The DARPA High Performance Knowledge Bases Programme

John Kingston*

AIAI
University of Edinburgh

Abstract. The HPKB programme aims to enable the development of very large knowledge bases which are comprehensive, reusable, maintainable, and address areas of interest to DARPA. These knowledge bases are expected to require foundation knowledge, knowledge acquisition facilities, and efficient problem solving techniques. The programme encourages development of multiple applications which will be tested against "challenge problems" at the end of each year of the programme.

1 Introduction

The goal of the DARPA-funded High Performance Knowledge Base programme is to produce the technology needed to enable system developers to construct rapidly (i.e. within months) knowledge bases of hundreds of thousands of axioms, rules or frames that provide comprehensive coverage of topics of interest, are reusable by multiple applications with diverse problem-solving strategies, and are maintainable in rapidly changing environments. It is envisioned that the process for constructing these knowledge bases will involve three major steps:

- **Building foundation knowledge:** creating the foundation knowledge (e.g. selecting the knowledge representation scheme, assembling theories of common knowledge, or defining domain-specific terms and concepts) to enable the construction and population of large, comprehensive knowledge bases for particular domains of interest – by selecting, composing, extending, specialising, and modifying components from a library of reusable ontologies, common domain theories, and generic problem-solving strategies.
- **Acquiring domain knowledge:** constructing and populating a complete knowledge base by using the foundation knowledge to generate domain-specific knowledge acquisition, data mining and information extraction tools. This will enable collaborating teams of domain (non-computer) experts to easily extend the foundation theories, define additional domain theories and

* John Kingston and Professor Austin Tate are the co-Principal Investigators on AIAI's "Knowledge Engineering Window on Europe" project, which is part of the HPKB programme. Thanks are due to Dr. Dave Gunning, the HPKB Program manager for DARPA, for giving his permission to describe the programme to EKAW-97 participants.

problem solving strategies, and acquire domain facts to populate a comprehensive knowledge base covering the domains of interest
– **Efficient problem solving:** enabling efficient problem solving, either by providing efficient reusable inference and reasoning procedures to operate on a knowledge base, or by providing tools and techniques to select and transform knowledge from a knowledge base into optimised problem solving modules tailored to the unique requirements of an application.

The objective of HPKB is to develop, integrate and test the technology needed to enable this process, beginning in 1997 and ending in 2000. The intention is to produce knowledge base development environments which combine the necessary foundation building, knowledge acquisition and problem solving technologies into an integrated development environment, and to use these environments to build reusable knowledge base components for multiple DARPA application projects.

The candidate applications are a set of new and on-going DARPA initiatives, all of which are building advanced information systems to improve some aspect of military operations, and all of which need knowledge-rich components. These projects include:

– **Dynamic Multi-user Information Fusion (DMIF)**, which is providing the joint warfighter with a clear and actionable picture of the battlespace through development of advanced technology to support active fusion management and generation of information products tailored to specific and evolving needs;
– **Joint Task Force (JTF) Advanced Technology Demonstration (ATD)** which is developing a portable and supportable technology base to provide command, control, communications and computer information (C4I) to a deployed Joint Task Force Commander for JTF crisis management, planning and execution.
– **Technology Development for the Joint Force Air Component Commander (JFACC)** which is developing information technologies to revolutionise the air operations planning processes, in order to enhance significantly the JFACC's ability to plan and execute and air operation at any level of crisis – major regional conflicts, coalition warfare, and operations other than war.
– **Advanced Logistics Program (ALP)** which is developing advanced information technology to support planning, execution, monitoring and replanning throughout the logistics pipeline, enabling the warfighter to project and sustain overwhelming combat power sooner and with less reliance on large DoD inventories.
– **Battlefield Awareness Data Dissemination (BADD)** which is providing warfighters at echelons from the Task Force Commander down to Battalion or lower, and especially mobile warfighters, with advanced battlefield awareness applications that are driven by near-real time data that is delivered by advanced data dissemination methods.

