Two Expert System Applications: Implications for Knowledge Representation for Explanations and Justifications

Bob Jansen

CSIRO Division of Information Technology,
PO Box 1599,
North Ryde,
New South Wales,
Australia.
Phone: (+61 2) 887 9489
Fax: (+61 2) 888 7787
Email: jansen@syd.dit.csiro.au

ABSTRACT

This paper discusses interim results from a research project into knowledge representation facilitating explanation and justification in knowledge-based systems. The research program has its roots in the re-development of an expert system Siratac, and the effect of this application on the acquisition and representation of the knowledge to facilitate explanations and knowledge justification.

The paper discusses an initial prototype, an expert system for the evaluation of a dark fibre risk in the selling of wool, which explored the problem of providing suitable explanations of the domain knowledge for a naive end user. The prototype represented the mapping between the various representations of the knowledge, or knowledge sources, as found in the expert system, namely the inference tree itself; research papers used as a knowledge source for parts of the inference tree; and a database also used as a knowledge source for parts of the inference tree.

This work led to the recognition of the importance of an assertion as a representation of one aspect of the knowledge embedded in research papers. There is a discussion of some of the problems associated with this simplified representation, and a description of work in progress. The work examines some of these problems with particular emphasis on intelligent assertion recognition and generation. The results of this work will impact conceptual modelling and thesaurus construction.

The discussion completes with a description of hypothesis testing to be applied to this environment.

1. INTRODUCTION.

The development of Knowledge-Based Systems (KBS) has led to an upsurge of interest in the explanation facilities supporting computer systems. The expert system paradigm, namely that of separating the declarative knowledge from the procedural knowledge (Hayes-Roth *et al* 83), opens up the possibility of explaining the behavior of the system in terms of the declarative knowledge. Facilities of this kind are thought to improve ease of use for the naive or casual end user, as well as for the expert.

Since their introduction, expert systems have tackled this problem in a variety of ways, ranging from simple hard coded textual strings embedded in the particular knowledge representation (Bolam 85), to complex analysis of the inference tree and the particular subsection currently under consideration (eg. Mycin, Buchanan & Shortliffe 84), and in some instances to the display of the entire inference tree.

Our experience in the application of expert systems technology in three application areas, Siratac (Hearn *et al* 86) a cotton pest management system; Garvan ES1 (Horn *et al* 85) a laboratory report interpretation expert system; and a prototype system for the evaluation of the dark fibre risk in wool (Jansen & Robertson 89), shows that the support for explanation and justification required knowledge in excess of that embedded in the inference tree and thus affects knowledge acquisition in increasing the type of knowledge acquired. Kornel 87 hypothesizes that there are two types of thought accessible via knowledge acquisition, *formal* thought and *narrative* thought. Formal thought is that body of knowledge capable of being represented in a formal framework, eg. formal logic, frames, semantic nets, production rules, etc., whilst narrative thought is knowledge that is too diffuse and ephemeral to be easily acquired and represented. It is hypothesized that in order to satisfactorily support explanation and justification facilities for a KBS, the narrative thought surrounding the formal thought is a mandatory product of knowledge acquisition. The narrative thought includes hunches, feelings, intuition, leaps of insight, etc. It could be described as the common sense knowledge that others, eg. the CYC Project (Lenat & Guha 88, Guha & Lenat 90), claim is necessary for intelligent behavior.

In common with the CYC project, this project uses knowledge embodied on paper, specifically research papers, as the source of both the formal and narrative knowledge supporting the human expert. The aim is to cull from the research papers the various abstract knowledge forms that encapsulate the knowledge the author is attempting to convey to the reader, and in a sense, to "animate" this knowledge via its use in the explanation/justification sub-system of the KBS.

Animating traditionally static forms of knowledge representation, viz. written text, tables, graphics, raises philosophical questions regarding the method of information dissemination in society, its relationship to the medium used to effect this dissemination, and how new media, eg. hypermedia, may affect this process. A process of producing a hypermedia form of the author's PhD thesis in parallel with a traditional paper copy, is in progress as a collaborative venture with a major publisher. The aim of this collaboration is to begin to answer two basic questions :-

- how good is hypermedia at representing the intent of the author;
- how good is hypermedia in transferring the author's intent to the reader.

This project is still in its infancy and will be the subject of a subsequent report. Marcure 90 discusses some of these issues and describes an experiment which attempted to measure the usefulness of hypermedia (in the form of Hypercard[†]) as an information dissemination or publishing tool.

