IEEE TRANSACTIONS ON

KNOWLEDGE AND DATA ENGINEERING

A publication of the IEEE Computer Society

MARCH / APRIL 1999 VOLUME 11 NUMBER 2 ITKEEH (ISSN 1041-4347)

REGULAR PAPERS

CORRESPONDENCE

Dr. Naphtali Rishe
Florida International University
School of Computer Science
Southwest 8th St. and 107th Avenue
University Park
Miami, FL 33199

IEEE COMPUTER SOCIETY

http://computer.org • tkde@computer.org

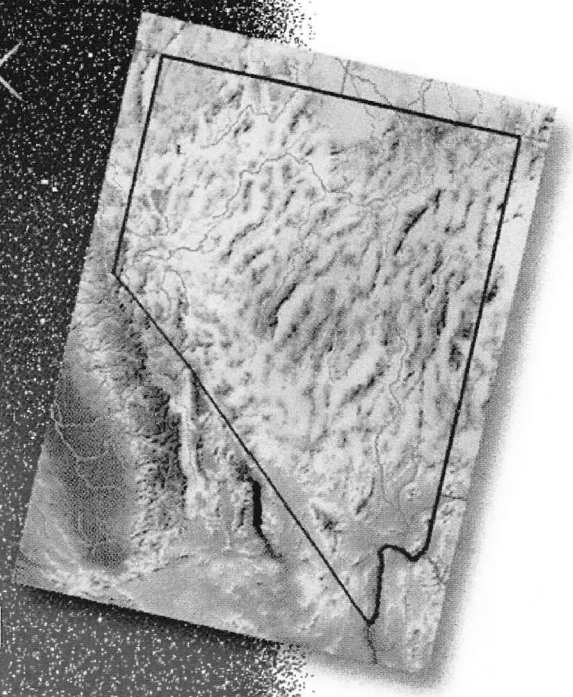
f= U

8532

99-SD

ISSSR

SYSTEMS AND SENSORS FOR THE NEW MILLENNIUM



LAS VEGAS, NEVADA

THE TROPICANA HOTEL

OCTOBER 31 – NOVEMBER 4, 1999

TUTORIALS / SYMPOSIUM PAPERS
SYMPOSIUM POSTERS / WORKSHOPS / EXHIBITS

International Society for
Photogrammetry & Remote Sensing

A Semantic Database (Sem-ODB) Solution for Efficient Storage and WWW Retrieval of Meteorological Data to Facilitate Weather Research and Forecasting*

Naphtali Rishe, Martha E. Gutierrez, Alejandro Gonzalez, Debra Lee Davis
High Performance Database Research Center
NASA Regional Application Center
Florida International University, Miami, Florida
hpdrc@cs.fiu.edu

ABSTRACT

Weather forecasting is very important in the world today. By analyzing meteorological data, one can formulate weather predictions and investigate weather phenomena like El Niño, or La Niña. The GOES-8 satellite, which is one component of the GOES system for U.S. weather monitoring and forecast operations, generates about 3.1 GB of data over the United States per day. This massive amount of data requires an efficient system to facilitate retrieval, storage, and data analysis for scientists wishing to study this data over time.

The High Performance Database Research Center (HPDRC) at the Florida International University has developed semantic database technology (Sem-ODB) capable of efficiently storing and retrieving these types of spatial data sets. This system can retrieve weather data, not only from our GOES-8 ground station but also from numerous machines throughout the Internet, and automatically load this data into a Semantic Database, which can be viewed via the World Wide Web.

1. INTRODUCTION

Weather forecasting is important; it helps save lives and mitigate property loss by issuing watches, warnings, and analyses of hazardous tropical weather. One major aid in weather forecasting is the use of imagery from meteorological satellites. There are a variety of meteorological satellites and satellite systems. GOES (Geostationary Operational Environmental Satellite) is one such system. The data provided by many different satellites is one of the most important tools that forecasters use. This data is archived in order to study long term phenomena such as El Niño, or La Niña. In order to analyze thousands of data sets and images, these must be archived efficiently and must be easily accessible.

