

AFRL-IF-RS-TR-2002-2
Final Technical Report
January 2002



LEVERAGING CYC FOR IAA

Veridian Engineering

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

20020507 110

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2002-2 has been reviewed and is approved for publication.



APPROVED: PAUL M. ENGELHART
Project Engineer



FOR THE DIRECTOR: JOSEPH CAMERA, Chief
Information & Intelligence Exploitation Division
Information Directorate

If your address has changed or if you wish to be removed from the Air Force Research Laboratory Rome Research Site mailing list, or if the addressee is no longer employed by your organization, please notify AFRL/IFED, 32 Brooks Road, Rome, NY 13441-4114. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

| REPORT DOCUMENTATION PAGE | | | Form Approved OMB No. 0704-0188 |
|---|--|--|------------------------------------|
| <small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small> | | | |
| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE Jan 02 | 3. REPORT TYPE AND DATES COVERED Final Jan 00 - Jul 01 | |
| 4. TITLE AND SUBTITLE LEVERAGING CYC FOR IAA | | 5. FUNDING NUMBERS C - F30602-99-D-0050/Task 0004 PE - 62702F PR - 459E TA - TX WU - T1 | |
| 6. AUTHOR(S) Jeanette Neal, Benjamin Rode, Chris Crowner, Dave Gunning and John Guwa | | 8. PERFORMING ORGANIZATION REPORT NUMBER | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Prime: Veridian Engineering 4455 Genesee St PO Box 400 Buffalo NY 14225 Sub: CYCORP, Inc. 3721 Executive Center Drive Suite 100 Austin TX 78731 | | 10. SPONSORING/MONITORING AGENCY REPORT NUMBER AFRL-IF-RS-TR-2002-2 | |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFRL/IFED 32 Brooks Road Rome NY 13441-4114 | | 11. SUPPLEMENTARY NOTES AFRL Project Engineer: Paul M. Engelhart, IFED, 315-330-4477 | |
| 12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution unlimited. | | 12b. DISTRIBUTION CODE | |
| 13. ABSTRACT (Maximum 200 words) This final technical report summarizes research conducted by Veridian Engineering and Cycorp, Inc. to access the feasibility of leveraging the combined capabilities of the Intelligence Analyst Associate (IAA), an information extraction (IE) system, and the Cyc Knowledge Base (KB), a very large knowledge base, for monitoring domains of interest to intelligence analysts. This research examined mechanisms to help alleviate the textual data overload that intelligence analysts experience. The IAA system has capabilities for processing large volumes of unstructured text, extracting information relevant to intelligence analysts, such as entities and simple events, storing the extracted information in a structured database, and enabling the use of analysis and visualization tools. By utilizing the Cyc KB, more intelligent processing using the context of the documents and that of the analysts' persistent knowledge bases to automatically generate new information and knowledge was accomplished. The inheritance feature of the Cyc KB ontological hierarchies provides a strong benefit in the efficient representation of knowledge. It was also shown that many of the information extraction tasks have interdependencies and that there is an advantage to accomplishing them in a KB using inference. | | | |
| 14. SUBJECT TERMS Information extraction, knowledge base, ontological engineering, intelligence analysts, microtheory development | | 15. NUMBER OF PAGES 40 | 16. PRICE CODE |
| 17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED | 18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED | 19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED | 20. LIMITATION OF ABSTRACT UL |

LEVERAGING CYC FOR IAA

TABLE OF CONTENTS

| | | |
|-------|---|----|
| 1 | Introduction | 1 |
| 2 | Referenced Documents | 1 |
| 3 | Driving Problems and Project Goals | 2 |
| 3.1 | Problems Driving the Project Objectives | 2 |
| 3.2 | Project Goals | 2 |
| 3.3 | Project Scope | 4 |
| 4 | IAA and Cyc KB System Overviews | 5 |
| 4.1 | IAA Overview | 5 |
| 4.2 | Cyc KB Overview | 6 |
| 5 | Technical Approach Overview | 8 |
| 5.1 | Leveraging the Cyc KB | 9 |
| 5.2 | Building on IAA Capabilities | 10 |
| 6 | Summary of Accomplishments | 11 |
| 6.1 | IAA-Cyc IE Software Development | 11 |
| 6.2 | Cyc KB Ontological Engineering | 11 |
| 7 | Technical Approach and Accomplishments | 12 |
| 7.1 | IAA-Cyc IE Software Development | 12 |
| 7.1.1 | Information Identification - References | 12 |
| 7.1.2 | Information Normalization - References | 14 |
| 7.1.3 | Information Normalization - Coreferences | 14 |
| 7.1.4 | Semantic Interpretation – Entities | 15 |
| 7.1.5 | Semantic Interpretation – Disambiguation | 16 |
| 7.1.6 | Semantic Interpretation – Entity Attributes | 16 |
| 7.1.7 | Semantic Interpretation – Event Interpretation | 18 |
| 7.1.8 | Information Inferencing – Entity Attributes and Relations | 19 |
| 7.2 | Cyc KB Ontological Engineering | 20 |
| 7.2.1 | Design Legacy of the IAA-Cyc Project | 20 |
| 7.2.2 | General Microtheory Structure | 20 |
| 7.2.3 | Military Positions | 22 |
| 7.2.4 | Military Ranks | 24 |
| 7.2.5 | Facilities and Postings | 25 |
| 8 | Lessons Learned | 26 |
| 8.1 | IAA-Cyc IE Software Development - Lessons Learned | 26 |
| 8.2 | Cyc KB Ontological Engineering - Lessons Learned | 27 |
| 9 | Future Directions | 28 |
| 9.1 | IAA-Cyc IE Software Development | 28 |
| 9.2 | Cyc KB Ontological Engineering | 29 |
| 10 | List of Acronyms | 30 |

TABLE OF FIGURES

| | |
|--|----|
| Figure 1 The goal of IAA-Cyc is to process prose text to extract information and automatically populate information tables in an analyst's BOK..... | 3 |
| Figure 2 IAA-Cyc uses techniques that support the representation (modeling) of entities and their attributes as well as links (relationships) between entities | 4 |
| Figure 3 The high level design concept for the IAA-Cyc system..... | 8 |
| Figure 4 The IAA-Cyc Project leveraged the Cyc KB capabilities to the benefit of IAA and its analyst users | 10 |
| Figure 5 The Cyc KB General Microtheory Structure..... | 21 |
| Figure 6 A more detailed view of the PLA/PLAAF force structure microtheories | 22 |
| Figure 7 Part of the predicate hierarchy of functional roles created in parallel with the IAA-Cyc Project..... | 23 |

LEVERAGING CYC FOR IAA

Final Technical Report

1 Introduction

The objective of this project was to assess the feasibility of leveraging the combined capabilities of the Intelligence Analyst Associate (IAA), an information extraction (IE) system, and the Cyc Knowledge Base (KB), a very large knowledge base, for monitoring domains of interest to intelligence analysts.

The goal of IAA is to help alleviate the textual data overload that intelligence analysts experience. IAA has capabilities for processing large volumes of unstructured text, extracting information relevant to intelligence analysts, such as entities and simple events, storing the extracted information in a structured database, and enabling the use of analysis & visualization (A&V) tools.

However, IAA needs the ability to perform further, more intelligent processing, using the context of the documents and that of the analysts' persistent knowledge bases or "bodies of knowledge" (BOKs) to automatically generate new information/knowledge and add this new knowledge to the analysts' BOKs in their domains of interest.

This report comprises the Final Technical Report for the project focusing on leveraging the Cyc KB for the IAA system. Section 2 lists the referenced documents. Section 3 presents the driving problems and project goals, Section 4 provides a brief overview of IAA and the Cyc KB, Section 5 presents an overview of our technical approach, Section 6 summarizes the project accomplishments, Section 7 provides more detailed information on technical approach and accomplishments, Section 8 summarizes lessons learned, Section 9 presents future directions, and Section 10 provides a list of acronyms.

2 Referenced Documents

The following is a list of the relevant documents referenced within this Report.

1. Allen, Kenneth W., Krumel, Glenn, Pollack, Jonathan D., *China's Air Force Enters the 21st Century*, RAND, 1995.
2. Cycorp, Inc., Cycorp web site providing information on the Cyc Knowledge Base and other knowledge based products at: <http://www.cyc.com>
3. Mulvenon, James C., *Professionalization of the Senior Chinese Officer Corps*, RAND, 1997.