2. EXPLANATION IN CURRENT EXPERT SYSTEMS

The Expert System (ES) paradigm requires the system to be capable of explaining its reasoning (Hayes-Roth *et al* 83). This requirement can be said to be responsible for the separation of the declarative knowledge from the procedural knowledge, which forms the backbone of this paradigm.

In some ES, this facility is supported by the inclusion of a static textual 'explanation' string, incorporated into the structure of the production rule representation, that could be displayed on request (Bolam 85). The extent of the explanatory power is limited by the semantics of the text that comprised the 'explanation' text and is not context dependent. An example is shown in figure 2.1. In this case, the 'WHY' string would be displayed as a result of the user requesting an explanation.

Other ES support explanation by rewording the question being posed. The assumption made here is that the rewording will make clear why the question is being asked. Obviously, the explanatory capability of

---

† Macintosh & Hypercard are registered trademarks of Apple Computers.

this scheme is limited by the apparent intelligence of the rewording system. Examples range from that shown in figure 2.2, which is an exemplar of the 'unintelligent' extreme. In this case, the question text is preceded by the string "Because I am trying to determine". In this case, the system is not explaining its reasoning so much as attempting to justify why the question is being asked, albeit in a terse, unhelpful form.

RULE I6
IF the animal is a mammal, and
it has pointed teeth, and
it has claws, and
its eyes point forward
THEN
it is a carnivore
WHY "This rule is one that helps to determine whether the animal is a carnivore. It is generally used when the animal has not been observed eating, and so it cannot be decided directly whether it eats meat (see also rules I5 and I9."

(From Bolam 85)

FIGURE 2.1 - Example of simple explanation text string included in the production rule

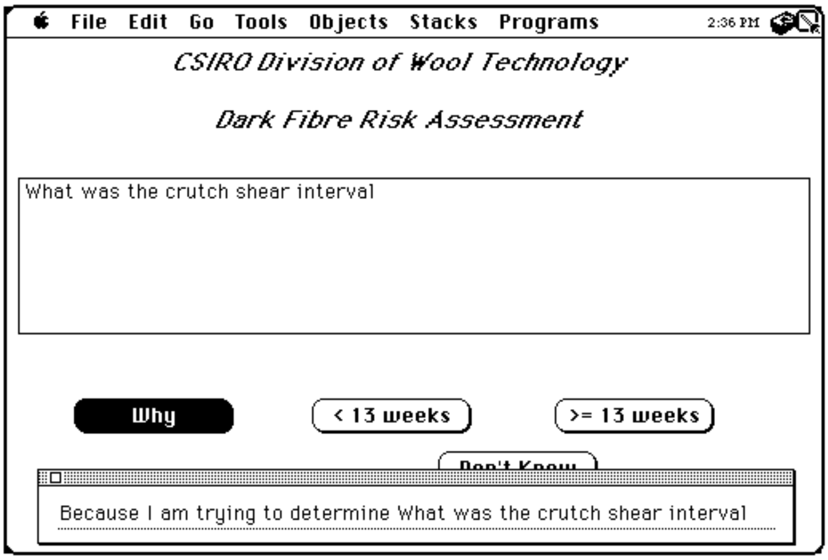

FIGURE 2.2 - Example of question rewording for explanation

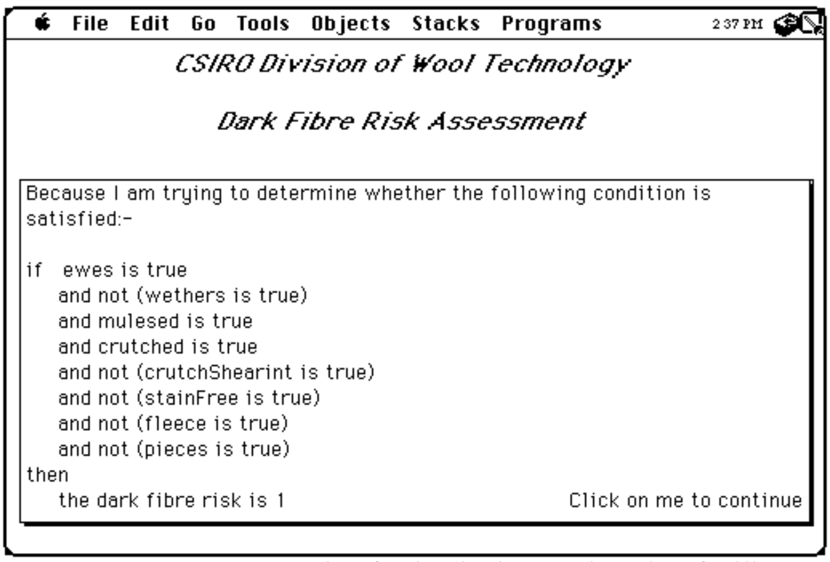

FIGURE 2.3 - Example of rule display explanation facility

Other ES may display the current rule under consideration, possibly in pseudo-natural language form. This scenario assumes that the knowledge encoded in the production rule paradigm is suitably visible to the user. Limitations inherent in this scheme include: the explanatory power is limited by the semantics of the production rule; the scheme assumes that all relevant knowledge has been captured and represented in the production rule; the scheme is not context dependent. An example is shown in figure 2.3.

