Knowledge Base Formation Using Integrated Complex Information

Douglas S. Lange
SSC San Diego

ABSTRACT

An intelligence support system has been developed using open hypermedia architecture. This approach integrates information from distributed disparate sources into a knowledge base. A public interface supports access by external applications. Filtering and change detection functions have also been implemented. The approach has shown promise in multiple domains, indicating possible wide application. This paper discusses the principles of the hypermedia framework for this system and how these principles may influence command, control, communications, computers, intelligence, surveillance, and reconnaissance (C^4ISR) systems in general.

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 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 (C^4ISR) systems have evolved, system integration has been the general theme. Stand-alone systems, each with its own database, were first interfaced to allow some data transfer. Data management schemes 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 not only allows multiple applications to share hardware, operating system, and network platforms, but also uses a layered service architecture that eliminates 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 broad systems made up of many small applications, any of which may be accessible through the user's workstation. Some applications work on common data managed through centralized services. Many data categories still form separate stovepipes since they are maintained in separate data repositories because of their differing technical natures and programmatic backgrounds. Users must associate the tactical situation shown in one application with the results of a logistical query conducted through another application.

Information Complexity

The focus on systems integration ignores the true goal in decision support. Information is of ultimate value to the decision-makers. Integrating the information is the next step. Unlike data-warehousing applications, military information is not just collecting and crunching sales and inventory figures from various branch offices. The military environment is complex. The variety of concepts, events, and situations that can be

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **AUG 2001**	2. REPORT TYPE	3. DATES COVERED -
4. TITLE AND SUBTITLE **Knowledge Base Formation Using Integrated Complex Information**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Space and Naval Warfare Systems Center,53560 Hull Street,San Diego,CA,92152-5001**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT
An intelligence support system has been developed using open hypermedia architecture. This approach integrates information from distributed disparate sources into a knowledge base. A public interface supports access by external applications. Filtering and change detection functions have also been implemented. The approach has shown promise in multiple domains, indicating possible wide application. This paper discusses the principles of the hypermedia framework for this system and how these principles may influence command, control, communications, computers, intelligence, surveillance, and reconnaissance (C4ISR) systems in general.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**7**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

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, bringing all data into a relational or object database will not completely accomplish information integration.

Pattern of Analysis

In researching the requirements for an intelligence support system for the U.S. 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. Analysts rarely create data, but search, filter, and review all available information. As they do, they form networks of related information [1].

DIA intelligence analysts spend some of their time building up a private model of their area of expertise. They spend the rest of their time responding to queries from DIA's various customers. The responses are typically linear essays. Analysts also periodically produce background reports on particular matters of interest. These reports also take a strictly linear, book-like form, even when delivered over a computer network.

Analysis of the current approach yielded the following problems:

- Products were static or updated using a paper publishing schedule.
- Customers with local information have no mechanism to share it with others.
- Only a particular question was 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 built by the intelligence analyst and communicating this knowledge to intelligence consumers. The goal was to move away from the linear essay to a more collaborative communications method. This method would allow for continuous update of the knowledge jointly held between the intelligence agency and its customers.

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 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.

HYPERMEDIA ARCHITECTURES

Hypermedia systems automate the management of information that is structured as described previously. Such systems provide the capability to work with a wide variety of data, while using the powerful information available through the structures created by the connections made among the various data items [2]. Hypermedia accurately records information, but its non-linearity allows the reader to access information in ways that the author did not necessarily expect. Users of analysis results can make new discoveries from the same body of data [3]. 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 create relationships among nodes.
- *Anchor.* Anchors connect nodes to the actual media that make up their content.

Open Hypermedia

From 1987 to 1991, researchers noted that the hypermedia systems did not support the needs of collaborative work groups and could not be integrated into computing environments used in large enterprises [4 and 5]. Requirements were found for hypermedia systems that were not addressed. These requirements included the following:

- 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 important to the 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 [6]:

- A system that does not impose any markup on the data. By marking up data to create hyperlinks, the data are changed, making the data 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 hardware platforms.