4. Veridian Engineering, *Intelligence Analyst Associate Software User Manual*, September 2000.
5. Veridian Engineering, Veridian Knowledge Management Internet site with information on analyst support tools/systems developed by Veridian such as IAA and the Document Content Analysis and Retrieval System (DCARS): <http://www.dcars.com>
6. Veridian Engineering, Intelligence Analyst Associate (IAA) Brochure at <http://www.dcars.com/infotech/html/products/IAA.html>

3 Driving Problems and Project Goals

3.1 Problems Driving the Project Objectives

The problems that drove the program objectives are based on discussions with analysts at the National Air Intelligence Center (NAIC) and the Joint Warfare Analysis Center (JWAC). The driving problems include:

- Analysts are plagued by information overload, especially the large volume of text documents and message traffic that they must examine in order to find and extract relevant information.
- Analysts cannot afford to miss information that impacts their analyses.
- Analysts need tools that focus on specialized information.
- Analysts require precise and reliable extracted information.
- Analysts have difficulty in converting and organizing extracted information into a form or tool that will support their analysis activities.
- Analysts do not have enough control over the information stored and manipulated by some of their tools/systems such as IAA.

3.2 Project Goals

The goals of the IAA-Cyc Project were to:

- Automatically populate analysts' bodies of knowledge (BOKs) or information level database tables from information extracted from text documents, especially unstructured prose text. Figure 1 below illustrates an example type of table that the IAA-Cyc software would be designed to fill. This table holds information on a PLA person.
- Take a domain-independent approach to information extraction to the extent possible.
- Focus on extracting information on persons, organizations, equipment, and facilities.
- Focus on candidate high priority software capabilities including:
 1. Identify missed persons, organizations, geopolitical entities.
 2. Normalize persons and organizations.
 3. Extract attributes and relations for identified entities.
 4. Infer attributes and relations for identified entities.
 5. Identify actors, actions, affected of events (for a small class of events).

| Person | | | |
|----------------------|------------------------|--|------------------------|
| HENGMEI HUANG | | | |
| # | Title | Job Position | |
| 1 | <i>Captain to Col.</i> | <i>Deputy Commander</i> | |
| 2 | <i>Captain to BGen</i> | ? | |
| 3 | <i>BGen to MGen</i> | <i>Commander</i> | |
| 4 | <i>MGen</i> | <i>Commander</i> | |
| 5 | <i>MGen</i> | <i>Deputy</i> | |
| 6 | <i>MGen</i> | <i>Commander</i> | |
| 7 | <i>MGen</i> | <i>Deputy Commander</i> | |
| <i>Continued</i> | | | |
| # | Main Org. | Administrative Unit | Address |
| 1 | <i>PLAAF</i> | <i>Air Group or Squadron</i> | ? <i>China</i> |
| 2 | <i>PLAAF</i> | <i>Suborg. of 7th Air Army</i> | <i>Guangxi, China</i> |
| 3 | <i>PLAAF</i> | <i>Command Post</i> | <i>Shanghai, China</i> |
| 4 | <i>PLAAF</i> | <i>Command Post</i> | <i>Shanghai, China</i> |
| 5 | <i>PLA</i> | <i>8th National People's Congress</i> | ? <i>China</i> |
| 6 | <i>PLAAF</i> | <i>Chengdu MR Air Force</i> | <i>Chengdu, China</i> |
| 7 | <i>PLA</i> | <i>Chengdu MR</i> | <i>Chengdu, China</i> |

Figure 1 The goal of IAA-Cyc is to process prose text to extract information and automatically populate information tables in an analyst's BOK

Figure 2 below illustrates the type of data structures used to hold information about persons, organizations, countries, and job positions. Analogous structures are used for other entity types. The data structures essentially consist of slot-value pairs. The slots may represent attributes such as name, type, and gender for which the filler would be a data type such as a text string or number. Additionally, some slots may represent attributes, such as affiliation, spouse, or residence, whose values are links or pointers to other data structures representing other persons, organizations, etc. These links represent relationships between the different entities. The figure illustrates a link representing a relationship between a person and his/her job position, and indirectly to the organization within which the job position exists.

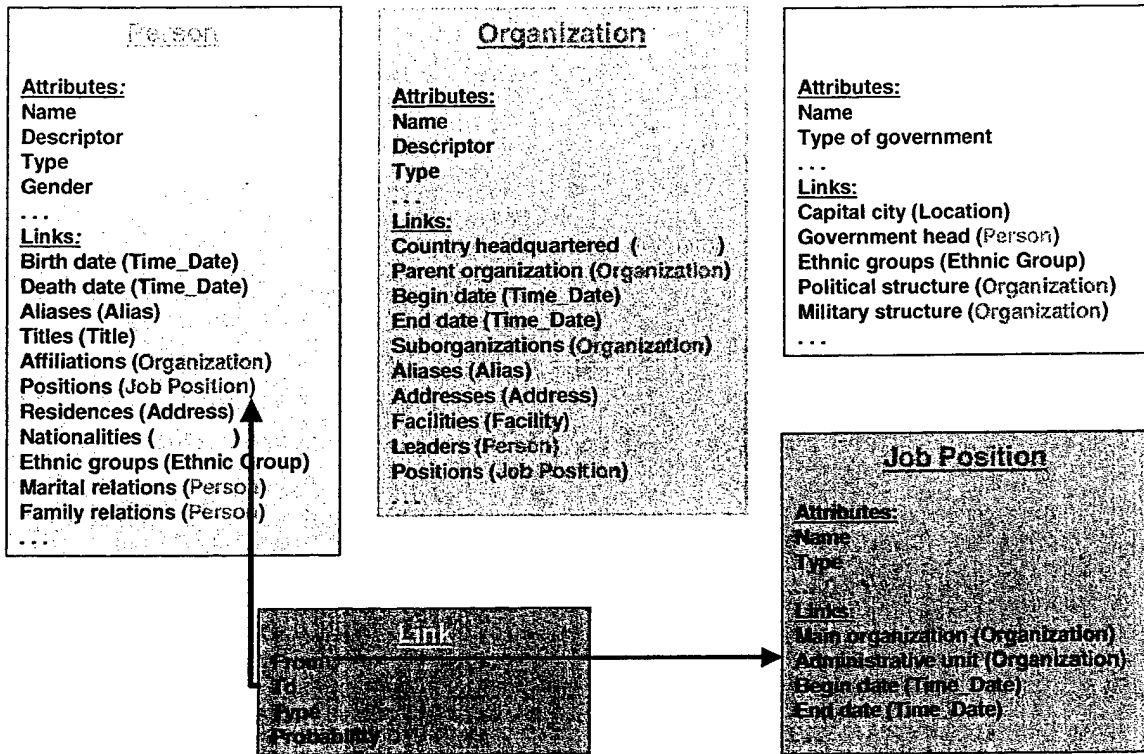


Figure 2 IAA-Cyc uses techniques that support the representation (modeling) of entities and their attributes as well as links (relationships) between entities

3.3 Project Scope

The following sources were studied for candidate attributes for this project for automated extraction and insertion into an analyst's BOK:

- The NAIC Dynamic Information Operations Decision Environment (DIODE) Model and Database.
- Sample documents provided by analysts.
- Cyc Knowledge Base (KB).

Based on this study and consultation with the Government, it was decided that the target attributes for automatic extraction for this project would be:

- Name (including aliases)
- Position (present and past)
- Military rank (present and past)
- Branch of military service
- Billet/military address

It was also agreed that the near term focus would use the Chinese military as the domain of interest. Sources of information on the Chinese military included:

1. Allen, K.W., Krumel, G., Pollack, J.D., *China's Air Force Enters the 21st Century*, RAND, 1995.
2. Mulvenon, J.C., *Professionalization of the Senior Chinese Officer Corps*, RAND, 1997.