A hypothesis, based on experiences with domain experts as well as end users, is that any of these simple facilities can only suffice to baffle the user in the majority of cases

It is possible that such terse unhelpful types of explanation may only serve to alienate the user from the ES, thus violating one of the most important principles of the ES paradigm.

3. EXPLANATION IN KNOWLEDGE-BASED SYSTEMS

The trend in ES today is towards Knowledge-Based Systems (KBS), systems combining the power and facilities of conventional information systems (IS) with those of ES. As shown in figure 3.1, Siratac (Hearn *et al* 86) is an exemplar of a KBS.

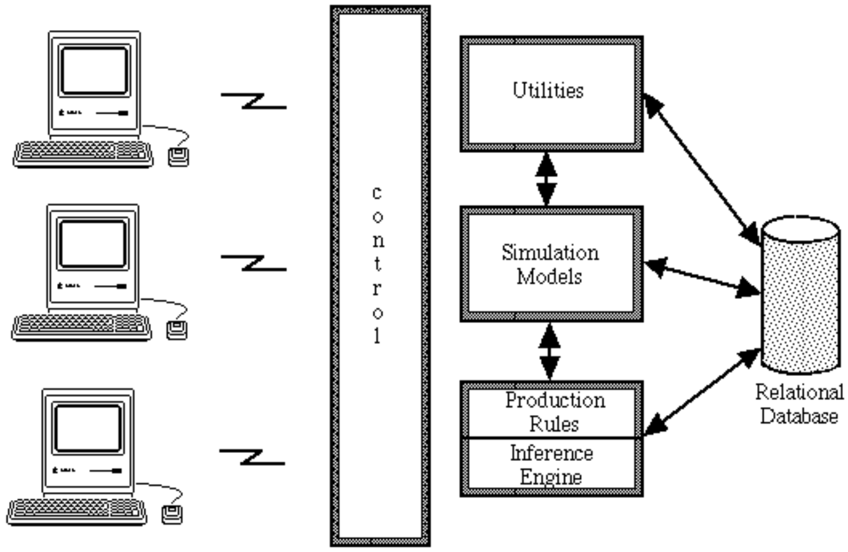

FIGURE 3.1 - Siratac system architecture

Siratac is composed of a relational database, a set of simulation models written in Fortran, a suite of utilities implemented in the 4th generation language VAX-RALLY, an overall control module, a graphic based user interface (implemented on a Macintosh computer linked to the central Vax computers), and a knowledge-based component implemented in the OPS5 production language.

The experience with Siratac suggests that this architecture is more likely to be manifested in KBS instead of the traditional ES architecture, a system implemented entirely in a knowledge language (eg. Prolog, OPS5, etc.).

This claim is based on studies (eg. Jansen 90, Debenham 89) that demonstrate that:-

- KBS show software life-cycle properties equivalent to that of conventional IS
- KBS will expand over time either as a result of domain expansion or knowledge refinement
- the integration of artificial intelligence (AI) and IS technologies will lead to KBS technologies subsuming both AI and ES.

One result of the development of KBS is that properties associated with either major component will now be associated with both. It is maintained that this development will ensure that explanation and justification facilities will be required for the IS components as well. This will have a major impact on the software engineering strategies employed for the development for IS. These strategies will have to be modified to enable the acquisition and representation, for subsequent processing, of a larger amount of knowledge in various more declarative forms. Support for explanation and justification will extend to the database schema, simulation models, the human-computer interface (HCI) component, in fact any and all components of the KBS. The effect of this will be to place even more emphasis on the acquisition and representation of the narrative thought associated with the formal thought, as well as the development of intelligent strategies for context-dependent explanation and justification. In conventional system analysis, this may be recognized as the movement to incorporate design rationales in the system documentation.