- A system in which there is no artificial distinction between readers and authors. This requirement 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 linked together by the analysts may be coming from other applications and databases integrated with the hypermedia system. These applications will not understand changes imposed on the data to support linking. The links must be separated from the content. This separation is the basic premise of an open hypermedia system. It has been demonstrated in many research systems [7].

The prototypical open hypermedia system is structured as shown in Figure 1.

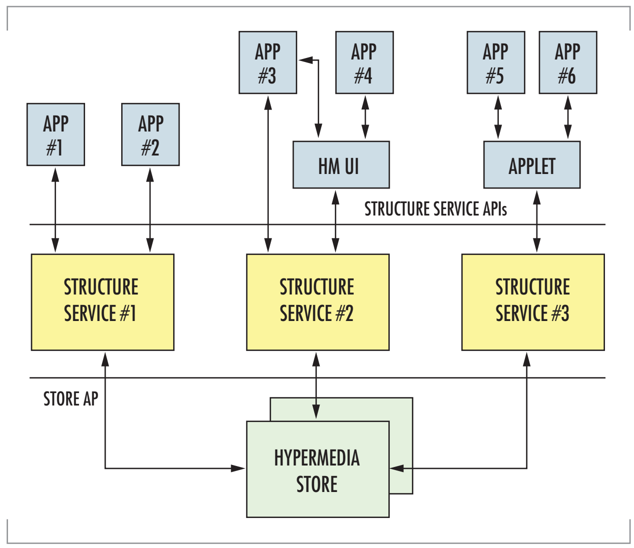


FIGURE 1. Open hypermedia architecture [9].

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 are based on set and graph theory, providing mathematically defined filter, search, and navigation methods. This category of hypermedia also includes human–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 [8]. The relationships among schema types and between schema entries and the instances created in the hypermedia closely mirror the relationships in object-oriented design.

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 provide the concept of a perspective. A perspective contains three elements. The first element is the perspective pattern. 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, a subset of the instance set satisfies both of the constraints.

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 to support analysis of the kind performed by DIA. These same modifications would appear to be important for related C^4ISR systems.

General Architecture

HOPS follows the open hypermedia form with the architecture shown in Figure 2.

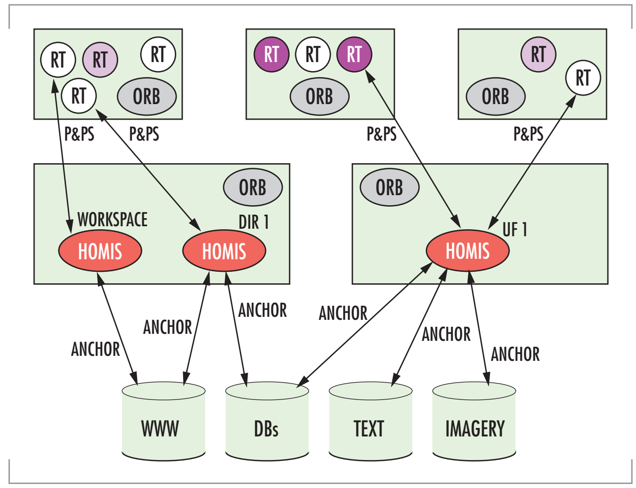


FIGURE 2. HyperObject Processing System.

In this figure, circles labeled "RT" represent runtime applications supporting a user directly or automated processing. "HOMIS" (HyperObject Multimedia Information Systems) are modified multimedia information systems [8] that can handle hypergraphs rather than simple graphs [10]. HOMIS function as structure servers, as called for in 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. Object request brokers allow the system to be distributed over multiple platforms.

Unique Hypermedia Features

Most hypermedia systems found in research literature work with information spaces constrained by either the level of diversity and quantity of the information, or by 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 are necessary to allow HOPS to handle the dynamic unbounded nature of military information integration.

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. A semantic network forms that describes the relationships among media elements rather than the tactical situation. To remedy this problem, HOPS uses multiple anchors per atomic node. Use of multiple anchors allows the nodes to define concepts or real-world objects and allows the links to represent relationships among them rather than relationships among the content elements.