4 IAA and Cyc KB System Overviews

4.1 IAA Overview

IAA performs extraction of entities and simple events from the high volume of available documents. IAA accepts ASCII documents from any source of text-based information: raw message traffic, reports, or open source text. IAA first applies a Text Zoner to locate the relevant parts of documents and messages and filter out the extraneous material from a document/message such as page breaks, headers, and footers. IAA then extracts the names of entities such as people, organizations, locations, dates and times. IAA also extracts shallow events in the form of subject, verb, direct and indirect objects. The extracted information is automatically loaded into a structured database for search and analysis.

In the A&V area, IAA provides a suite of eight (8) tools for analysts to use. These tools are:

- The Query Tool enables the user to create, edit, and execute queries that search the IAA database of information extracted from the documents/messages.
- The Statistics Tool enables the user to view information about the occurrence of single terms or phrases in a data set retrieved from the IAA database. The occurrence data is provided for each of the fields of the set of retrieved records.
- The Data Browser provides tabular visual displays of data sets retrieved from the IAA database. The Browser, for example, enables the user to view a dynamic table displaying the participants in simple events along with the location and date/time of the events, if available.
- The Document Browser enables the user to view and read the full text of any document in the IAA database, and view the location of the extracted information in the context of the full document/message.
- The Timeline Tool provides temporally-based visualizations of data sets retrieved from the IAA database. Each item in the data set (e.g., event) is represented by an icon on the timeline display with an associated descriptive text phrase and an associated horizontal bar that illustrates the duration or extent of the event or activity represented by the icon.
- The Geographic Display Tool provides geographical visualizations of data sets retrieved from the IAA database. The Geographic Display Tool displays icons for the data items on a map overlay display, placed appropriately to illustrate the location attribute of each item.
- The Topic Areas Tool enables analysts to save IAA database queries in a flexible and extensible hierarchical tree of folders that represent domains and topics of interest. Saved

queries and the folder hierarchy are represented graphically using icons. Queries may be moved, copied, renamed, edited, run and displayed from within the tool. In this way, the tool provides a centralized topical organization of an analyst's work in IAA.

- The IAA Concept Domain Tool enables the analyst to define conceptual domain areas for which he/she is responsible, define different forms of the questions that he/she is tasked to answer, and edit the concept domain information to develop it over time. The purpose of the Concept Domain Tool is to enable the analyst-user to more quickly find and discover information on topics of interest and to enable the analyst to better control the precision of his/her search.

For more information on IAA, visit the Veridian Knowledge Management Internet site containing information on analyst support tools/systems developed by Veridian such as IAA and the Document Content Analysis and Retrieval System (DCARS). The web site is at: <http://www.dcars.com>.

For more information on IAA in particular, visit the Intelligence Analyst Associate (IAA) Brochure at <http://www.dcars.com/infotech/html/products/IAA.html> at the Veridian web site.

4.2 Cyc KB Overview

The Cyc knowledge base (KB) is a formalized representation of a vast quantity of fundamental human knowledge: facts, rules of thumb, and heuristics for reasoning about the objects and events of everyday life. The medium of representation is the formal language CycL. The KB consists of terms, which constitute the vocabulary of CycL, and assertions which relate those terms. These assertions include both simple ground assertions and rules. Cyc is not a frame-based system. Instead, the Cyc team thinks of the KB as a sea of assertions, with each assertion being no more "about" one of the terms involved than another.

The Cyc KB is divided into many (currently hundreds of) "microtheories", each of which is essentially a bundle of assertions that share a common set of assumptions. Some microtheories are focused on a particular domain of knowledge, a particular level of detail, a particular interval in time, etc. The microtheory mechanism allows Cyc to independently maintain assertions which are *prima facie* contradictory, and enhances the performance of the Cyc system by focusing the inferencing process.

At the present time, the Cyc KB contains tens of thousands of terms and several dozen hand-entered assertions about or involving each term. New assertions are continually added to the KB by human knowledge enterers. The aforementioned numbers do not include (i) non-atomic terms such as predicates that express relationships between entities, nor (ii) the vast number of assertions added to the KB by Cyc itself as a product of the inferencing process.

The Cyc inference engine performs general logical deduction (including modus ponens, modus tolens, and universal and existential quantification), with AI's well-known named inference mechanisms (inheritance, automatic classification, etc.) as special cases. Cyc

performs best-first search over a proof-space using a set of proprietary heuristics, and uses microtheories to optimize inferencing by restricting search domains.

For more information on the Cyc Knowledge Base and other knowledge based products, visit the Cycorp web site at: <http://www.cyc.com>

5 Technical Approach Overview

The high level design concept for the IAA-Cyc system is illustrated in the figure below. The main processing steps include the following:

- Text Zoning is the identification of the various parts of a message or document (e.g., header, addressee list, source, title, body) as well as extraneous items such as page breaks, headers, and footers.
- Information Identification is the recognition of text segments comprising expressions for items such as entities, entity attributes, and simple events.
- Normalization means the conversion of the text expressions into standard expressions for the entities or concepts; normalization was applied to identified text segments expressing the entity names (e.g., "Senator Clinton," "Clinton," and "Hillary" would all be mapped into a standard name such as "Hillary R. Clinton").
- Semantic interpretation refers to the process of transforming text expressions into meaning representations.
- Information Inference refers to the process of inferring items of information from the text that were not expressed explicitly.
- The Loader is responsible for loading extracted information into the analyst's database.

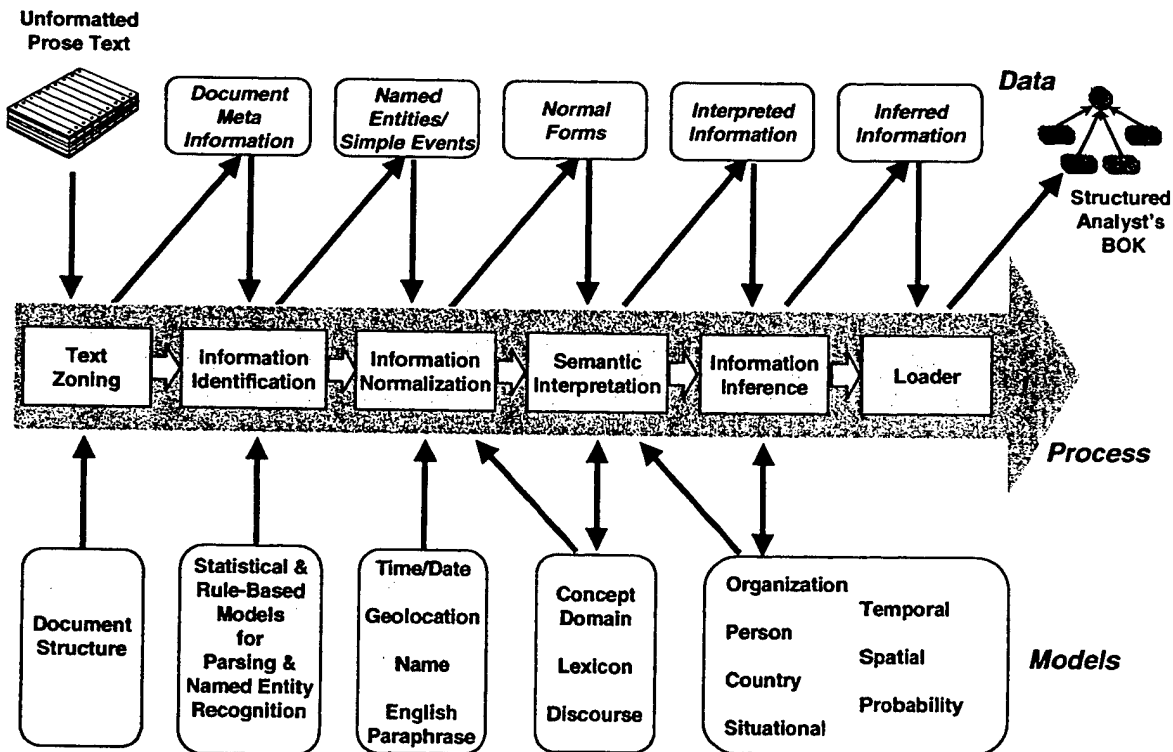


Figure 3 The high level design concept for the IAA-Cyc system

5.1 Leveraging the Cyc KB

The Cyc KB provides significant capabilities that can be leveraged to the benefit of IAA and its end users. The capabilities that were exploited in this project include:

- The ability to represent domain dependent facts in the Cyc KB to identify, classify and specify knowledge concerning relevant entities.
- The ability to represent rules in the KB and use the Cyc KB inference engine to allow information to be derived from identified entities and entity classifications.
- The ability to represent attributes of entities and their classifications.
- The ability to represent entity types and relations between the types.
- The ability to use microtheories for the representation of contexts.
- The ability to make use of ontological knowledge representation, permitting:
 - Different levels of generality in analysis allowing various degrees of domain independence to be maintained.
 - The ability to exploit inheritance and thereby gain benefits such as economy in the statement of rules.
 - Use of the existing wealth of knowledge previously developed and implemented in the Cyc KB, including both general common-sense knowledge and more domain-specific specialized knowledge in relevant areas.