Siratac again offers a suitable example. Siratac was able to explain its decisions, based on data and conclusions as a result of the inferencing process, but if a conclusion was drawn based on data supplied by a simulation model, the value of the data item was unexplainable and not justifiable. Neither was it possible to inform the user, expert or otherwise, of the assumptions that underlay any of the simulation models. Apart from the actual Fortran code itself, there was no representation of the knowledge that was encoded in the simulation models enabling an explanation of what actually occurred, or should have occurred, in the run of the model.Thus it is concluded that:-

- facilities similar to those described in Bratko *et al* 90 for Kardio would be necessary
- qualitative, or deep reasoning, models of the simulations would provide more suitable explanations of the process implemented in simulation models to be generated
- enhanced information retrieval access to large databases of loosely structured information, ie. research papers, would support the explanation and justification facilities.

4. THE NEXT GENERATION

4.1. The Wool Technology Prototype

The research program currently underway aims to develop the tools and techniques to support the requirements described above. A prototype environment[1] that integrates the ES with a database of research papers has been produced, enabling the user to be provided with relevant sections of any of the research papers thought to explain, justify, or describe underlying assumptions of the current topic. Figure 4.1 shows the system architecture.

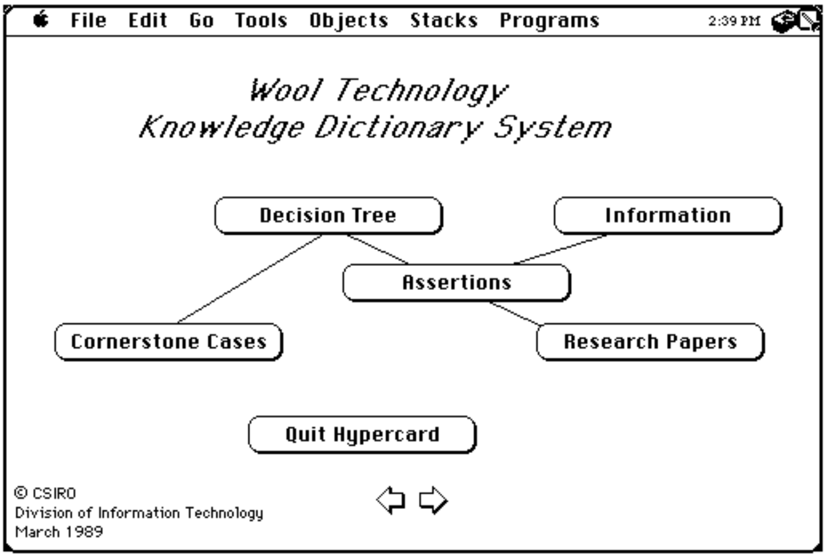

FIGURE 4.1 - System architecture of prototype environment

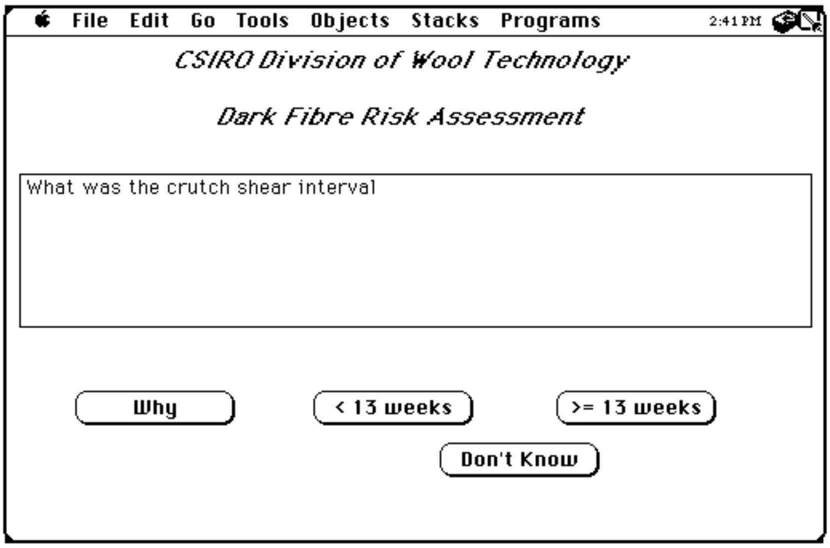

FIGURE 4.2 - Question posed by expert system with the availability of a "Why" facility

This development has hinged upon the concept of an *assertion*, an important meaningful statement made by an author to convey their hypotheses, experimental data, and conclusions to the reader. The project captured, in full, five research papers[2] associated with the domain of the dark fibre risk in wool[3]. These papers were manually processed by the authors and experts to extract the relevant assertions[4]. The set of assertions is meant to represent the important statements, taken out of context and edited to ensure readability, that represent one aspect of the domain knowledge.

