Knowledge Base Formation Using Integrated Complex Information

Douglas S. Lange
SPAWARSYSCEN
53560 Hull Street
San Diego, CA 92152-5001
dlange@spawar.navy.mil

Abstract

An intelligence support system has been developed using an open hypermedia architecture. This approach provides a means to integrate information from distributed disparate sources into a knowledge base. Access by external applications is supported through a public interface. Filtering and change detection functions have also been implemented. The approach has shown promise in multiple domains indicating that it may be widely applicable. This paper discusses the principles of the hypermedia framework used for this system and how these principles may influence C4ISR systems in general.

1 Introduction

Command and control involves three fundamental processes that fit together in a tight cycle. Situation analysis provides the context on which to act. Decisions are made based on the analysis results. These decisions constitute planned movements, engagement orders, and many other possible actions. Decisions must be communicated to those who are to carry out the actions. The results of these actions are observed as part of a new situation analysis.

As Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) systems have evolved, system integration has been the general theme. Stand-alone systems, each with its own database, were first interfaced to allow some transfer of data. Data management schemes were implemented to provide some consistency among databases and operational units. System federation gradually allowed multiple applications to run on users' workstations, preventing the need for specialized hardware and support software for large numbers of individual systems. The current state of system integration allows not only allows multiple applications to share hardware, operating system, and network platforms, but also utilizes a layered service architecture to eliminate redundancy of some capabilities.

The evolution of system integration has broadened the stovepipes that were so narrow in previous system generations. The resulting view is of a few very broad systems made up of many small applications, any of which, may be accessible through the workstation in front of the user. Some applications work on common data managed through centralized services. However, many categories of data still form separate stovepipes as they are maintained in separate data repositories due to their differing technical natures and programmatic backgrounds. It is left to mental effort by the users to associate the tactical situation shown in one application with the results of a logistical query conducted through another.

1.1 Information Complexity

The focus on systems integration ignores the true goal in decision support. It is the information that is of ultimate value to the decision-makers. Integrating the information is the next step. However, military information is not a simple matter of collecting and crunching sales and inventory figures from various branch offices as found in data warehousing applications. The domain of the military environment is complex. The variety of concepts, events, and situations that can be described subjectively or measured and reported objectively is probably limitless. No ontological study can *a priori* determine all of the possible data types needed to describe the military environment. Therefore, information integration is not going to be completely accomplished through bringing all data into a relational or object database.

1.2 A Pattern of Analysis

In researching the requirements for an intelligence support system for the United States Defense Intelligence Agency (DIA), a pattern of analysis was uncovered that was common to those used in some other domains. The primary feature of this pattern is that an analyst's role is to create associations among existing data. An analyst rarely creates data, but searches, filters, and reviews all available information. As they do, they form networks of related information. Lange, 1999]

Current practice involves DIA intelligence analysts spending a portion of their time building up a private model of their area of expertise. The remainder of their time is spent responding to queries from DIA's various customers. The responses typically take the form of linear essays. Analysts also periodically produce background reports on particular matters of interest. These too take a strictly linear book-like form even when delivered over a computer network.

Analysis of the current approach yielded the following problems:

- The products were static or updated on a paper publishing timeframe.
- There is no mechanism for customers with local information to share it with others.
- Only a particular question gets answered, even if it was not the correct question.
- Analyst turnover causes a large loss of knowledge.

As a result of these insights, work was initiated to find a way of recording the knowledge being built by the intelligence analyst and communicating this knowledge to intelligence consumers. The goal was to move away from the linear essay and its strict segregation of reader and writer roles to a more collaborative method of communications that would allow for continuous update of the knowledge jointly held between the intelligence agency and its customers.

1.3 Recording Decisions

Decisions also take the form of associations among data or information elements. A classic example may be the order for a surface combatant to engage a hostile aircraft. The decision-maker did not create the aircraft or the positional and attribute data known about that aircraft. Likewise, the decision-maker did not generate the information related to the surface combatant. The value added by the decision-maker is that an engagement relationship (perhaps with other amplifying information) should exist between the two.