The figure below illustrates some of the knowledge areas represented and used in the IAA-Cyc system. The figure indicates some of the ontologies used and the types of entities represented. These ontologies include military positions, ranks, and facilities. Links (relations) between the different entity types were also represented. Example relations include the relationship between a person and his/her position, as well as the relation between a person and the organization with which the person is affiliated.

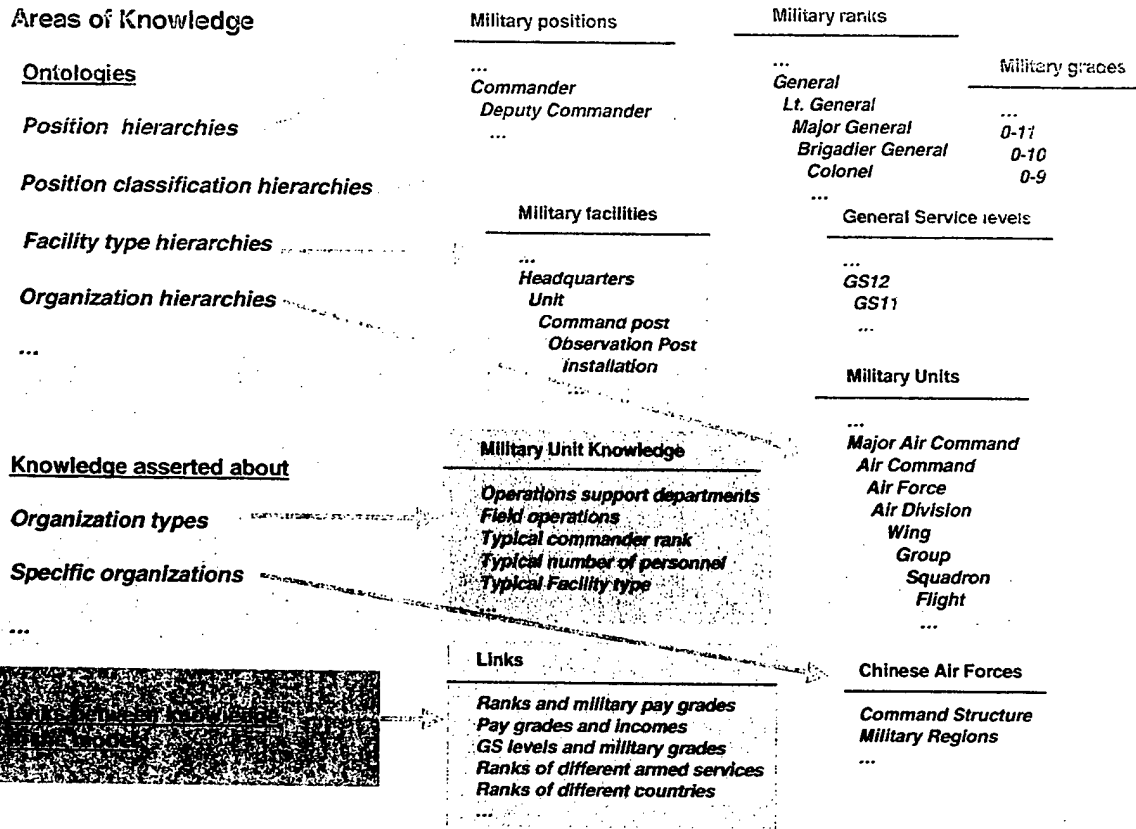


Figure 4 The IAA-Cyc Project leveraged the Cyc KB capabilities to the benefit of IAA and its analyst users

5.2 Building on IAA Capabilities

In this first phase research and development effort, the development of IAA-Cyc built on and utilized some of the text processing capabilities of IAA, such as the Text Zoner and entity identification capabilities. The Project also explored alternative technical approaches such as partial parsing rather than full parsing and new technology that will advance and extend the capabilities of IAA.

6 Summary of Accomplishments

6.1 IAA-Cyc IE Software Development

In the area of software development for information extraction, project accomplishments included:

- Automatic extraction of information concerning attributes and relationships about persons and organizations involving positions, units, ranks, postings and facilities.
- Noun phrase analysis to extract the above mentioned attributes and relationships.
- Coreference resolution and normalization of certain types of references (personal pronouns, proper names, and limited forms of descriptions).
- Analysis of clauses that express directly the above mentioned attributes. This clause analysis includes an identification and normalization of a time/date that the attribution is associated with.

Software components were developed to implement the following IE capabilities:

1. Information Identification - Reference
2. Information Normalization - Reference
3. Information Normalization - Coreference
4. Semantic Interpretation - Entity Interpretation
5. Semantic Interpretation - Disambiguation
6. Semantic Interpretation - Entity Attributes
7. Semantic Interpretation - Event Interpretation
8. Information Inferencing - Entity Attributes and Relations

For this IAA-Cyc effort, all these components were primarily implemented within the Cyc KB, along with an associated C program that performed some querying and processing.

6.2 Cyc KB Ontological Engineering

New ontological engineering (OE) work specific to the IAA-Cyc Project can be classified as falling under the following general headings:

- Military Positions
- Anticipated and actual tenures (in ranks and in positions)
- Rank comparatives
- Faceting functions for military organizations
- Organizational facilities and postings
- Rank-to-position mappings
- Command structure of the Chinese PLA and PLAAF

CycL specifications of the internal command hierarchy and military force structure of the PLA and PLAAF deserve mention as presenting special technical considerations. Specifically, our source documents contained distinct descriptions of the PLA command

structure at five distinct phases or levels of development: early history, 1947-1954, 1954-1970, 1970-1985, and 1985-present. Assertions that were true in one time frame were not necessarily true in any of the others.

Our solution to this problem was to sequester period-specific assertions into temporally indexed microtheories that were specialized microtheories of a general ChineseMilitaryForceStructureMt microtheory whose assertions were presumed to hold throughout the 'early history through 2001' time frame. Although this solution was acceptable for the purposes of the project, it should be noted that it proved possible to implement only because the developer could make a fairly hard-and-fast distinction between assertions that held true for the PLA/PLAAF command structures generally (throughout all the time periods referenced) and assertions that held true in exactly one of the specified time periods. Had it been the case that we had to deal with 'intermediate' assertions that covered proper, non-singleton subsets of the set of time frames (e.g., 1947-1985), it would have been necessary to reify a more complex partial order of microtheories.

7 Technical Approach and Accomplishments

7.1 IAA-Cyc IE Software Development

7.1.1 Information Identification - References

The software for the identification of references to information items, namely entities, addressed the following phenomena:

- Names
 - Multi-token names: "Military Region", "People's Liberation Army"
 - Prepositional compounds: "Secretary of State", "Commander of the PLAAF"
- Pronouns (with person, number and gender attributes)
- Descriptions
 - Definite: "the Guangzhou Military Region"
 - Indefinite: "a responsible officer"
- Lists
 - Qualification lists (subparts)
 - Organizations: "Political Department, PLAAF Headquarters"
 - Locations: "Paris, Texas"
 - Temporal: "May, 1999"
 - Conjunctions and uniform lists
 - "Korea, Japan", "Clinton, Barak, Arafat", "Clinton and Barak" etc.