The concept of an assertion as implemented in this prototype is a simplified version of what will be implemented in the final environment. It was assumed that, from a data modelling point of view, the assertion is used to map a particular node of the inference tree to a particular area of a research paper. The situation in reality is more complex than this, an assertion may refer to many nodes or paths through the inference tree as well as to many section of the papers. This was however deemed beyond the scope of the prototype which aimed at showing the usefulness of the assertion as a knowledge mapping concept. As a further simplifying measure, the system does not recognize duplicate assertions, instead each assertion was manually edited to ensure non-duplication.

---

[1] This prototype was produced as part of a collaboration with the CSIRO Division of Wool Technology.

[2] In total, 56 research papers are represented, but of those 51 are references from the initial 5.

[3] The ES is used to rate a bale of wool for the amount of dark fibre. Dark fibre is undesirable, as wool which contains a great deal of it is unsuitable for white and pastel garments which command the highest price. The current threshold is 100 dark fibres per kilogram of wool.

[4] The total number of assertions represented is about 100. See section 5.2 for more details of future research in this area.

The production of the assertion representation of the papers enabled a more intelligent, context dependent explanation and justification facility to be implemented in this ES. Functionally, the user, when asked for the value of a data item, may elect to request explanations.

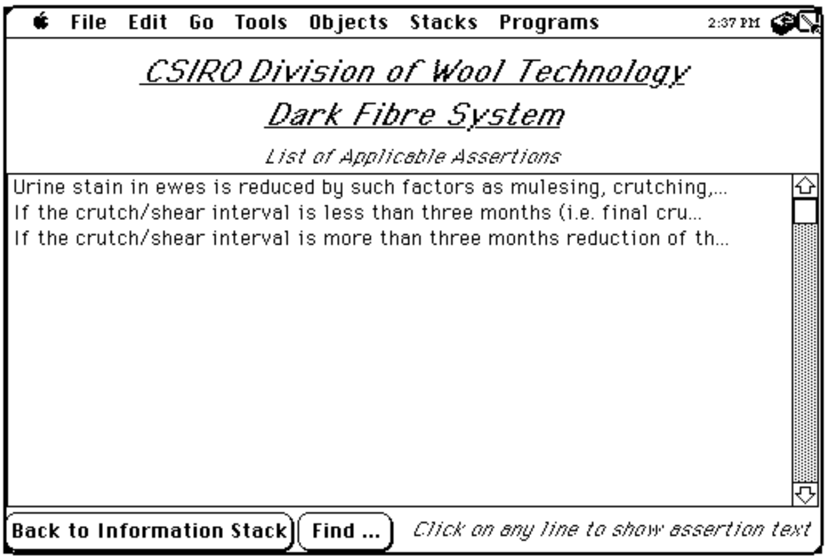

FIGURE 4.3 - List of assertions relevant to the question

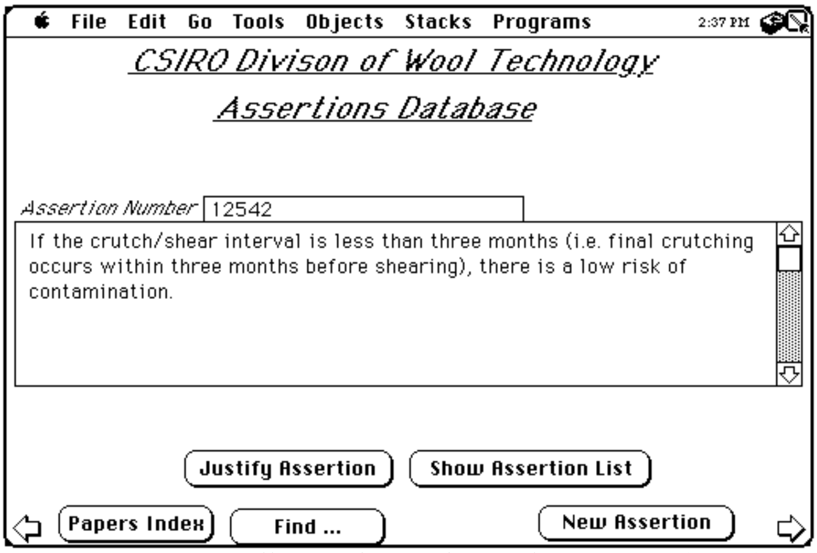

FIGURE 4.4 - Full text of one of the relevant assertions, ie. number 2 in figure 4.3

At this point, there are interesting philosophical questions regarding the semantics of the 'why' facility, namely what does 'why' actually mean in this context? Can it be answered by re-phrasing the question, as described above? Can it be answered by showing the current rule, as described above? Is the user asking what the data item represents, for example what does the term "crutch-shear interval" in figure 4.2 actually mean? What is the importance of the value 13 for crutch-shear interval, why not 42? Is the user interested in the relationship between the data item, the value to be supplied, and the result of the ES (ie. the dark fibre risk rating supplied)? Each of these possibilities can be supported by the assertion representation, assuming of course that all the assertions have been represented.

