An Architecture for Organising Dynamic Information About Space and Time

Andreas Björklind, LIBLAB, Dept. of Computer and Information Science, Linköping University, Sweden

Abstract This article describes an architecture for handling heterogeneous information about space and time. The proposed system consists of a meta-database which is handled by a meta-data engine. The engine functions as the filter between the crisis management system (based on a geographical information system) and the large heterogeneous raw database. The meta-data handler is a HyTime engine, based on the international standard ISO/IEC 10744. The internal query language is the HyQ notation from the HyTime standard.

1.0 Introduction

Today’s information systems are stretched to their limits in situations of continuous, large quantity flows of spatio-temporal data. Such situations are encountered in, for example, assessment and control of crises and emergencies. Meta-databases for the management of data about data have become a necessity.

There are four main differences between the meta-databases for assessment and control systems and traditional information retrieval systems; e.g. bibliographic information systems. The first is the need for speed in generating and updating the meta-database. Secondly there are spatio-temporal aspects of the information. The third difference is the large volume and variety of information to be managed and to be transformed to symbolic form. Fourth is the collaborative interaction in the assessment and control process.

It is evident that traditional approaches are not satisfactory in organising multi-media spatio-temporal data. In this paper an architecture for organising heterogeneous spatio-temporal multi-media data is proposed. This paper presents some early ideas on spatio-temporal meta-data management in a crisis management system. The underlying GIS (Geographical Information System) is working at FOA (Swedish Defence Research Establishment). The proposal in this paper is to be seen as a continuation of the work already done at FOA. Section 8.1 of this paper is a further discussion of future work within this project.

2.0 Geographical Information Systems

Geographical Information Systems (GIS) is a quickly expanding area for research and development. There are many journals and conferences on the subject, and recently there are beginning to be textbooks available. The early GIS work was geared toward digitising maps and other cartographic information; see for example the collection (Peuquet & Marble 1990) for a survey of research and development up to 1990. The research and
development was being done within the cartographic and geographic disciplines. The awareness of the need to find a more general information view has emerged during the last few years. GIS must be able to handle both cartographic and "ordinary" information. Currently important topics are integrated information handling and generalisation of information (see further discussion about generalisation in section 6.0). Several people in the GIS community have begun to talk about Spatial Information Systems to accentuate a more general information science approach. See for example the recent textbook of Spatial Information Systems (Laurini & Thompson 1992).

2.1 Assessment and control systems

Assessment and control systems can be regarded as knowledge organising systems. They are usually intended for a large number of users. In the typical assessment and control system the user's task is to monitor, contain and bring to an end an emergency. The focus of this paper is on crisis and emergency management teams and the assessment and control systems they use. To carry out their task they require extensive amounts of information of widely different character; ranging from verbal documents by telephone to data from different types of remote sensors. In this paper all types of incoming data will be called documents regardless of source (e.g. sensor data from radars, reports in the form of verbal communications, data delivered on paper or electronic mail).

The outgoing information from a system of this sort is, like it's input, heterogeneous. There can be orders delivered by verbal or written media, new instructions to automatic devices (e.g. sprinkler systems or traffic signals), inquiries for new information, part of plans in form of instructions to field personnel.

3.0 Information flow

In large scale assessment and control operations the scale of information flow is vast. One of the most important problems is not just to allow for the flow itself, but rather to automatically interpret the structure and contents of documents, and relate them to each other. There are at least two major problems with the traditional way of human interpretation and knowledge management. First, after the crisis is over and dealt with it is usually very difficult to perform a subsequent analysis and evaluation of the work. "Playback" functionality is vital in assessment and control systems. Second, the training situations that in a realistic way mirror the controlled chaos of a real life situation are difficult and costly to set up. Therefore simulation functionality is also very important in a system for assessment and control.

Generalisation of information is an important topic. In GIS research the emphasis has been on generalisation of map information. In assessment and control operations generalisation of information related to geographical space and to time is vital. Humans generalise when handling large amounts of information. An important issue is to give decision support for the process without withholding vital information. Generalised information must clearly be noted as such when presented.

4.0 Spatio-temporal information

In the project which this paper describes, the information is coming from various kinds of sources. Common to all documents are that they are linked to one or more points or regions in space, and in time. For this reason all information flowing into the system can be indicated on a map together with a time stamp indicator. A geographical information
system (GIS) will be required for the digital map, and the mentioned GIS at FOA in Linköping (FOA is the Defence Research Establishment in Sweden) will be used. The basic spatial unit in the system is Run-Length-Code (RLC) which is an object oriented data structure (Jungert 1986).

Two kinds of information will be processed by the system: sensor data and reports. Sensor data are automatically generated by remote sensors and could for example be radar images or video shots. Reports are created and delivered by humans, for example telephone messages or electronic mail. All incoming information must be aligned; i.e. adopted to the actual map used, with respect to scale, orientation and extension. The alignment to a map is not of concern in this project. All incoming information is considered to have undergone the transformations needed. Both kinds of incoming information are basically documents — using the broad definition of the word.