As the data on the two combatants changes, the association must be reviewed but is not necessarily invalidated. Likewise, a reversal of the decision changes the relationship among the combatants but does not change any of their individual data. This fundamental distinction between the structural representation of the associations among concepts or real-world objects and the content that describes them is common between the knowledge created by analysts and decision-makers.

2 Hypermedia Architectures

One mechanism for automating the management of information that is structured in the manner described above is through hypermedia systems. Such systems provide an ability to work with a wide variety of data while utilizing the powerful information present through the structures created by the connections made among the various data items [Nurnberg, et. al., 1997]. Despite accurately recording information, the non-linearity of hypermedia allows the reader to access information in ways not necessarily anticipated by the author. In this way, users of analysis results can make new discoveries from the same body of data [Nielson, 1990]. Likewise, distribution of responsibilities in a large command and control environment is aided by ensuring that not all uses of the data must be preconceived though accurate representation of constraints is essential.

The basic features of most hypermedia systems are as follows:

- *Node*. A node is an object that represents a document or some other media element.
- *Link*. Links are used to create relationships among nodes.
- *Anchor*. Anchors connect nodes to the actual media that make up their content.

2.1 Open Hypermedia

From 1987 to 1991 researchers noted that the hypermedia systems of the time did not support the needs of collaborative work groups and that they could not be integrated into computing environments being used in large enterprises [Halasz, 1987][Malcolm, et al., 1991]. Requirements were found for hypermedia systems that were not being addressed. These included:

- Interoperability to access and link information across arbitrary platforms, applications, and data sources.

- Link and node attributes to record the author of a link, what the permissions are for the particular link or node and other management information.
- Anchors that allow attachment to the exact data desired.
- Link types to provide more information about the meaning of a particular link and what functions the link is intended to support.
- Public and private links to support collaborative environments.
- Templates for automating routine analysis tasks.
- Navigational aids that can act as filters and supply powerful querying mechanisms.
- Configuration control so that information of importance in an analysis effort can be developed and managed in hypertext.

To address these requirements, open hypermedia systems evolved. Open hypermedia systems have been defined as those that exhibit the following characteristics [Davis, et. al., 1992]:

- A system that does not impose any markup on the data. By marking up data in order to create hyperlinks, the data is changed making it inaccessible to systems that cannot handle the markup.
- A system that can be integrated with any tool that runs under the host operating system. This can be extended to mean a system that can be integrated with distributed object environments.
- A system in which data and processes may be distributed across a network, and across hardware platforms.
- A system in which there is no artificial distinction between readers and authors. This is quite important for systems supporting analysis.
- A system to which new functionality can be easily added.

Since analysts and decision-makers are simultaneously readers and authors of node contents and links, these characteristics are vital in an information support environment. Likewise, the ability to link objects without changing them is critical. The information being linked together by the analysts may be coming from other applications and databases with which the hypermedia system has been integrated. These applications will not understand changes imposed on the data in order to support linking. The links must be separated from the content. This is the basic premise of open hypermedia system and has been demonstrated in many research systems [Will, 1997].

The prototypical open hypermedia system is structured according to figure 1 below.

2.2 Graph-Based Hypermedia

Several other hypermedia system types contribute capabilities necessary to support analysis functions. Chief among these is graph-based hypermedia. Graph-based hypermedia is based on set and graph theory, providing mathematically defined filter, search, and navigation methods. This category of hypermedia also includes huma-computer interaction methods featuring graphical depictions of the hypermedia.

The idea of a schema made of node and link types provides the basis for much of this method's power. [Lucarella and Zanzi, 1996] The relationships among schema types and between schema entries and the instances created in the hypermedia mirror the relationships in object-oriented design closely.