In this prototype, the request for an explanation, expressed by pressing the "Why" button, leads to a search of the assertion list using a fixed keywords - out - of - context (KWOC) facility, retrieving assertions containing predefined keywords. The retrieved list is displayed to the user (figure 4.3).

If required, the user can select a particular assertion from the list and request its full display, instead of the first 70 characters as in the assertion list (figure 4.4). In this case, the user is presented with the full text of the assertion and its unique reference number.

Once the full text of the assertion is displayed, the user can request a justification of the assertion, leading to a display of that section of text from a research paper that caused the assertion to be represented (figure 4.5). If the text of the assertion is an exact match of any part of the text in the section of the paper being displayed, then this subsection is highlighted. This situation is however, rare, as in most cases the assertion has been edited to ensure readability in its out-of-context representation. Note the hypertext browsing capability as the link to another representation of the knowledge, a graph, from the reference point “figure 4”.

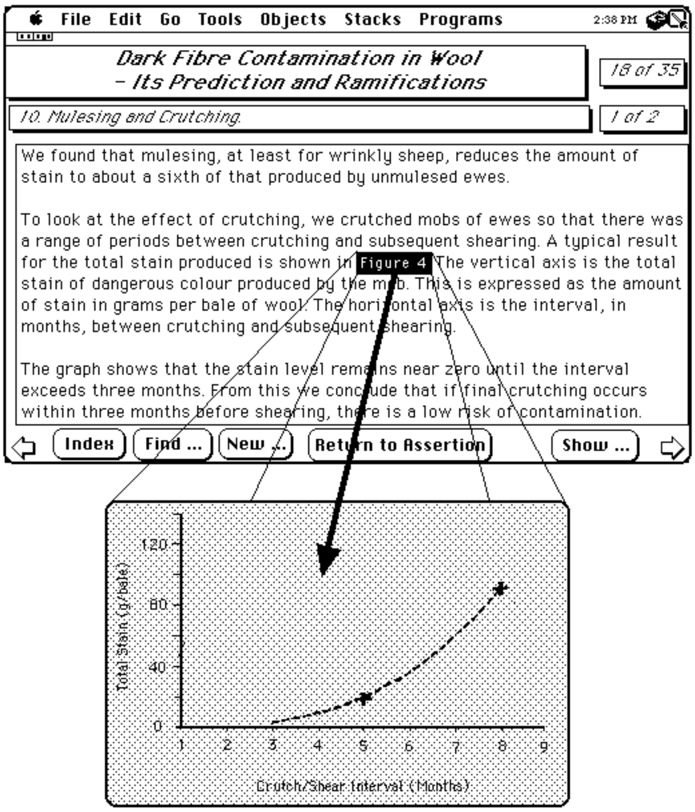

FIGURE 4.5 - Section of text that produced the assertion and hypertext link to accompanying graph

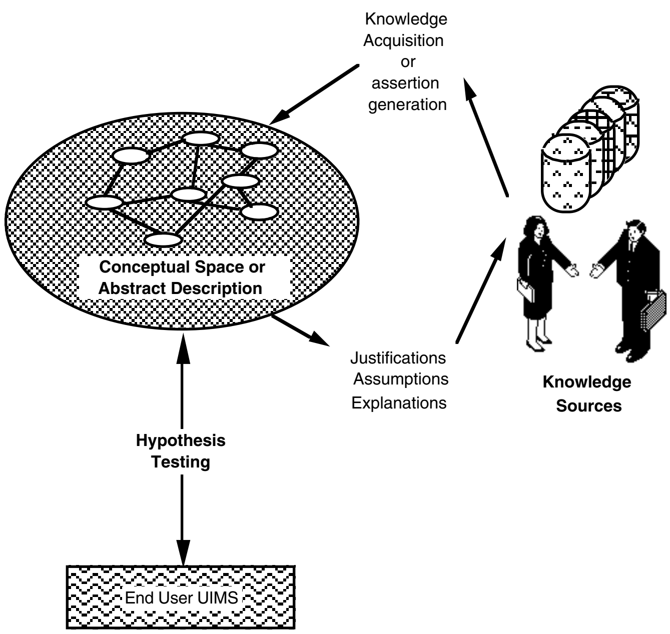

FIGURE 5.1 - System diagram of future research

5. FUTURE WORK

Future research aims at extending the basic work of the prototype by initially concentrating on: assertion recognition and generation; production of a conceptual model of the domain; and hypothesis testing. Figure 5.1 shows a plan for the future work.