- Appositives
 - Comma delimited: “PLAAF Commander, General Yu Zhenwu”
 - Not comma delimited: “PLAAF Commander General Yu Zhenwu”
- Parentheticals
 - Acronyms: “Military Region (MR)”

Where references are extracted from:

- References from identified noun groups (Preprocessing)
 - Verbs, conjunctions, prepositions and punctuation break groups
- References from identified named entities (IdentiFinder)
 - Persons, organizations, locations, times from IdentiFinder
- References from recognized proper names (KB)
 - In the KB
- References from interpretation of descriptions (IE)
 - Determination of how a word or phrase modifies another
 - Determination of how a word or phrase narrows or qualifies the meaning of an entity
- References serve as temporary constants
 - For making assertions concerning meaning
 - For making assertions concerning coreference

Example Input and Results

Input:

“MGEN XU CHENGDONG, DIRECTOR, POLITICAL DEPARTMENT (PD), PLAAF HEADQUARTERS. “

Results of the various stages of processing:

- *Tokenizer*: “MGEN” “XU” “CHENGDONG” “,” “DIRECTOR”, “POLITICAL” “DEPARTMENT” (“ “PD” “)” “,” “PLAAF” ...
- *Entity identification*: “MGEN XU CHENGDONG”
- *Partial parser*: “MGEN XU CHENGDONG” “DIRECTOR” “POLITICAL DEPARTMENT (PD)” “PLAAF HEADQUARTERS”
- *Recognized proper names*: “POLITICAL DEPARTMENT” and “PLAAF”
- *Recognized common nouns*: “DEPARTMENT” and “HEADQUARTERS”
- *Recognized qualification relations*: “PLAAF” qualifies “HEADQUARTERS” “POLITICAL DEPARTMENT” qualifies “PLAAF HEADQUARTERS”
- *Recognized modification relation*: “MGEN” modifies “XU CHENGDONG”

7.1.2 Information Normalization - References

By “normalization” we mean the conversion of a referring expression (a text string) into a standard representation.

Normalization of names includes recognizing parts of names

- Common name aliases; e.g., “Jim” for “James”
- Aliases with initials; e.g., “J. Homer” for “James Homer”
- Acronyms identified using parentheticals in documents; e.g., “Military Region” for MR

Normalization of Dates/Times includes recognizing the components of time expressions

- Months, days and years are composed into a normal form; e.g., “May 6, 1998” is translated into (DayFn 6 (MonthFn May (YearFn 1998)))

Entities are normalized by locating a constant or creating one in a normalized form

- A KB constant serves as a normal form; e.g., for “James Wu” the constant “WuJames-Person1”

7.1.3 Information Normalization - Coreferences

- Pronoun coreference
 - Person, number, gender attributes of references must match for a pronoun to corefer with another reference
 - Coreference can depend upon other attributes, e.g., “General Smith met with Prime Minister Major today. He told the Prime Minister”
- Names
 - Names that share normal forms corefer
 - Acronyms in parenthetical compounds coref to the preceding reference, e.g., “Political Department (PD)” PD corefers with Political Department
- Descriptions
 - Appositive parts corefer; e.g., “PLAAF Commander John Doe”
 - Two references corefer if they have common heads and are both in definite references; e.g., “the Anhui network”, “the network”
 - A reference to a position corefers with a reference that has been assigned the position attribute; e.g., “Commander” corefers to “Commander Wu”

Example Inputs and Results

Inputs:

“MGEN XU CHENGDONG, DIRECTOR, POLITICAL DEPARTMENT (PD), PLA AF HEADQUARTERS.

LITTLE IS KNOWN OF XU'S PAST.

HE WAS FIRST NOTED IN PRESS REPORTS IN MARCH 1992 AS A RESPONSIBLE OFFICER IN THE GUANGZHOU MILITARY REGION (MR).”

Results:

- The parenthetical “PD” corefers with “POLITICAL DEPARTMENT”.
- The named entity “XU” corefers with “XU CHENGDONG”.
- The pronoun “HE” corefers with the named entity “XU”.
- The parenthetical “MR” corefers with “MILITARY REGION”.

7.1.4 Semantic Interpretation – Entities

The IAA-Cyc software searches the KB for existing constants using the strings associated with the input linguistic elements (e.g., tokens and references):

- Case insensitive string matching within a specialized domain (“microtheory”).
- Normal forms are used to extend the search.

Semantic relations between recognized constants are identified to create new denotational terms:

- Denotational Functions: meaningful compound expressions used as arguments to predicates and functions.
- Qualification relations based upon the head noun in a noun phrase.
- Specialized forms such as temporal expressions.

New KB constants are created for references that cannot be associated with an existing constant or denotational term:

- Internal normal forms are created based upon classification of the references.
- Other areas of the KB may be searched for existing constants.

Example Inputs and Entity Interpreter Results

Inputs:

“MGEN XU CHENGDONG, DIRECTOR, POLITICAL DEPARTMENT (PD), PLA AF HEADQUARTERS.

HE WAS FIRST NOTED IN PRESS REPORTS IN MARCH 1992 AS A RESPONSIBLE OFFICER IN THE GUANGZHOU MILITARY REGION (MR).”

Results:

- XuChengdong-Person1 is derived as the interpretation for “XU CHENGDONG”
- PeoplesLiberationArmyAirForce-China is derived as the interpretation for “PLAAF”
- (HeadquartersFn PeoplesLiberationArmyAirForce-China) is derived as the interpretation for “PLAAF HEADQUARTERS”
- (PoliticalDepartmentFn (HeadquartersFn PeoplesLiberationArmyAirForce-China)) is the interpretation for “POLITICAL DEPARTMENT PLAAF HEADQUARTERS”
- XuChengdong-Person1 is derived as the interpretation for “HE”
- (MonthFn March (YearFn 1992)) is the interpretation for “MARCH 1992”
- (MilitaryRegionFn Guangzhou) is the interpretation for “GUANGZHOU MILITARY REGION”

7.1.5 Semantic Interpretation – Disambiguation

By “disambiguation” we mean the following: When a name, descriptive reference, or pronoun can refer to more than one item, disambiguation refers to the determination of the best choice for the item to which it refers (e.g., if both “Bill Clinton” and “Hillary Clinton” are mentioned, which one does just “Clinton” refer to?).

Because of limited resources, very little effort was applied to the issue of disambiguation to date.

The context surrounding a reference provides information that will be used for disambiguation:

- Coreference determination can narrow the possible meanings.
- Attributes may be used to select a meaning; e.g., select the person that has the military title that was recognized in the text.

The organization of knowledge into separate microtheories within the Cyc KB limited the amount of disambiguation that needed to be performed in our limited project to date. A preference for “closer” microtheory meanings may be implemented (not currently implemented)

7.1.6 Semantic Interpretation – Entity Attributes

Entity attributes:

- Are determined from the surrounding context of the reference:
 - Modifiers of the reference
 - Preceding/following references
- Are determined from information from the clause:
 - Clauses are classified according to their main verb
 - Currently, only main verbs that directly express attributes are handled

- The attributes of persons that are extracted:
 - A person's name and aliases
 - A person's position and their position in an organization
 - A person's title and/or military rank
 - A person's age
- The attributes of organizations that are extracted:
 - An organization's names and aliases
 - The positions within an organization
 - The location of an organization
 - The suborganizations of an organization

Example Inputs and Entity Attribution Results

Input: "MGEN XU CHENGDONG"

Derived result:

(rank-Military XuChengdong MajorGeneral)
 (hasTitle XuChengdong MajorGeneral)

Input: "MGEN XU CHENGDONG , DIRECTOR"

Derived result:

(hasPosition XuChengdong Director)

Input: "MGEN XU CHENGDONG , DIRECTOR, POLITICAL DEPARTMENT, PLAAF HEADQUARTERS"

Derived result:

(hasPosition XuChengdong (DirectorFn
 (PoliticalDepartmentFn
 (HeadquartersFn PeoplesLiberationArmyAirForce-China))))

Input: "MGEN XU CHENGDONG, DIRECTOR, POLITICAL DEPARTMENT"

Derived result:

(hasPositionIn XuChengdong Director PoliticalDepartment)

Input: "MGEN XU CHENGDONG, DIRECTOR, POLITICAL DEPARTMENT, PLAAF HEADQUARTERS"

Derived result:

(hasPositionIn XuChengdong Director (PoliticalDepartmentFn
 (HeadquartersFn PeoplesLiberationArmyAirForce-China))))

Input: "DIRECTOR, POLITICAL DEPARTMENT"

Derived result:

(positionInOrg Director PoliticalDepartment)

Input: "DIRECTOR, POLITICAL DEPARTMENT, PLAAF HEADQUARTERS"

Derived result:

(positionInOrg Director (PoliticalDepartmentFn
(HeadquartersFn PeoplesLiberationArmyAirForce-China)))

Input: "GUANGZHOU MILITARY REGION"

Derived result:

(orgInLocationIAA MilitaryRegion Guangzhou-China)

7.1.7 Semantic Interpretation – Event Interpretation

Interpretation of a restricted type of simple "event" was implemented to extract certain attributes of entities, with the entity occurring as the subject of the clause.

