a question or a problem statement. The strong hypothesis evolving from these observations is that topical knowledge of different domains or subdomains is semantically localized in the body of knowledge a human possesses.The *principle of keyword-dependent access to knowledge* formulated in the second section reflects this hypothesis. It suggests associating vocabulary terms (concept signifiers) with packets of knowledge elements that deal with those concepts. The resulting interaction between term words and knowledge elements may also be involved in choice making among competitive knowledge packets when structure-dependent access is faced with alternatives.

The above observation which attributes a semantic locality to significant vocabulary seems further supported by findings in brain research reported on by Ritter (1989). In a famous case of brain accident by stroke, the ability to use concept words (from the category of fruits and vegetables) was impaired selectively. This was interpreted as the result of focal brain damage. Although no direct evidence has been found so far, current work in neural networks suggests assuming a highly structured organization of semantic knowledge which has a spatial manifestation in the brain. Experiments by Ritter and Kohonen (1989) training an array of neural units showed that semantic meaning correlates with spatial grouping when words are presented in a minimal grammatically and semantically well-formed sentence context. Although hypotheses on higher-level brain functions are presently demonstrable only on a very modest scale, it seems motivated to explore new techniques emerging from neural networks as to how they can help to organize symbolically represented semantic knowledge and to guide retrieval in a knowledge base.

Assuming the validity of these assumptions, retrieval in large and diverse knowledge bases would be guided if significant clusters of concept words could be associated with domain-specific knowledge elements. This effect seems equally relevant for focussing on highly specific bodies of domain knowledge (e.g., *heart* vs. *kidney*) as it is for zooming in on levels of granularity (e.g., *texture quality* vs. *outer appearance*). "Focussing" or "zooming" relates to the fact that a dynamic partitioning of a knowledge base in visible, reachable, and invisible knowledge is achieved. The principle of *structure-dependent access* allows to identify further subbodies of knowledge as necessary when they are topically close to domain knowledge currently in focus. The principle of *persistence of access conditions* serves to keep the current "perspective" on the knowledge base that is brought about by access tagging.

Needless to say, setting up a structured knowledge base according to these principles requires a careful analysis of a domain and its significant vocabulary in relation to a task-relevant scenario of knowledge use. The plan is to integrate an associanist network that accepts groupings of domain concept words and generates entry focuses (ACCESSED tags) with a topcically organized hierarchical knowledge base that represents domain knowledge symbolically (see Figure 2 for illustration). The vocabulary network is trained by according sets of concept word clusters associated with topic words indicating entries in a knowledge packet structure. Work is in progress to evaluate the feasibility of this approach in an experimental system.