5.1 Conceptual Modelling

The conceptual model, or abstract description, will act as the central pivot mapping between the knowledge chunks represented within the various subsystems. The aims are to :-

- discover the ontology of an assertion
- discover the structure of an assertion
- discover assertion subtypes
- show the nature of the mapping function
- represent more fully the semantics of an assertion.

This will be tested by producing the conceptual model of each assertion using the semantic net representation[5]. It is hypothesized that the KWOC facility currently implemented corresponds to the nodes of the appropriate semantic net of an assertion. More importantly, this will also include the semantics of the relationships between the concepts, thus increasing the number of retrieved text sections. In addition, the conceptual model will enable the detection of possible duplicate assertions, those assertions that describe the same, or nearly the same, conceptual models. The conceptual model will act as a filter to the research papers akin to the semantic net model in some text retrieval applications (eg. Cohen & Kjeldsen 87).

We are also in the process of producing a thesaurus of this domain, and representing the thesaurus in the conceptual modelling tool, thus forming the core of the domain conceptual model[6]. An interesting offshoot of this research will begin to uncover how the advent of hypermedia technology may affect the production of a thesaurus, and its knowledge/indexing structures[7].

5.2. Assertion Generation

In collaboration with Professor Ann Henderson-Sellers of Macquarie University[8], Sydney, Australia, work has begun aimed at discovering heuristics describing how experts recognize assertions in research papers.

---

5 In collaboration with Professor Colette Rolland, of the Université Paris 1 - Sorbonne, we are extending a conceptual modelling

An outcome of this research will be more intelligent tools supporting the experts in the reverse engineering of the assertions from existing research papers. The knowledge could be sourced from any parts of the paper, ie. text, tables, graphics, etc., and also from the various abstract structures that form part of the paper, ie. rhetorical structures and devices, etc. (see for example, Alvarado *et al* 90).

5.3. Hypothesis Testing

In collaboration with Paul Compton of the University of New South Wales, a project to support hypothesis testing (Feldman *et al* 89) on the developed conceptual models and underlying data from the research papers representation[9] has begun.

Hypothesis testing is an activity that enables a researcher to trial new hypotheses using conceptual models and data accessible via conventional research papers. Conventional information retrieval does not cater for this type of activity. Research into improving electronic publishing and retrieval is generally aimed at conventional notions of scientific publication. That is, the publication is intended to convey the hypothesis and findings that the authors wish to communicate using experimental data to support these hypotheses. However reading papers to see if there is any data in the paper which relates to a hypothesis that is quite different from the concerns of the original author is not supported. This research aims to tackle these different requirements starting with the representations of publications resulting from the assertion recognition and representation.

The work on conceptual modelling and hypothesis testing is seen as an adjunct to that work in information retrieval, and as such will be influenced by, and in turn influence, this technology.

6. CONCLUSIONS

This paper has described interim results of a research program aimed at producing knowledge representations supporting intelligent explanation and justification for Knowledge-Based Systems.

We have hypothesized that Knowledge-Based Systems, namely those systems subsuming conventional information systems and current expert systems, will require extensive support for intelligent explanation and justification for all the components of those systems. This will impact on knowledge acquisition process as this must be capable of acquiring both the formal knowledge, as currently captured and represented in system documentation and expert systems, as well as the narrative knowledge that will be required to support intelligent explanation and justification. The research so far has indicated that other available knowledge sources must be brought to bear on to this problem, and there is work in progress attempting to utilize databases of loosely structured information, namely research papers.

We have produced an abstract representation of one aspect of the knowledge embedded in the text of research papers, a representation are termed assertions. It has been shown, by way of a prototype system, that this representation does indeed function successfully as a mapping representation between the knowledge embedded in expert systems, and the representation of the same knowledge in research papers. The paper has detailed some of the limitations of this representation, and has described work in progress and future work aimed at compensating for these limitations.