Many problems exist that limit the availability of this data to consumers in the outside world. Due to the amount of data, and the organization of the data it is difficult to quickly view and study the data. Many of these datasets require expensive processing program in order to view the images from these satellite systems. Due to weather, and human nature, it is possible to miss some important data during the collection process. The amount of meteorological data generated by the satellites every day, make it difficult to maintain, organize, and utilize the storage space.

We have developed a suite of solutions to these problems that are implemented using Sem-ODB technology, and the World Wide Web. Using the Internet as a platform, we are able to load Datasets into the Semantic Database from many different machines. The World Wide Web is then used as an interface to view images from the GOES Series of Satellites and to perform Algorithms on these images. Using the Semantic Database Technology as a solution greatly increases the speed of data retrieval, processing, and allows scientific consumers to use the data efficiently.

1.1 HPDRC

The High Performance Database Research Center (HPDRC) is a division of Florida International

* This research was supported in part by NASA (under grants NAGW-4080, NAG5-5095, NAS5-97222, and NAG5-6830) and NSF (CDA-9711582, IRI-9409661, HRD-9707076, and ANI-9876409).

University (FIU), School of Computer Science. It conducts research on database management systems and various applications, leading to the development of new types of Database Management Systems (DBMS), new database techniques, and the refinement of existing ones.

HPDRC's largest project is the development of Sem-ODB, a prototype massively parallel Semantic/Object Oriented DBMS. Our system is useful for most typical database applications, as well as for specialized domains such as Earth Sciences [RISH95].

1.2 RAC

The NASA Regional Applications Center (RAC) at FIU is a subdivision of HPDRC. NASA has established approximately 15 RAC's across the country at this time. The RAC Program was initiated by NASA Goddard Space Flight Center's (GSFC) Applied Information Sciences Branch, to extend the benefits of its information technology research and cost-effective system development to a broader user community. The RAC objectives are based on the goal of fostering the use of environmental and Earth resource data by regional institutions. The ultimate goal of the RAC is to establish a fundamental set of remote sensing technologies that can be assembled by a specific user community, to meet the information needs of that community.

2.0 SEM-ODB

Sem-ODB, HPDRC's Semantic DBMS, is based on the Semantic Binary Model. In the Semantic Binary Model, the information is represented by logical associations (relations) between pairs of objects and by the classification of objects into categories. The Semantic Binary Model is a natural and convenient way of specifying the logical structure of information and for defining the concepts of an application's world [RISH92].

2.1 DESCRIPTION

The Semantic database models are potentially more efficient than the conventional models for two main reasons. The first is, all the physical aspects of the representation of information by data are invisible to the user. The second is that the system knows more about the meaning of the user's data and about the meaningful connections between such data. The first reason creates a potential for optimization by allowing more changes without affecting the user programs. The second allows this knowledge to be utilized to organize the data so that meaningful operations can be performed faster at the expense of less meaningful operations [RISH92].

The mathematical abstraction of the relational model has allowed the introduction of powerful and easy-to-use languages for retrieval and updates of databases. The semantic model however, offers a higher degree of abstraction, which results in more concise user programs, speedier processing (due to optimization), and a wealth of other features. Relational databases are good for general conventional database applications. However, in situations where the structure of information is complex, or where greater flexibility is required (objects with unknown identifiers, or objects moving from one category to another, etc.), or where non-conventional data is involved (spatial data, long text, images, etc.), semantic databases need to be considered.

The efficient retrieval and updates are a requirement of the semantic database. Requests are maximized by decomposing queries into atomic retrieval operations and each atomic retrieval request normally requires only one disk access. A transaction is composed of a set of facts. These facts can state that the objects belong to a category, they can state that there is a relationship between objects or they can be fact relating objects to data, such as numbers, texts, dates, images, etc. HPDRC's Semantic DBMS contains semantic facts and inverted semantic facts. This fact inversion scheme assures efficiency of queries including range queries and content access and also exhibits low entropy of data blocks, which facilitates compression [RISH92].

3.0 METEROLOGICAL DATA AND THE GOES SATELLITE SERIES

Meteorological Data is useful to study many things. Such products derived from the data are: Cloud Cover, Water Vapor, and Sea-Surface Temperature. These are useful to produce Weather Forecasts and to study the climate and global environment. The GOES Satellite System provides most of the useful information required by Meteorologists around the world.