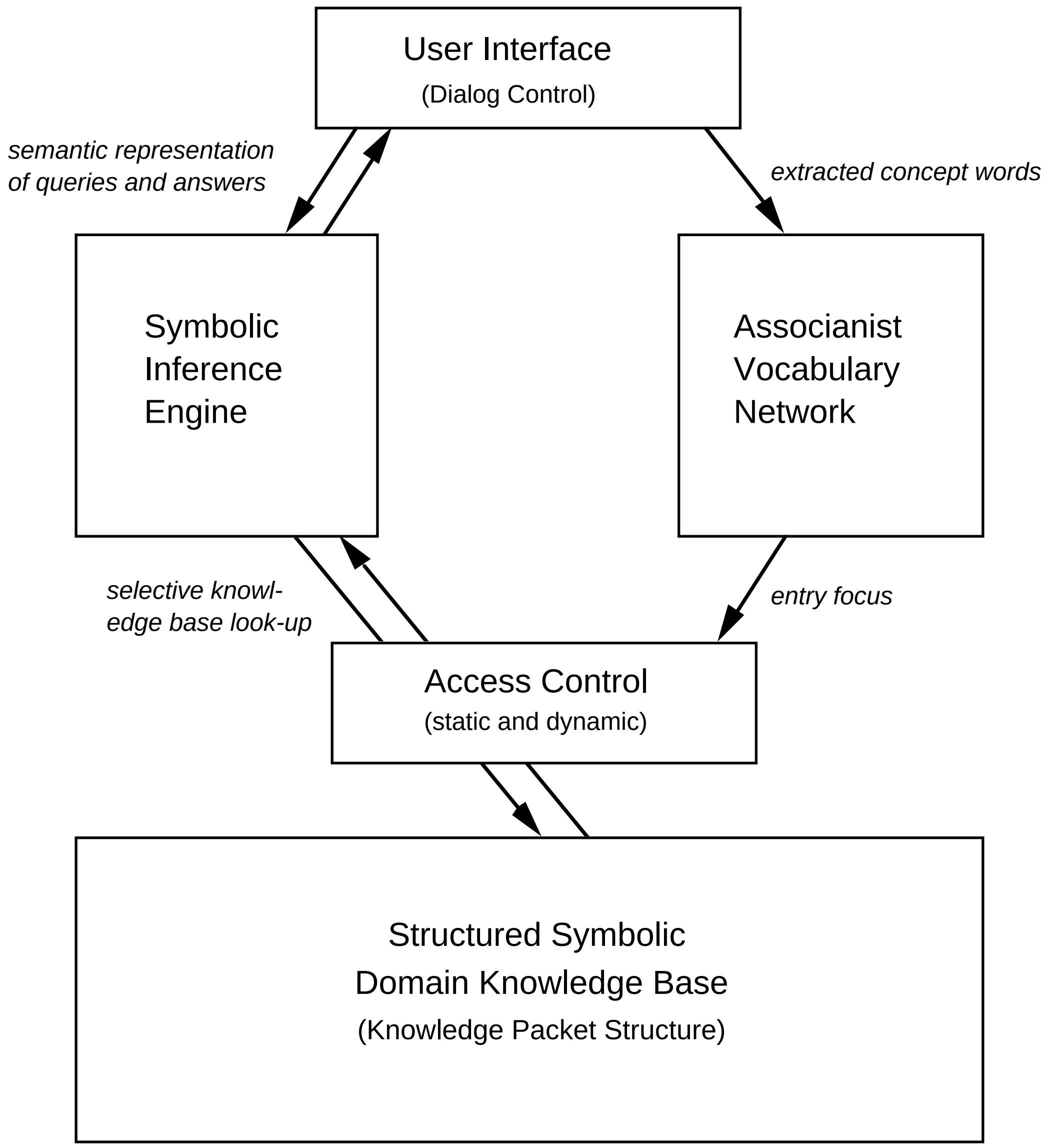

Figure 2. Rough sketch of hybrid system architecture integrating symbolic knowledge base with subsymbolic component realizing keyword-based access to knowledge. Combinations of concept words extracted in linguistic preprocessing generate plausible entry focuses to be used in access control.

CONCLUSION AND PLANNED WORK

The fact that larger and increasingly diverse knowledge bases become manageable seems a primary issue for further progress in AI techniques in general, and in particular in the field of expert systems. Principles of domain-oriented knowledge structuring were discussed in this paper with respect to the construction and use of wide-focussed domain knowledge bases. These principles guide the organization by bundling knowledge elements related to a common context, and they provide means for accomodating conflicting theories in a common system. Static access conditions permit selective knowledge use and prevent the interpretative procedure from attempting all syntactically applicable knowledge elements. Dynamic access conditions describe methods to identify relevant portions of knowledge and strategies how to keep track of them. These ideas are particularly relevant for retrieval in large knowledge bases, however, they are not yet fully worked out; future work will explore how techniques from neural networks can be exploited. The importance of granularity aspects for structuring and accessing knowledge was pointed out; coping with the difficulties of different granularity levels is part of the research agenda.

The retrieval problem stressed in this paper is embedded in more general research aimed at modular development of knowledge-based systems. It is argued that the modularity of an expert system cannot be defined solely in a way that is convenient for software development. It must roughly follow the modularity of an expert's knowledge to aid in the reconstruction and maintenance of that knowledge; see also (Prerau *et al.*, 1990). On the other hand, the system must be modularized in such a way that each developer can work on a module without depending too much on other developers. To achieve this, the system should be divided into several knowledge bases which correspond to recognizable subaspects of the expert domain knowledge. Experimental implementations are planned using a cluster of several UNIX workstations, each of which is equipped with a current generation multiparadigm development tool. Modular parts of the system are to be realized as individual knowledge bases which are saved to and loaded from the shared file server during development. Responsibility for knowledge bases is spread among the team of developers, however, for each knowledge base, a single developer is responsible.

REFERENCES

Chandrasekaran, B. (1987). Towards a functional architecture for intelligence based on generic information processing tasks. Proc. 10th Int. Joint Conf. on Artificial Intelligence (IJCAI-87).

de Wardener, H. E. (1990). Editorial Review: The primary role of the kidney and salt intake in the aetiology of essential hypertension: part I. Clinical Science 79(3), 193-200.

Guyton, A. C. (1981). Textbook of Medical Physiology, Philadelphia 1981.

Heller, B., Kauffmann, K., & Wetter, Th. (1989a). HYPERTON - Prototyp eines wissensbasierten Systems zur Diagnose und Therapie der Hypertonie. Wissenschafts-Journal Forschung und Praxis der Ärztezeitung, Jahrgang 8, Nr. 92.

Heller, B., Kauffmann, K., & Wetter, Th. (1989b). Second generation expert systems: An application for chronic diseases. Proc. Sixth World Conference on Medical Informatics, Part 1 (MEDINFO 89), pp. 204-207. Amsterdam: North-Holland.

Kuipers, B. (1987). Abstraction by Time-Scale in Qualitative Simulation, Proceedings AAAI-87, pp. 621-625.

Lenat, D. B. & Feigenbaum, E. A. (1987). On the thresholds of knowledge. Proc. 10th Int. Joint Conf. on Artificial Intelligence (IJCAI-87), 1173-1182.

Prerau, D. S., Gunderson, A. S., Reinke, R. E., & Adler, M. R. (1990). Maintainability techniques in developing large expert systems. IEEE Expert June 1990, 71-79.

Rettig, R. (1989). Role of the kidney in primary hypertension. University of Heidelberg: Dept. of Pharmacology.

Ritter, H. (1989). Self-organizing maps for internal representations. Bielefeld: University of Bielefeld Research Group on MIND AND BRAIN, Report No. 7/1989.

Ritter, H. & Kohonen, T. (1989). Learning "semantotopic" maps from context. Proceedings of the International Joint Conference on Neural Networks (IJCNN-90), Washington D.C.

Steels, L. (1990). Components of expertise. AI Magazine 11(2), 28-49.

Wachsmuth, I. (1987). On structuring domain-specific knowledge. Stuttgart: IBM Deutschland, LILOG Report 12, March 1987.

Wachsmuth, I. (1989). Zur intelligenten Organisation von Wissensbeständen in künstlichen Systemen. Stuttgart/Heidelberg: IBM Deutschland, IWBS Report 91, November 1989.

Wielinga, B. & Schreiber, G. (1990). KADS: Modelbased KBS development. In Marburger, H. (Ed.): Proc. 14th German Workshop on Artificial Intelligence (GWAI-90) (pp. 322-333). Berlin: Springer-Verlag.