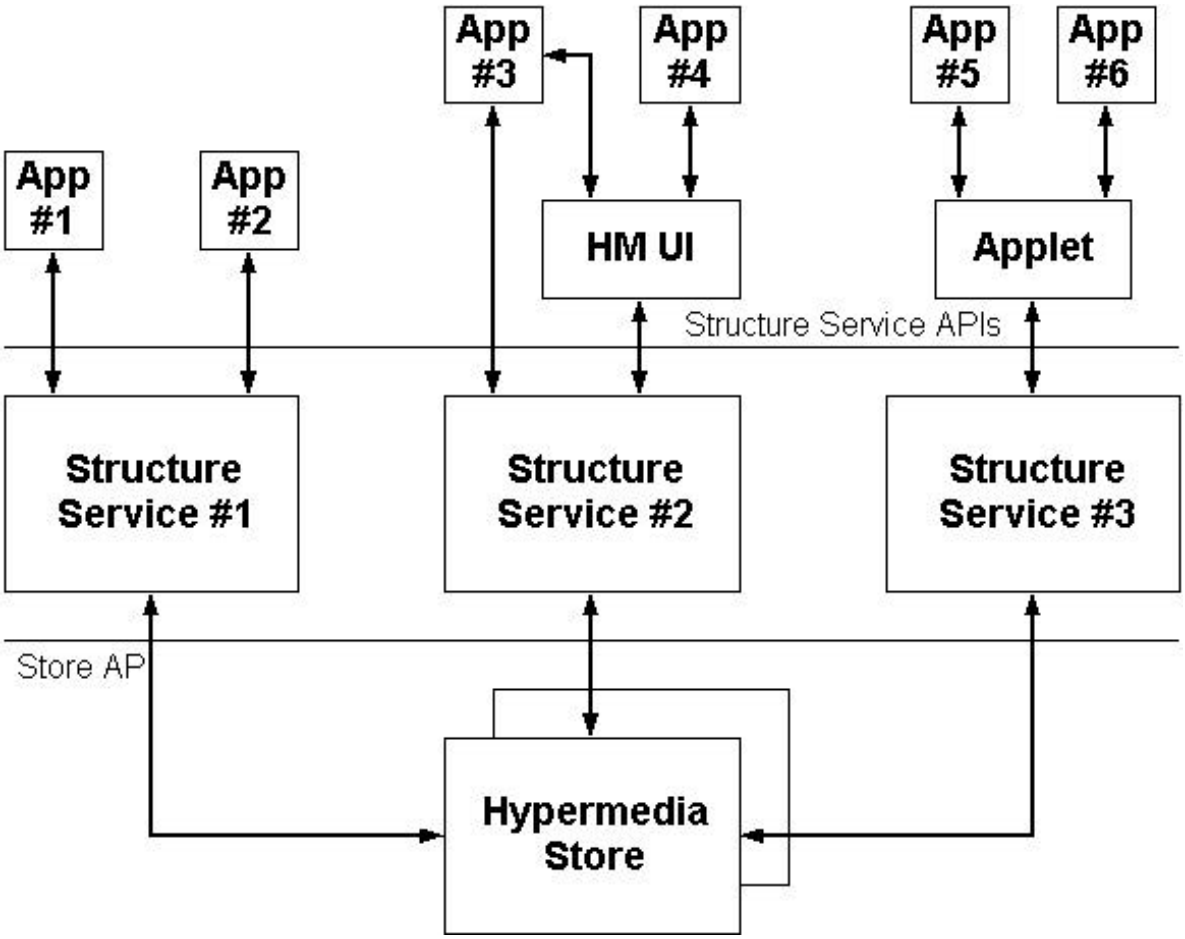

Figure 1. Open Hypermedia Architecture [Wiil and Nurnberg, 1999]

One result of the typing found in graph-based hypermedia systems is that the resulting hypermedia forms a semantic network. Semantic networks are used to model concepts and real-world situations, making them a natural tool for modeling a tactical situation or the results of intelligence analysis.