– **Information Gathering, Processing and Analyses in support of Crisis management (Project Genoa)** which is developing information tools to assist crisis managers at the highest decision levels of Government in discovering, organising, visualising and understanding information about nascent and emerging crises.

2 Development approach

The HPKB program is building multiple knowledge base development environments which will be exercised against two *challenge problems*. At the end of each year of the project, the knowledge base developers will be presented with a hypothetical situation which describes the domain to be covered, defines (using a formal ontology) the terms to be used in specifying input requests or test questions, and includes appropriate sample questions and training data. The developers will then be asked to build a knowledge based system to the problem specifications, recording how long it takes; answer a set of test questions, and record the percentage of correct answers, then modify their system to deal with a changed requirement and re-evaluate it.

3 Challenge problems

The two challenge problems will be in the following areas:

1. Crisis understanding. The aim will be to build a broad but shallow knowledge base of general geo-political knowledge to support information retrieval for crisis detection and understanding. The knowledge of the crisis understanding process will probably be drawn from sources within the CIA, either directly or via knowledge acquired for Project Genoa; the hypothetical scenario will provide some data, and further data can be gathered from open source environments such as the CIA World Factbook and the Internet. In addition, it is expected that the systems will draw on a "corporate memory" of previous crises; the form of this is yet to be determined.
2. Battlefield awareness. The aim will be to build an in-depth knowledge base of battlefield situations, operations, and logistics to support multiple military applications. The sources of knowledge for this challenge problem are yet to be defined, but are likely to include the DMIF, ALP and JFACC programs.

4 Program organisation

The organisation of the program is shown in Figure 1.

The challenge problems are being developed by two organisations, Alphatech and IET. The remaining participants are being directed towards two integration environments, in order to ensure that the multiple applications which are produced are capable of interoperability. In addition, at the HPKB kick-off meeting

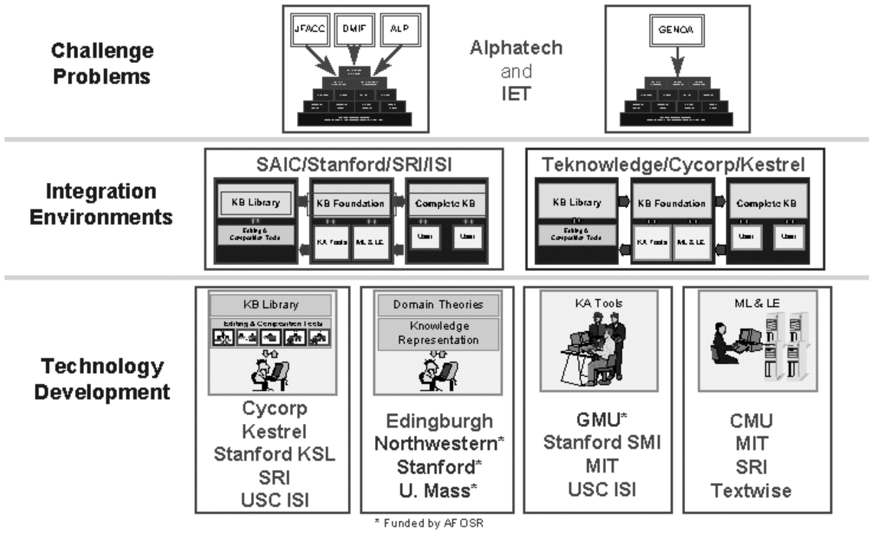

Fig. 1. HPKB Program Organisation

in June 1997, it was agreed that the Generic Frame Protocol (GFP) from SRI [SRI1997] would be used as a common basis for interoperability.

The work of some of the participants – those with whom AIAI will collaborate closely in its own work programme under HPKB – is outlined below. For more details of these and other projects, see [DARPA1997].