Large Open-Ended Schema

Schemas imply an ability to predict all 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 [1]. An example can be demonstrated in terms of exercise plans. During Tandem Thrust 97, 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 adapt on the fly to allow analysts and decision-makers to see and interpret information and record and inform regarding 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 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 filter both in terms of the complexity and in terms of the size of the knowledge base that they work with to avoid being overwhelmed. HOPS allows this capability through adaptations of the graph-based hypermedia concepts of perspective patterns and filters [10]. Perspective patterns allow the user to limit the kinds of information being worked with, while filters focus attention on information with particular content.

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, 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.

Link Equality

Although hypermedia relies on associations between elements for its character, many interaction techniques found in research literature still focus 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 decision-makers is found in 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 for navigation. They themselves provide critical information. HOPS handles this by making links special types of nodes, allowing all the mathematics of filtering, searching, and browsing to work on links. [10].

Framework

HOPS itself is not a command and control system or an analysis system. HOPS is a hypermedia framework designed to support analysis and to provide some generic applications for interacting with the hypermedia. HOPS is intended to be used by adding domain-specific applications along with an initial schema to create an analysis system of the type needed. Such work is in support of DIA's mission.

In the Military Operations in the 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 within the framework, users have a variety of ways to work with the information.

PROSPECTS FOR INFORMATION INTEGRATION

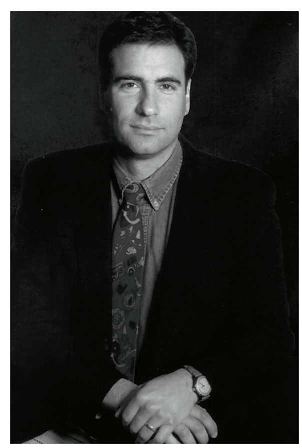
Hypermedia systems hold promise for information integration. Any number of decision-support tools can access the semantic network formed of the associations and nodes. Decision-makers can have access to all the information they need because the hypermedia can be made from information elements from all available systems. While the semantic network is serving higher level decision tools, the content is left untouched, and is 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. An integration method that brings the information into a semantic network can allow meaningful access to human beings and autonomous agents. The goal of command and control systems should be to integrate information rather than just the applications. 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

1. Lange, D. 1999. "Hypermedia Potentials for Analysis Support Tools," *Proceedings of Hypertext '99*, Association for Computing Machinery (ACM), pp. 165–166.
2. Nurnberg, P., J. Leggett, and E. Schneider. 1997. "As We Should Have Thought," *Proceedings of Hypertext '97*, ACM, pp. 96–101.
3. Nielson, J. 1990. *Hypertext and Hypermedia*, Academic Press, San Diego, CA.
4. Halasz, F., 1987 "Reflections on NoteCards: Seven Issues for the Next Generation of Hypermedia Systems," *Proceedings of the ACM Conference on Hypertext*.
5. Malcom, K., S. Poltrock, and D. Schuler, 1991. "Industrial Strength Hypermedia: Requirements for a Large Engineerng Enterprise," *Proceedings of the Third ACM Conference on Hypertext.*
6. Davis, H., W. Hall, I. Heath, G. Hill and R. Wilkins. 1992. "Towards an Integrated Information Environment with Open Hypermedia Systems," *Proceedings of Hypertext 1992*, ACM, pp. 181–190.
7. Wiil, U. 1997. "Message from the OHSWG Chair," Open Hypermedia Systems Working Group Web Site, http://www.ohswg.org/intro/chair.html (December).
8. Lucarrela, D. and A. Zanzi. 1996. "A Visual Retrieval Environment for Hypermedia Information Systems," *ACM Transactions on Information Systems*, vol. 14, no. 1 (January), pp. 3–29.
9. Wiil, U. and P. Nurnberg. 1999. "Evolving Hypermedia Middleware Services: Lessons and Observations," *Proceedings of the 1999 ACM Symposium on Applied Computing*, pp. 427–436.
10. Lange, D. 1997. "Hypermedia Analysis and Navigation of Domains," Master's Thesis, Computer Science Department, Naval Postgraduate School, Monterey, CA.

❖



Douglas S. Lange

MS in Software Engineering, Naval Postgraduate School, 1997

Current Research: Software generation; knowledge bases; enterprise architectures.