Another result is that sophisticated filtering mechanisms can be defined. Graph-based hypermedia provides the concept of a *perspective*. A perspective contains three elements. First is the perspective patter. A perspective pattern is a hypergraph that is a subset of the schema hypergraph. The second element is a filter, which is a constraint on the instance set. The filter may

constrain either through the node and link attributes, or the content attached through the anchors. Finally there is a subset of the instance set that satisfies both of the constraints.

3 HyperObject Processing System

The design of the HyperObject Processing System (HOPS) inherits features from both open hypermedia systems and graph-based hypermedia systems. Some modification to the established research architectures was required in order to support analysis of the kind performed by DIA. These same modifications would appear to be important for related C4ISR systems.

3.1 General Architecture

HOPS follows the open hypermedia form with an architecture shown in figure 2 below.

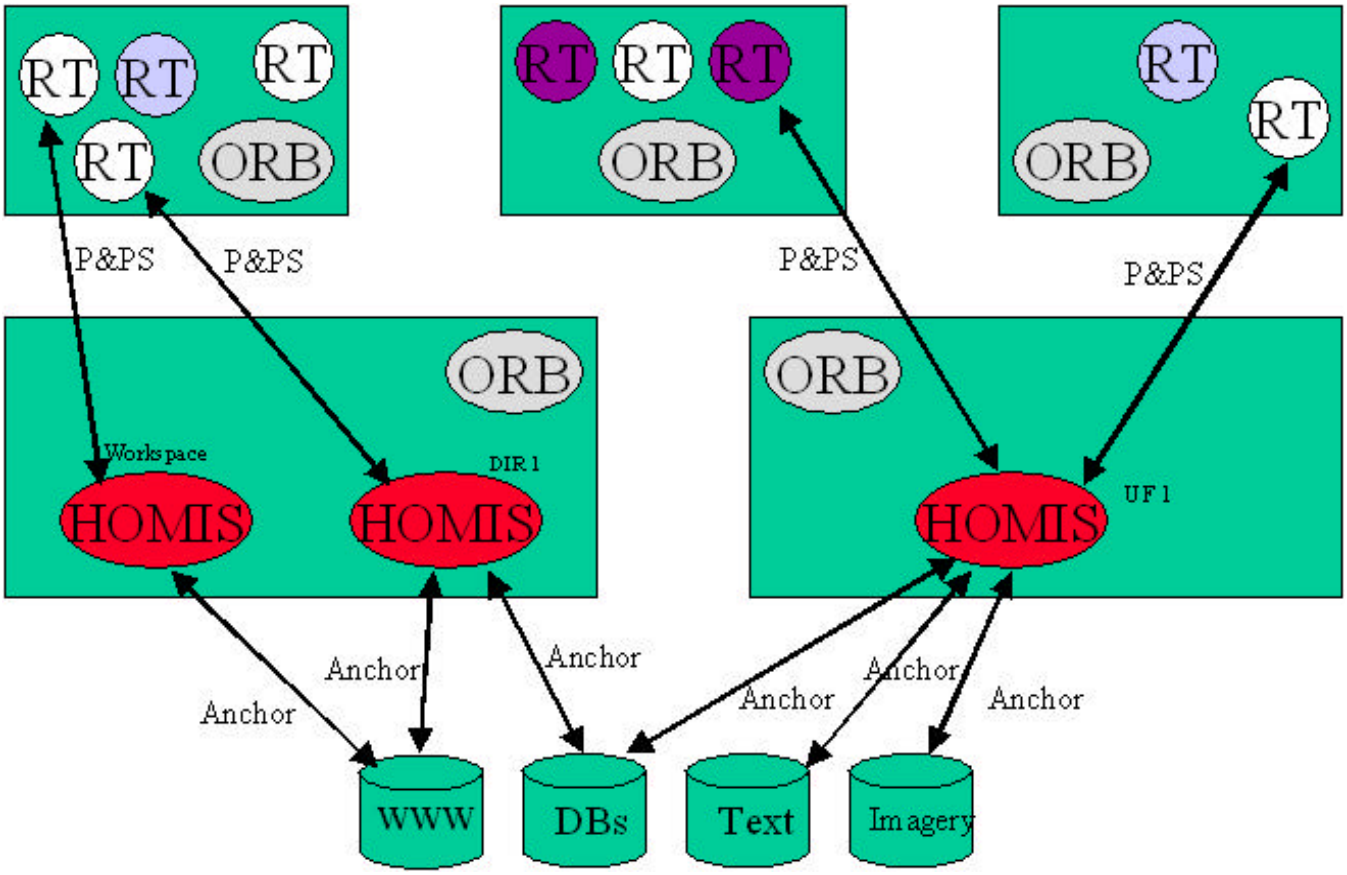

Figure 2. HyperObject Processing System

In the figure above, circles labeled 'RT' represent run-time applications either supporting a user directly or automated processing. 'HOMIS' are HyperObject Multimedia Information Systems, which are modified Multimedia Information Systems [Lucarella and Zanzi, 1996] able to handle

hypergraphs rather than simple graphs [Lange, 1997]. HOMIS function as structure servers as called for in the generic open hypermedia systems. However, they provide graph-based hypermedia functions. Each HOMIS has a schema and instance set, perspectives (‘P’ in figure 2) and filters can be defined, and graph-based navigation interactions are possible. 'ORB' represents an object request broker, in our case supporting Java Remote Method Invocation. These allow the system to be distributed over multiple platforms.

3.2 Unique Hypermedia Features

Most hypermedia systems found in the research literature work with information spaces constrained by the level of diversity and quantity of the information, restrictions on the structure of information, or by limited change of the underlying data. Several aspects of HOPS are unique among hypermedia systems. The features have been found to be necessary to allow HOPS to handle the dynamic unbounded nature of military information integration.

3.2.1 Multiple Anchors.