GOES (Geostationary Operational Environmental Satellite) records cloudcover to aid in monitoring weather patterns and severe storms. It is also used to estimate rainfall during thunderstorms and hurricanes

for flash flood warnings, as well as estimate snowfall accumulations and the overall extent of snow cover.

The GOES satellite accomplishes its mission using two different instruments, the Imager and the Sounder. Scan control and data collection for the instruments are independent of each other and of most other activity in the spacecraft. [NESD94] The imager is a multi-channel instrument that contains five different channels. Channel one covers the visible range, channel two is the near infrareds range, channel three studies water vapor, channel four and five are in the thermal infrared range. The sounder contains many other channels and is used to determine the vertical temperature, the moisture profile of the atmosphere, surface, cloud top temperatures, and determines the ozone distribution. The imager is the most widely used data by scientific consumers and is the one used for the WWW weather system described in this paper.

The GOES Satellite Series has two different resolutions for the Imager Instrument; it has 1KM resolution for the visible channel, and 4KM resolution for the other 4 channels.

4.0 PROBLEM

Scientists and researchers wishing to analyze weather patterns can sometimes encounter limitations, due to the availability of the data, the various formats in which the data is produced and finding a matching processing software that conforms to the data format. With these limitations, it is often difficult to perform real-time analysis of data unless you are equipped with a ground ingest system.

At HPDRC, we have a GOES-8 GVAR ground ingest system, and although monitoring weather patterns becomes less complicated, putting these images on the WWW for others to access in an efficient manner is not as trivial. The GVAR data format is rather complex and consumers wishing to quickly view and produce products from this meteorological data, would have to have a GVAR ingest system or the proper software to convert the format. They would also have to deal with the time-consuming task of ordering the data to perform the algorithms. In addition most sites offer only recent data and algorithms cannot be performed until the data is ordered.

Continual data archival may also present additional problems. Data collection can be disrupted by a computer crash, a receiver malfunction, and weather storms can sometimes interfere with reception. These problems can lead to missing datasets that may be of importance to the scientific community. Continually storing data takes up a huge amount of space. GVAR transmits about 3.3 gigabytes of data a day. To store data for a year, would require over one terabyte of storage space.

5.0 SOLUTION

HPDRC has developed a solution to the problem and semantic database technology is an integral part of this solution. By using Sem-ODB, we are storing meteorological data in an effective and organized manner. An efficient design and implementation of a Sem-ODB weather database and a suite of applications in conjunction with web forms to facilitate querying of the data, make up this system.

5.1 DATABASE

The weather database is currently comprised of GOES-8 imagery for the hurricane season. Therefore, it is often referred to as the hurricane database. This database was developed to provide an environment for scientists and non-scientists to have access to GOES-8 imagery for an entire hurricane season. The database is categorized by hurricane season, which in turn is categorized by date.

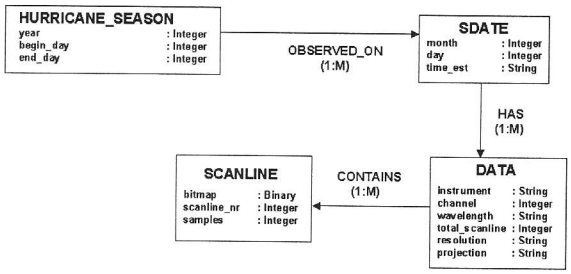


Figure 1 – Hurricane Database Semantic Schema

The semantic schema contains four Categories: HURRICANE_SEASON, SDATE, SCANLINE, and DATA. The HURRICANE_SEASON category contains the year of the hurricane season, and the starting and ending Julian day for the season, this distinguishes one hurricane season from the next and allows storage of multiple season in one database. The HURRICANE_SEASON category contains many SDATE, which describes the month, day, and time of the data set. Each DATA object may contain many channels and many instruments. The data is stored by

scanline, so it makes sense that each channel's image is divided into scanlines. This is an example of the power, and ease of understanding of Sem-ODB technology.