ACKNOWLEDGEMENTS

We acknowledge the advice and guidance of: Dr. Bob Colomb of the CSIRO Division of Information Technology (CSIRO DIT); Ms. Judy Marcure of CSIRO DIT for many interesting discussions about publishing; Dave Charlton and Rob Rottenbury and their colleagues from the CSIRO Division of Wool Technology who provided the stimulus for the initial prototype; Professor Ann Henderson-Sellers and Dr. Mervyn Jones and their colleagues from the Macquarie University School of Earth Sciences who were the stimulus for extending this research into the present prototype; Paul Compton of the University of New South Wales Department of Computer Science, for many stimulating discussions

---

tool based on the Remora IS design methodology (Rolland & Richard 82) to cater for the conceptual modelling of knowledge. This tool will be used to produce the conceptual models of each assertion.

6 This work is a collaborative project with the University of Technology (Kuring-gai).

7 This environment is being implemented using the Hypercard† software on Macintosh† computers.

8 This collaboration will focus on the domain of the role of carbon dioxide in the greenhouse effect.

9 It is planned, wherever possible, to represent graphs and tables from research papers using their underlying unprocessed data, thus enabling the reader: to critique the author's processing functions of the data; to display the data in a range of supported graphic representations (ie. pie charts, bar charts, etc.); perform 'what-if' type enquiries by altering data values and re-graphing the data.

and access to his hypothesis testing work; Ms. Hillary Yerbury and Ms. Linda Whitford of the University of Technology (Kuring-gai) for the thesaurus preparation; and various colleagues at the Division of Information Technology for their help, advice and support.

REFERENCES

Alvarado SJ, Dyer MG & Flowers M, *Argument Representation for Editorial Text*, Knowledge-Based Systems, Vol. 3, No. 2, June 1990, pp87-107

Bolam WJ, *Explanation in Expert Systems*, in Proceedings of the First Pan Pacific Computer Conference, Melbourne, Australia, 1985, pp 898-917

Bratko I, Mozetic I & Lavrac N, *Kardio: A Study in Deep and Qualitative Knowledge for Expert Systems*, The MIT Press, 1990, ISBN 0-262-02273-7

Buchanan BG & Shortliffe EH (eds), *Rule Based Expert Systems, The MYCIN Experiments of the Stanford Heuristic Programming Project,* The Addison-Wesley Series in Artificial Intelligence, reprinted 1984

Cohen PR & Kjeldsen R, *Information Retrieval by Constrained Spreading Activation in Semantic Networks*, Information Processing & Management, Vol. 23, No. 4, 1987, pp255-268

Debenham JK, *Knowledge Systems Design*, Prentice Hall, Advances in Computer Science Series, Brent RP (ed), 1989, ISBN 0-130516428-1

Feldman BZ, Compton PJ & Smythe GA, *Hypothesis Testing: An Appropriate Task for Knowledge-Based Systems*, Proceedings of the Fourth Knowledge Acquisition for Knowledge Bases Systems Workshop, J Boose & B Gaines (eds), SRDG Publications, University of Calgary, Canada, 1989, pp 10-1 10-20

Guha RV & Lenat DB, *Cyc: A Midterm Report*, AI Magazine, Vol. 11, No. 3, Fall 1990, pp32-59

Hayes-Roth F, Waterman DA & Lenat DB (eds), *Building Expert Systems*, Advanced Book Program, Addison-Wesley Publishing Company Inc. 1983

Hearn B, Brook K, Ashburner A & Colomb R, *SIRATAC, A Cotton Management Expert System*, in Proceedings of the First Australian Artificial Intelligence Congress, Melbourne, Australia, 1986

Horn K A, Compton P, Lazarus L, Quinlan R, *An Expert System for the Interpretation of Thyroid Assays in a Clinical Laboratory*, The Australian Computer Journal, Volume 17, No 1, February 1985

Jansen R, *The Knowledge Dictionary: A New Methodology and Tool for the Engineering of Expert Systems*, PhD Thesis, University of New South Wales, 1990.

Jansen B & Robertson J, Management of Wool Dark Fibre Risk Knowledge Using Hypertext, CSIRO Division of Information Technology, Technical Report TR-FD-89-05, May 1989

Kornell J, *Formal Thought and Narrative Thought in Knowledge Acquisition*, International Journal of Man-Machine Studies (1987) 26, pp203-212

Lenat D & Guha R V, *The World According to CYC*, MCC Technical Report ACA-AI-300-88

Marcure J, *Dynamic Documents and Print Conventions*, CSIRO Division of Information Technology Technical Report TR-FD-90-08, 1990.

Rolland C & Richard C, *The Remora Methodology for Information Systems Design and Management,* Information Systems Design Methodologies: A Comparative Review, T W Olle, H G Sol, A A Verijn-Stuart (eds), North Holland Publishing Company, 1982, pp 369-426