The middle layer of HOPS holds the semantic network. Classical hypermedia systems use a node to represent a piece of media and anchor to a single media element to provide content. This forms a semantic network describing the relationships among media elements rather than describing a tactical situation. To remedy this, HOPS uses multiple anchors per atomic node. This allows the nodes to define concepts or real-world objects and the links to represent relationships among them rather than relationships among the content elements.

3.2.2 Large Open-Ended Schema

Schemas imply an ability to predict all of the types of information to be used and the entire range of associations that will exist among the elements. In some domains this is possible, but not in the military information domain [Lange, 1999]. An example can be demonstrated in terms of exercise plans. During Tandem Thrust 97 (TT97), one of the primary requirements concerned protecting the Great Barrier Reef. Environmental mitigation strategies and environmental reports are not typically found in the command and control systems of our armed forces. There will always be unpredicted situations in warfare and military exercises. Information systems must be able to adapt on the fly to allow analysts and decision makers to see and interpret information and record and inform of decisions. The HOPS design allows users to include information not accounted for in the schema through the object-oriented method of deriving all nodes and links from common ancestors. This allows users to bypass rules in the schema and connect nodes and links in ways not previously predicted. The user or an administrator can then update the schema, on-the-fly, to allow autonomous tools to process the information more easily.

Analysis schemas and instance sets can become quite large. The problems being modeled are quite complex. The size of the schema represents the complexity of the model, while the size of the

instance set represents the quantity of information. Consumers of the analysis model must be able to filter both in terms of complexity and in terms of size of the knowledge base that they work with in order to avoid being overwhelmed. HOPS allows this capability through adaptations of the graph-based hypermedia concepts of perspective patterns and filters [Lange, 1997]. Perspective patterns allow the user to limit the kinds of information being worked with, while filters focus attention on information with particular content.