The interpretation makes use of the following components:

- Actor of clause – A person that is referred to by the clause's subject
- Main verb of clause – The last element of a clause's predicate (the predicate head)
- Time of clause – A time expression in a clause's verb modifier
- The verb modifier is restricted to having the preposition "in" as its first element

Example Inputs and Event Interpretation Results

Input:

"HE WAS FIRST NOTED IN PRESS REPORTS IN MARCH 1992 AS A RESPONSIBLE OFFICER IN THE GUANGZHOU MILITARY REGION (MR)."

Results:

- The actor of the clause is identified as "HE". Since "HE" corefers to "XU CHENG DONG" the actor of the clause has the denotation "XuChengdong-Person1"
- The main verb of the clause is derived as "NOTED". The semantic type of "NOTED" is ReportVerb.
- The verb modifier reference "A RESPONSIBLE OFFICER" has the semantic type PositionType derived for it.
- The time of the clause is derived as "MARCH 1992".
- The verb modifier reference "THE GUANGZHOU MILITARY REGION" has a semantic type Organization derived for it.

Input:

“HE WAS FIRST NOTED IN PRESS REPORTS IN MARCH 1992 AS A RESPONSIBLE OFFICER IN THE GUANGZHOU MILITARY REGION (MR). “

Results:

- The attribute extracted in a clause is qualified to hold within the time associated with the clause:

(holdsIn (MonthFn March (YearFn 1992))

(hasPositionIAA XuChengdong-Person1 Officer))

(holdsIn (MonthFn March (YearFn 1992))

(hasPositionIn XuChengdong-Person1

(OfficerFn

(MilitaryRegionFn Guangzhou)))

If there is no time associated with a clause, then the current discourse time may be used to qualify the attribution.

7.1.8 Information Inferencing – Entity Attributes and Relations

Additional attributes can be inferred from extracted attributes:

- What follows from having a certain position
- What follows from having a certain title or rank

Additional attributes can be inferred from extracted attributes using knowledge of the Chinese military. This Chinese military knowledge includes:

- Hierarchies of units, positions, and ranks
- Mapping between ranks and positions
- Typical tenure at a rank and position
- Mappings between positions and facilities

Examples of inferred attributes:

- If you have the position of a MilitaryOfficer, then your position is within a MilitaryOrganization.
- Having a certain rank implies a salutation. If the salutation is a MilitaryTile, then the person is a MilitaryPerson.
- Certain positions are associated with titles; e.g., heads of certain governments are presidents.
- Ranks are scaled so that it may be determined when someone outranks someone else.

7.2 Cyc KB Ontological Engineering

The technical approach to ontological engineering (OE) for the IAA-Cyc Project was a variant of the standard Cycorp OE approach, which incorporates the following stages in the order listed:

1. Close analysis of the target text by a member of the ontological engineering staff.
2. Preparation of English-0 paraphrase of the text content.
3. Translation of the English-0 paraphrase sentences into CycL assertions.
4. Revision and review of the CycL knowledge engineering (KE) files in consultation with Veridian personnel.
5. Loading of the final draft versions of the CycL files.
6. Application-testing of the new material in the Cyc Knowledge Base.

An important caveat is that our Cycorp developer found it eminently possible to dispense with the above Step 2 for most relevant text sections. KE files were subject both to in-house review at Cycorp and application-oriented vetting by designated personnel at Veridian.

Our document-driven knowledge entry also benefited from some higher-level ontological engineering programs being pursued concurrently at Cycorp, notably an ongoing OE effort to scope out and define the concept of a functional role, and a somewhat older effort to model requirements and expectations.

7.2.1 Design Legacy of the IAA-Cyc Project

The IAA-Cyc Project benefited primarily from three areas of prior work:

1. General work on modeling functional roles in the Rapid Knowledge Formation (RKF) project;
2. Legacy work from the Control of Agent-Based Systems project that modeled expectations in Cyc, and
3. Work on military ranks, echelons, command structures, and military specializations inherited from the HPKB Battlespace project.

The functional role modeling effort provided the groundwork for the OE defining military positions, and the work on expectation-modeling provided the basis for subsequent definition of the expectations associated with various positional roles, and also for the definition of vocabulary specifying “standard” associations between ranks and positions and expected tenures in positions and ranks.

7.2.2 General Microtheory Structure

Most of the high-level IAA-specific material was entered directly in the MilitaryForceStructureMt, which is currently defined to inherit from both the ReasoningWithExpectationsMt (expectation vocabulary) and the FunctionalRoleAnalysisMt (functional roles). Work on the PLA/PLAAF command structure was entered in a complex of

microtheories inheriting directly from the MilitaryForceStructureMt. These microtheories are illustrated in the figure below.

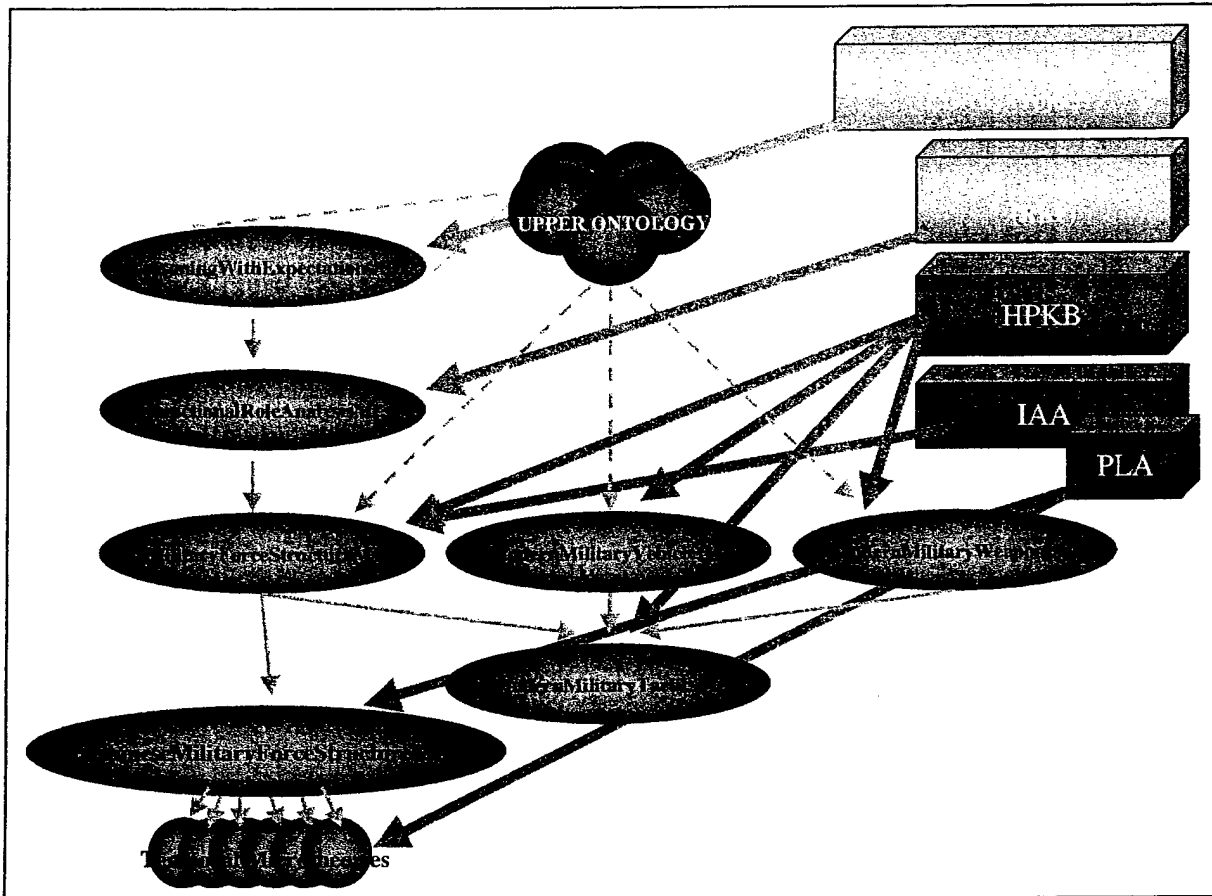


Figure 5 The Cyc KB General Microtheory Structure

The figure below shows a more detailed view of the PLA/PLAAF force structure microtheories. Note that each of the time frame microtheories inherits directly from the ChineseMilitaryForceStructureMt and from none of the others. Because all of the assertions that are true in a Cyc microtheory are inherited down a genlMt link, it would not do, e.g., to have the “early history mt” inherit to the “1947-1954” microtheory or vice versa: there are assertions which are true in either time frame that are not true in the other.