4.1 Formal description

The necessity for formal approaches to the description of objects in space and time, with complex structures and interrelationships, have become apparent with the growing interest in hypermedia. There is also a growing awareness of the necessity to be able to interchange information between experimental and commercial application programs.

The international standard for Hypermedia/Time-Based Structuring Language — HyTime (ISO 10744) — an extension of SGML (Standard Generalised Mark-up Language, ISO 8879), provides a formal framework in this project. It is suitable for the descriptive tasks arising in the context of assessment and control systems, both for the planning tasks and for the meta-database design.

HyTime can be seen as a formal language for describing hypermedia documents and their relations. The specific processing instructions in a hypermedia document are tied to specific products and applications. Many of these instructions are for presentation and are application specific. Some of the instructions are concerned with the internal description and representation of the information, about the interconnectivity between documents and other application unrelated instructions. By using HyTime one can describe the fundamental properties of documents and thus be able to use the same information on different platforms, by different applications. For further reading on HyTime there is the standard itself (ISO/IEC 1992) and the Newcomb article, which is slightly inaccurate due to being written before the standard was finalised (Newcomb, Kipp & Newcomb 1991).

The HyTime standard (ISO/IEC 1992) specifies a number of architectural forms which are grouped into HyTime modules. In a specific document or application you only need to be using the modules of the standard that are appropriate. In the project described in this paper all modules will be used (Base, Measurement, Location, Hyper links, Scheduling, Rendition, and Event projection modules). The foremost argument for using HyTime in this project is the great interoperability benefits between this and other hypermedia systems. In the HyTime standard a query language is defined: HyQ: HyTime Query Notation. This query language will be used for the internal query processes in the proposed system.

One of the problems with HyTime, like with it's parent standard SGML (Goldfarb 1990), is that there are no current commercial or otherwise applications supporting the standard. This means the initial effort in a new HyTime based project is big. There are several books and software announced for release in 1994.
4.2 Reasoning

Reasoning is becoming more and more important in applications dealing with large quantities of information where one of the aims is to identify types of object interrelationships; see the collection (Frank, Campari & Formentini 1992) and the article (Pissinou & Makki 1993).

Various kinds of reasoning techniques exist. When reasoning on spatial information one of the problems is the extremely large amounts of information. A solution is to filter out most of it by using methods that just use a fraction of the available information, mainly represented in a low resolution. This is generally called generalisation and is discussed in section 6.0. If necessary, it may also be possible to include more information (higher resolution) in later stages of the reasoning processes when good approximations of the final solutions are available, in order to refine the result even further.

The main goal is to reason on at least two levels. The first level is for reasoning on meta-information about documents that arrive through many different media, and are distributed both with respect to time and space. The second level will be concerned with reasoning on information that is the content of the incoming documents. On both these levels it will clearly be of importance that the methods are formal and qualitative. Formal methods are e.g. used to reason about spatial and temporal relations. Qualitative methods are e.g. used to "filter" which documents are important in a given situation. These levels of reasoning will make it possible to develop decision support aids. The user will have means for making his own decisions based on information given. It is not within the scope of this project to design a decision support system. That is a system to be built as a application interface to the GIS.

5.0 Databases for storing multi-media information and their logical structures

The multi-media databases of this project will become very large (one single document may well be as large as 1 giga byte), and must allow for the storage of a wide range of data types. The proposed way of organising the multi-media database is to store conventional data in one database and let images and similar types of multi-media data be stored outside the main database in a suitable form.

5.1 Meta-databases

In traditional database design the meta-database is the data dictionary. The data dictionary is static and is well suited for static databases. In the case of this project data is highly dynamic and a traditional data dictionary would not suffice as a way of describing it.

The meta-database in this project is concerned with the logical structure and access methods of documents in the database. The main problem is to design a meta-database that can handle a large flow of documents in which spatio-temporal data are of great concern. The meta-database cannot be fixed in its structure, but must be dynamic depending on the actual crisis. The meta-database must also include means for decision making based on various methods for spatio-temporal reasoning that can be used in other subsystems or applications that will be using the meta-database. It will be necessary to have means for selecting the information that is available in the database with respect both to time and space.
It has already been pointed out that the meta-database must be integrated with a geographical information system which will support the visualisation of the meta-information with which the meta-database will be concerned. The use of a GIS will, of course, support the visual aspect of the meta-database.

6.0 Generalisation

Information generalisation is a method which has been used in GIS. The most used generalisation is that of map images. The research and development effort is focused on generalisation of cartographic information (i.e. generating lower resolution maps from high resolution data).

It is important to remember that generalisations are not a true picture of reality. The strength of generalised information is also their weakness: a simplified version of the original information. The important issue is how to generalise in an appropriate way. This has for example been discussed by (Persson 1994).

7.0 The proposed system

The ideas presented in this article are to be integrated into the working GIS of Erland Jungerts project at FOA. The GIS already has a presentation module and a working query language. The following details are only concerned with the additional work to be done in order to integrate the system.

The remainder of this section gives a description of the proposed system's parts. The system is presented roughly in a process flow view. I.e. modules are describes in order of which sequence the incoming data gets processed.