3.2.3 Link and Anchor Integrity

When important decisions are being made based on the information presented, error is less tolerable than in our daily workings with the World Wide Web. Anchored content must not disappear unexpectedly. Likewise if content changes, then the model must be re-evaluated to determine if it is still valid. The typed links of the storage layer must also be carefully managed to prevent dangling links. HOPS accomplishes these goals by caching anchored content, and providing periodic checks using an autonomous change detection agent. Agents used for this purpose can use whatever rules suit the application.

3.2.4 Link Equality

Although hypermedia relies on associations between elements for its character, many of the interaction techniques found in research literature are still focused on the content (e.g., string matching filters and searches, searches on images). Links are primarily used for navigation. This may be because in many applications, links are addresses, used to point to more information or typed paths to get to related nodes. Since the primary value added by intelligence analysts and decisions makers is found in the associations among elements, authors and readers of the products will want the ability to interact with typed links in ways other than simply using them navigation. They themselves convey critical information. HOPS handles this by making links special types of nodes. This allows all of the mathematics of filtering, searching, and browsing to work on links.[Lange, 1997].

3.3 Framework

HOPS itself is not command and control system or an analysis system. HOPS is a hypermedia framework designed to support analysis and provides some generic applications for interacting with the hypermedia. The way that HOPS is intended to be used is by adding domain specific applications along with an initial schema to create an analysis system of the type needed. This has been done in support of DIA’s mission.

In the Military Operations in Built-up Areas project, HOPS was integrated with the Lightweight Extensible Information Framework (LEIF) to provide geographic and temporal views of the hypermedia. An intelligence product creation wizard and intelligence specific anchors were also

used. Together with the generic applications that exist within the framework, users have a variety of ways of working with the information.

4 Prospects for Information Integration

Hypermedia systems hold promise for information integration. Any number of decision support tools can access the semantic network formed of the associations and nodes. Since the hypermedia can be made from information elements from all available systems, decision-makers can have access to all of the information they need. While the semantic network is serving higher-level decision tools, the content is left untouched and still accessible by those tools that interact directly with content databases.

Beyond executing applications from a single workstation, integrated information could provide decision-makers with a competitive advantage. In particular an integration method that brings the information into a semantic network can allow meaningful access to humans and autonomous agents. The goal of command and control systems should be to integrate information rather than just the applications. An architecture, such as that used for HOPS, centered on the structure of information can accomplish this goal. Military plans, tactical situations, and their interaction can be described using hypermedia induced semantic networks.

References

Davis, H., Hall, W., Heath, I., Hill, G., and Wilkins, R., "Towards an Integrated Information Environment With Open Hypermedia Systems", *Proceedings of Hypertext 1992*, ACM, 1992.

Lange, D. "Hypermedia Analysis and Navigation of Domains", M.S. Thesis, Computer Science Department, Naval Postgraduate School, Monterey, CA. 1997.

Lange, D., "Hypermedia Potentials for Analysis Support Tools", *Proceedings of Hypertext '99*, ACM, 1999.

Lucarrela, D. and Zanzi, A., "A Visual Retrieval Environment for Hypermedia Information Systems", *ACM Transactions on Information Systems*, Vol. 14, No. 1 (Jan 1996).

Nielson, J., *Hypertext and Hypermedia*, Academic Press, San Diego, CA, USA, 1990

Nurnberg, P., Leggett, J., and Schneider, E., "As We Should Have Thought", *Proceedings of Hypertext '97*, ACM, 1997.

Wiil, U., "Message from the OHSWG chair", *Open Hypermedia Systems Working Group Web Site*, http://www.ohswg.org/intro/chair.html, Dec 1, 1997.

Will, U., and Nurnberg, P., "Evolving Hypermedia Middleware Services: Lessons and Observations", *Proceedings of the 1999 ACM Symposium on Applied Computing*, ACM, 1999.