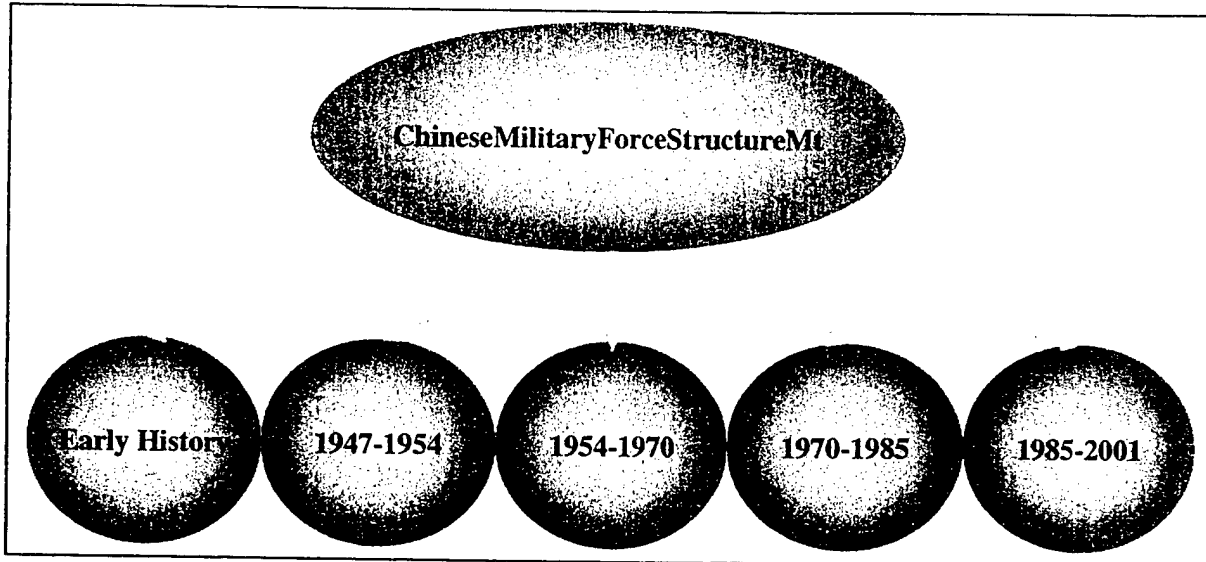


Figure 6 A more detailed view of the PLA/PLAAF force structure microtheories

7.2.3 Military Positions

Military positions were handled in terms of functional roles, which in turn were built on the model of the Cyc actorslot hierarchy. Actorslots are used to specify the role, which an existing thing plays in an event. Functional roles, in contrast, are used in specifying the role that a given agent plays in a functional system.

Part of the predicate hierarchy of functional roles created in parallel with the IAA-Cyc project is shown in the figure below. Indented relations are more specialized predicates of the relations under which they are listed. That is, their argument pairs inherit the more general relations. The most general relation in the scheme is componentInSystem-FunctionalRole shown at the top.

position, and the military organization in which the position is held, and the duration for which the position has been held.

Introduction of predicates for specifying both expected tenure and an individual's actual tenure in a position allows for *conformity checking*. Possible to formulate a CycL query in a given reasoning domain that will check for expectations that *can't* be proved, using pre-set inference parameters and constraints.

(and

(expected-ToBe ?PROP)

(unknownFormula ?PROP))

For example, suppose Huang Henmei held the position of deputy commander of the 29th Air Group for five years, and the expected tenure for a deputy commander of an Air Group in the Chinese Air Force was two years. The query shown above would detect this discrepancy.

An analogous approach was taken for ranks.

7.2.4 Military Ranks

Military ranks tend to be idiosyncratic with respect to country and branch-of-service. This leads to the natural question of what rank/position for a given country and branch-of-service is "comparable" to what other rank/position for some other country and/or branch of service. For example, what US Air Force rank is comparable to the U.S. Army rank of "Colonel"? What position in the Chinese Army corresponds to the North Korean position of "Commisar"? Of course, ranks or positions may be compared in several different ways.

An important respect in which ranks may be compared is in terms of what the holder of said ranks are *authorized* to do. This, in fact, was how rank and position comparisons were handled in this first phase of the IAA-Cyc Project. Accordingly, a rankAuthorizes predicate was introduced, in terms of which another predicate, comparableRanks, is defined. This latter is a quaternary predicate that relates two ranks and two classes of military organization as faceted by country and branch-of-service. The meaning of this predicate is that whatever roles a bearer of the first rank is authorized to play in the first class of organization, a bearer of the second rank is authorized to play in the second class of organization, and vice-versa. Note that this predicate's argument structure presupposes a means of faceting the class of military organizations by country and branch-of-service. This was another innovation introduced in the course of the IAA-Cyc Project.

As mentioned, predicates like comparableRanks required referencing military organizations by national allegiance and branch-of-service. This could have been done by simply having an argument position for allegiance and an argument position for branch-of-service. However, this would have meant comparableRanks would have to relate two ranks, two countries, and two branches of service. However, the preference at Cycorp is to not reify predicates with arity higher than five if it can be avoided. For this reason, and because it would be generally useful to reference military organizations by allegiance and branch of service, and by

allegiance, branch of service, and echelon, two faceting functions were introduced, `OrgTypeByGeoAndBOSFn` and `OrgTypeByEchelonGeoAndBOSFn`, together with two new type level collections to serve as their respective ranges: `MilitaryOrgTypeByGeoAndBOS` and `MilitaryOrgTypeByEchelonGeoAndBOS`. It is expected that they will be of great utility in the definition of new predicates and in parsing.

7.2.5 Facilities and Postings

The new work performed as part of the IAA-Cyc effort continued the longstanding conceptual distinction in CycL between an organization and the physical plant that it occupies. In the course of work on organizational facilities and postings, the following were introduced:

- predicates for relating units to their postings, both “generally” and within a specified time period,
- functions for reifying “the” posting of a given unit for a specified time,
- “expectancy” predicates for specifying the typical duration of posting for a given echelon, allegiance, and branch-of-service, and
- predicates for relating organization type to the types of facilities occupied.

Important considerations in this work involved taking into account the fact that:

- some “occupancies” are essentially one-to-one (one agent -- one facility), while others are many-to-one, one-to-many, and many-to-many,
- most occupancies have a definite duration, and
- in some cases we want to refer to occupancies that are “transitive with respect to super-regions”.

The purpose of the predicate `posting-Military` is to relate a military unit to the location where it is explicitly “posted”. Two important things to recognize about this predicate are the fact that it is functional in both arguments (i.e., it is a one-one relation), and the fact that it is not transitive with respect to super-regions. That is, even though a particular region is specified to be the `posting-Military` of a particular unit, super-regions of which that region is a part will not be inferred to be the `posting-Military` of the unit.

`MilitaryPostingFn` is the functional analog of `posting-Military`: it can be used to return “the” `posting-Military` of a particular military unit. Both this function and the corresponding predicate are parameterized to temporal context, insofar as the posting of a particular unit may change with time.

The predicate `postingForTemp-Military` is the predicate analog of `posting-Military` that is not parameterized to temporal context: i.e., the temporal thing that it takes in one of its argument positions is used to explicitly reference the time frame of the posting.