For GOES-8 GVAR, each CONUS data file is ~65 MB and stored by scanlines. This is important to the associated application that requires that this data be displayed in close-to-realtime. Storing the data by scanlines permits efficient scaling down of the image during retrieval, for better performance. To scale an image that is stored by scanline, let N = the scale factor. Query only every N^{th} scanline. Once you have this, when you process the raw data file, you get every N^{th} pixel. This method efficiently reduces retrieval time.

5.2 WORLD WIDE WEB

The associated multimedia application is an auto ingest system which fetches the data from the our GOES-8 ground station, loads it into the Sem-ODB database, automatically, and facilitates querying of this database through a multimedia web application. The "latest weather image" displayed in near-real-time is resampled by a factor of 4. The images displayed when the database is queried online are resampled in the same way. The web interface provides the user an option to create a weather loop (movie) from a selected time span as well as query the database for individual weather images. Both of these operations offer the user a resampled image to speed the retrieval process as well as the weather loop.

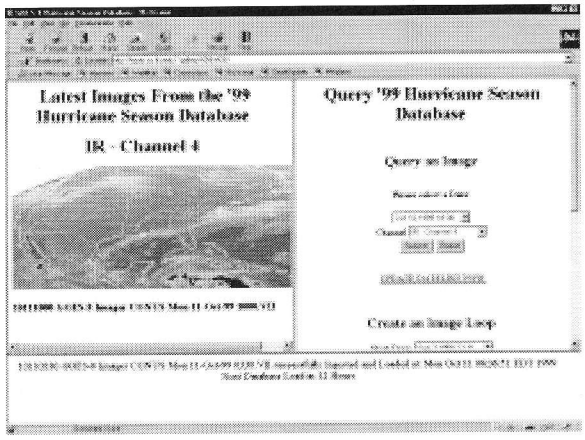


Figure 2 – World Wide Web Interface

The World Wide Web Page has 3 main frames: the latest image, the hurricane database query form, and the console.

1) The Latest Image.

The Latest Image shows the most recent processed image acquired by the ground ingest system, although some images shown here may not have been loaded into the database. This image is updated from the main auto-ingest system described in the next section.

2) The Hurricane Database Query Form

This form allows one to query the database by selecting a date and channel. The image will then come up scaled. This form also allows the user to view an image loop by selecting a starting and ending image date. This image loop is shown as an animated GIF and allows one to take a broad look at the weather pattern for a specified period of time. Should the user need to further analyze this data, the entire image can be retrieve from the database. This image loop is similar to those, which many TV news and weather station use.

3) The Console

The console shows a message describing the latest action taken by the auto ingest program. It describes what the server is doing, any errors it may have encountered, and the last file loaded into the database.

5.3 AUTO-INGEST

The auto ingest system uses FTP and UNIX scripts to communicate with many machines throughout the internet and automatically load data into the hurricane database. Both FTP and UNIX scripts are used by this system to access remote sites, determine which is the latest file added to that site, download that file, and load it into the database. This process can be timed to update the database as often as needed.

The Auto Ingest System is composed of four main modules. Their functionality is described below.

Module 1 - Check/Download from Ingest to Server
A list of the current files in the system is downloaded from the ingest machine to the server. From this list, the latest ingested file is selected. This file is then downloaded to the server for further processing. This is a timed process and varies depending on the requirements. The time interval is set by the first parameter to the System.

[NESD94] R. Komaid, K. Mckenzie. *NOAA Technical Report NESDIS 82,* US. DOC, 1994.

[RISH92] N. Rishe, *Database Design: The Semantic Modeling Approach*, McGraw-Hill, 1992.

[RISH95] N. Rishe, W. Sun, D. Barton, Y. Deng, C. Orji, M. Alexopoulos, L. Loureiro, C. Ordonez, M. Sanchez, A. Shaposhnikov, "Florida International University High Performance Database Research Center," *SIGMOD Record*, 1995, vol. 24, no. 3, pp. 71-76.

[SPAC94] Space Systems/Loral, *Operations Ground Equipment (OGE) Interface Specification*, DRL 504-02-1 Part 1, Palo Alto, CA, 1995.

[SAID93] A. Said, W. Pearlman, "An Image Multiresolution Representation for Lossless and Lossy Compression" *SPIE Symposium on Visual Communications and Image Processing*, Cambridge , MA, Nov. 1993.