- **Cycorp** [Cycorp1997] is a company formed to exploit the CYC knowledge base. CYC can be considered as a very large ontology, with nearly 200 microtheories, 18,000 concepts, and 175,000 axioms; the aim within HPKB is to allow technology developers to represent knowledge quickly and faithfully.
- **ISI** from the University of Southern California aim to integrate problem-solving methods and ontologies into applications. Their work is based around EXPECT [Gil & Melz1996], which is a framework for developing KBS applications; EXPECT checks the domain information it has been given, and prompts the user to fill in information which it needs to know but does not have. This knowledge acquisition is guided by a pre-defined Interdependency Model.
- **Stanford University** (specifically, Mark Musen's group in Medical Informatics) aim to extend their existing work on Protege [Gennari1997], which supports the construction of KBS systems and knowledge maintenance tools. Protege-II has been dissembled into a collection of problem-solving components which can be mixed and matched for a particular application; the aim is to make these components more widely available by integrating with others in the HPKB program, and by making the code more widely available (e.g. by adding CORBA-compliant features).

– **MIT's** Clinical Decision Making Group have developed MAITA (the Monitoring, Interpretation and Analysis Tool Arsenal) which is a rich library of implemented problem-solving components for monitoring tasks [Doyle1997].

5 AIAI's work

AIAI's work is organised around four themes: awareness, applications, research and responsiveness.

Awareness: AIAI's aim is to provide the HPKB program with a "Window on Europe", in order to perform a transatlantic technology transfer role. The aim is to inform the U.S. of related work, particularly European work on knowledge-level methods.

Applications: We aim to support the HPKB programme using multi-perspective modelling approaches to capture the generic problem solving processes used in the challenge problems. The target models will be ontologically underpinned, probably using the ontology of CommonKADS models, but possibly making use of other modelling techniques such as IDEF3 (see [Kingston, Griffith, & Lydiard1997] for some of our previous work in this area). Outstanding modelling issues will be captured to drive the modelling process itself, thus providing a top-down approach to model development.

Research: The major focus of our awareness work will be to collect problem solving methods from around the world and to make them available to HPKB participants in a single library. The appropriate format, content, and level of abstraction of the library are all issues which need careful consideration if the library is to be both accurate and usable.

We also plan to move ahead on the representation of generic problem solving *processes*, perhaps using ontologies of processes, activities, roles, agents, or process products.

Responsiveness: We aim to tailor our plans to the needs of the initiative. We have already responded to interest in some of our previous work; further work may include organisational modelling for the battlefield awareness challenge problem, a "deep slice" demonstrator of expertise modelling for crisis understanding, a broad view on knowledge asset management within the relevant military organisations, a survey of knowledge acquisition tools, or reports and briefings on European knowledge representation techniques (such as CommonKADS) and standards (such as those emerging from the EuroKnowledge initiative).

Acknowledgements

This work is sponsored by the Defense Advanced Research Projects Agency (DARPA) and Rome Laboratory, Air Force Materiel Command, USAF, under grant number F-30602-97-1-0203. The project is monitored by the USAF Rome Laboratory. The U.S. Government is authorised to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation hereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing official policies or endorsements, either express or implied, of DARPA, Rome Laboratory or the U.S. Government.

References

[Cycorp1997] Cycorp. 1997. Cycorp, Inc. http://www.cyc.com/.

[DARPA1997] DARPA. 1997. High Performance Knowledge Bases. http://www.teknowledge.com/HPKB/.

[Doyle1997] Doyle, J. 1997. The MAITA project. http://medg.lcs.mit.edu/projects/maita/.

[Gennari1997] Gennari, J. 1997. HPKB Project at the Section on Medical Informatics. http://www-smi.stanford.edu/projects/protege/Hpkb-web/.

[Gil & Melz1996] Gil, Y., and Melz, E. 1996. Explicit Representation of Problem-Solving Strategies to Support Knowledge Acquisition. In *Proceedings of AAAI-96*. AAAI.

[Kingston, Griffith, & Lydiard1997] Kingston, J. K.; Griffith, A.; and Lydiard, T. 1997. Multi-Perspective Modelling of the Air Campaign Planning process. In *Proceedings of IJCAI-97*. AAAI Press.

[SRI1997] SRI. 1997. Generic Frame Protocol (GFP). http://www.ai.sri.com/~gfp.