The function `MilitaryPostingForTempFn` is the function analog of `MilitaryPostingFn` that is not parameterized to temporal context—again, because it has an argument position that is used to explicitly reference the time frame.

The predicate posting-Military-Generic is like posting-Military except that it is transitive with respect to super-regions. For example, if a unit has San Antonio as its posting-Military-Generic, then it also has Texas as the posting-Military-Generic, also the United States, etc. The name of the predicate derives from the fact that the predicate is deemed to be in some sense looser than posting-Military.

In mapping positions to ranks, an "expectancy" predicate is used to specify the expected rank for a given position in a particular echelon, country, and branch-of-service. The definitional rule is cast using expected-ToBe, so that the relation could be used for conformity checking.

8 Lessons Learned

8.1 IAA-Cyc IE Software Development - Lessons Learned

The inheritance feature of the Cyc KB ontological hierarchies provides a strong benefit in the efficient representation of knowledge. Rules can be associated with a class at a relatively high level of the ontology and then inherited and applied to instances of its subclasses. This eliminates the need to have a separate version of the rule associated with each of the classes. For example, a rule associated with the Title class concerning the use of a person's title can be inherited and applied to more specific kinds of titles such as military titles or government titles. An example of such a rule might be something to the effect that if a person has a certain title such as "Senator", then the person could be referred to simply using the title in a definite noun phrase such as "the Senator".

A major drawback of the developed IAA-Cyc IE software is that it is slow. The software is too slow to realistically process an input document (6-12 minutes for 4 typical sentences). The reasons for this slowness include the proliferation of new Cyc KB constants that are created in the course of processing a clause. All these constants are considered as possible bindings to the variables in the antecedents of forward chaining rules, although only the constants created in the clause being processed can possibly meet the conditions in the antecedents of many rules in the Cyc KB. The possibility of applying heuristic procedures that would prevent the binding of out-of-scope constants to rule variables is being considered for future investigation.

Many of the information extraction tasks have interdependencies and there is an advantage to accomplishing them in a KB using inference. These IE subtasks include resolving coreferences, identifying modification and qualification relations expressed in noun phrases, identifying appositives in noun phrases, identifying entity attributions, and identifying semantic roles for events. Use of the KB does not require fixed sequential control of the processing for each task (in forward chaining rules inferences are triggered whenever antecedent information becomes available), so interdependencies between the inputs and outputs of each task are easily accommodated.

However, we also learned that it is inefficient to use the Cyc KB inference capability for certain processing tasks such as tasks that involved pattern matching for the purpose of

identifying the boundaries of references and the determination of normal forms for names of persons mentioned in a document.

8.2 Cyc KB Ontological Engineering - Lessons Learned

One aspect of the IAA-Cyc ontological engineering (OE) work that became apparent in retrospect, was that the representation of military positions was insufficiently fine-grained with respect to distinguishing between “officially held” positions and “actual” positions. Whether this necessitates integration with a wide-scale OE treatment of “actual” facts and “apparent” facts in Cyc needs to be considered.

Attention may also need to be paid to the orthogonal issue of whether “positions” in general are best treated as relations (roles) or as collections for purposes of parsing. This may relate to a general pattern becoming apparent in RKF, where concepts that are best treated as relations for purposes of inference and analysis are better regarded as collections for purposes of natural language generation and parsing. The RKF project is exploring several potential solutions to this difficulty which might plausibly be leveraged for IAA.

We may also need to revisit the issue of temporally indexed microtheories for PLA/PLAAF force structure representations. Although the current coarse-grained scheme is adequate for current purposes, it may be necessary to move to a finer grained scheme capable of accommodating assertions that overlap more than one, but not all, of the specified time frames.

9 Future Directions

9.1 IAA-Cyc IE Software Development

With respect to future development and the proper approach to take in the development of IAA-Cyc IE software, the following points apply:

- A two-tier approach should be taken which addresses:
 - Short-term practical goals, and
 - Longer term goals, keeping “aspirations high.”
- In general, an incremental approach should be taken where more involved tasks (and more sophisticated approaches to tasks) will not be attempted until it is determined that performance is sufficient on the basic tasks of most value to the analyst.

A suggestion was made that the follow-on IAA-Cyc II development could benefit from an examination of an analyst’s final reports. This examination might help in determining what information is important and useful to the analyst.

As a result of discussions and decisions made at the final Technical Interchange Meeting (TIM), the tasks listed below were agreed upon for the near-term follow-on effort:

1. Generate and fill person templates (structures or records) from document text.

These templates will have slots for the following attributes:

- Name (normal form)
- Aliases
- Titles
- Military rank
- Current position (with organization has position in)
- Date assumed position
- Past positions (with start and end dates and organizations)

2. Develop a graphical user interface so that a user can review and edit the person templates.

3. Develop KB rules that will:

- Determine the consistency of the information in the template slots.
- Infer the filler information for those slots that are empty (“knowledge gaps”).
- Flag anomalies (i.e., information that is unexpected) such as an unusually fast promotion of a person (a “shooting star”).

Note: Initially, confidence levels for the slot filler information will not be an R&D focus.

The production of person templates will require more concentrated approaches to the following problems:

1. Consolidation of entities: By "consolidation of entities" we mean the ability to determine when two names refer to the same entity.

This problem can be further broken down into:

- Name aliasing
- Pronoun coreference resolution
- Coreference of descriptions of persons

2. Disambiguation of entities: When a name, descriptive reference, or pronoun can refer to more than one entity, disambiguation refers to the determination of the best choice for the entity to which it refers.

3. Association of a date/time interval with any information regarding a person's position. The date/time would indicate the time period during which the information is believed to be true. It needs to be decided how this association should be handled. For example, should a date always be associated with position information? How much should document context be used to determine the date? Should KB rules be used to generate expected dates based upon past positions and their dates?

Approaches to these problems must be both efficient and accurate. Evaluation and extension of approaches developed as part of this first phase IAA-Cyc Project are necessary.

9.2 Cyc KB Ontological Engineering

Regarding near-term extension of the first phase Cyc IAA vocabulary and reasoning capabilities, the following are indicated as action items:

1. Devise use cases for testing "expectancy" predicates in various forms of conformity checking in order to further extend their utility in this regard.
2. Develop a suite of predicates for stating requirements and necessitating conditions, analogous to suite of "expectancy" preds already extant. More specifically, these predicates would be used for expressing conditions that must or must not exist in a given context. E.g., we might plausibly consider introducing a predicate, one of whose uses might be to state that two positions could not be held within the scope of a specified time frame. Such predicates would be reasonable candidates for integration with work on modal operators currently being undertaken at Cycorp, although their primary purpose for IAA would be for use in anomaly detection, and not in reasoning about necessity and possibility.
3. Revise military position representation to incorporate the concept of "official" vs. "actual" positions.

4. Assuming it exists, determine best solutions to putative “natural language versus ontological engineering” conflict over representation of positions as role relations, with reference to similar solutions adopted by the RKF dialog group.
5. Review temporal microtheories for PLA and PLAAF force structures to determine whether it is feasible to implement a finer-grained system for sequestering temporally qualified assertions. Also, consider integration with the on-going work being entered by the Cycorp Temporal Reasoning Special Interest Group.

10 List of Acronyms

| | |
|-------|---|
| AFRL | Air Force Research Laboratory |
| API | Application Programmer Interface |
| ASCII | American Standard Code for Information Interchange |
| BOK | Body of Knowledge |
| DIODE | Dynamic Information Operations Decision Environment |
| GAF | Ground Atomic Formula |
| HPKB | High Performance Knowledge Bases |
| IAA | Intelligence Analyst Associate |
| IE | Information Extraction |
| KB | Knowledge Base |
| KE | Knowledge Engineering |
| OE | Ontological Engineering |
| PLA | People’s Liberation Army |
| PLAAF | People’s Liberation Army Air Force |
| RKF | Rapid Knowledge Formation |
| RPC | Remote Procedure Call |

**MISSION
OF
AFRL/INFORMATION DIRECTORATE (IF)**

The advancement and application of Information Systems Science and Technology to meet Air Force unique requirements for Information Dominance and its transition to aerospace systems to meet Air Force needs.