The input to the system consists of a heterogeneous information flow described in section 3.0. When the documents enters this system they have been aligned (normalised) and stamped with time and space co-ordinates. Time may be a point in time or a period of time. The space element will be a point in space or an area or a three dimensional space. The process of alignment and stamping is beyond the scope of this project. There are other groups at FOA working on these issues.

The cataloguer is a fundamental part of the system. This is the unit that transforms the aligned and stamped input to meta-data. The original data itself is handed to the raw database system. The catalogue information will, other than time and space information, include information about data type, size of raw data, creator of the document, raw database location, a specific expiration time stamp used for quickly changing information (like in crisis emergency systems) and other information that is needed for the meta-data engine (the system module which handles the meta-data). The catalogues will have support data: a thesaurus and a document structuring engine. The thesaurus is in practice like a thesaurus in bibliographic information systems. The document structuring engine will take the raw data and make the necessary transformations for it to conform to the internal HyTime structuring standard. The cataloguer passes the modified raw data to the raw database, and the cataloguing information to the meta-data engine.

As indicated in the section about databases for multi-media information, the proposed approach is to have two main databases in the system. One for storing the original raw data and one for the transformed data, the meta-database. The raw database is the collection of the original documents, in their original form. The original form may have been slightly
altered in order to conform to the HyTime document structure definition. The raw database is used by the GIS when a document is to be accessed in its pure form.

The meta-data engine is the most fundamental component of the proposed system. This is the module which maintains the virtual objects database, the meta-database. It uses the data generated by the cataloguer to add new objects to the meta-database. It generates necessary relations (links) between existing objects and to the map. This is in principle what in HyTime literature is called a HyTime Engine (ISO/IEC 1992).

The virtual objects database is the meta-database. This database is only used by the meta-data engine. It consists of all the meta-data of the raw data. The map is stored in this database. Relations between meta-objects and the map are stored and the relations between objects.

The existing GIS at FOA uses a query language called Graqula. In order to translate these queries into the internal language there is need for a query translator. The internal meta-data query language (HyQ) is not suited for direct human use. It's connection to HyTime makes it interchangeable, but there has to be an application based query language for the system. Graqula and the interface to it in the GIS provides the appropriate means for query input.

Finally there is the Geographical Information System itself. This system is discussed in length in other publications (see the reference section). The GIS will send Graqula queries to the meta-data engine via the translator, and receive the ensuing results. If a document is to be viewed, the raw database location is used for retrieval and viewing. In this system queries like these should be handled: "Give me all relevant information about the forest fire in Valla Forest, not more than 15 minutes old"; "Show me all information relating to the NW part of Valla Forest".

8.0 Further work

This paper has discussed the fundamental ideas for the expansion of the existing GIS at FOA in Linköping. There are other research and development areas which may find the proposed work useful. The major application for the proposed architecture is the assessment and control systems in crisis and emergency management. The architecture is also highly interesting in other application areas where heterogeneous dynamic multi-media documents are to be received, analysed, stored and retrieved. One example is Computer Supported Co-operative Work (CSCW) in systems analysis and design situations.

The following two sections will discuss further work possible from the ideas presented in this paper. First is the work planned within the current joint project at LIBLAB and FOA. Second, the research areas open for further exploration which have been touched on herein.

8.1 The current project

For more information on the GIS of this project and the ideas permeating it, please refer to the material published about it. For example (Jungert 1990), (Jungert 1993) and (Persson & Jungert 1992).

One of the primary tasks in the project is to analyse the various kinds of incoming information with regard to its various characteristics. The resulting analytic information will be used to design methods of symbolic description for the meta-database. This will
result in a series of HyTime document definitions. The analytic information will also be used in the design of appropriate indexing and classification schemes, and for creating a thesaurus for the cataloguer module. The transformation of the information to symbolic form, using HyTime, will enable it to be used both in reasoning and for subsequent presentation in the GIS-system. By making the raw data compliant to the HyTime standard several commercially available display systems will be suitable. By making the internal meta-data HyTime structured (formally described), formal reasoning is more readily applicable.

Following the initial document analysis the following topics are to be addressed for continued work: meta-database technology; HyTime engine design; Graquila to HyQ translation.

8.2 Further research

Meta-databases and their design will be of continued interest for the information processing field. As the amount of electronic documents grows there will be a need for working abstractions of the information mass. Generalisation techniques for information in the broad sense are one way of combining meta-database technology and expert systems. How do we make an abstracted video of 15 minutes footage of a forest fire? These and related questions has to be solved in order to make large amounts of electronic documents accessible.

In this volume, Björklund discusses the use of meta-database technology in the area of CSCW in her "The Potential of Using Knowledge Organizing Tools in Collaborative System Development" (Björklund 1994). The environment of the emergency management system involves group-work. The team consists many different kinds of professionals involved. Therefore the ideas and problem areas discussed in Björklund's paper are important for the future work of this project with regard to both the design of the meta-database, and the design of the user interface system to the GIS intended for emergency management.

References


