The WorldInfo Assistant: Spatio-Temporal Information Integration on the Web*

Jose Luis Ambite‡, Craig A. Knoblock†‡, Mohammad R. Kolahdouzan†
Maria Muslea‡, Cyrus Shahabi† and Snehal Thakkar†

†University of Southern California
Department of Computer Science
Los Angeles, CA, 90089

[kolahdoz, shahabi, snehalth]@usc.edu

‡ University of Southern California
Information Sciences Institute
4676 Admiralty Way, 10th floor
Marina del Rey, CA 90292

[ambite, knoblock, mariam]@isi.edu

Abstract

Due to the recent growth of the World Wide Web, numerous spatio-temporal applications can obtain their required information from web sources. In this demonstration we show *The WorldInfo Assistant*, an application that extracts and integrates spatial, temporal and other information about different regions of the world from different web sources and databases. This application also provides integration of different vector data with the satellite images of different regions of the world. Finally, We demonstrate several approaches for efficient querying moving objects with predefined paths and schedules.

1 Introduction

On both the public Internet and private Intranets there are vast number of data sources available and many of these sources contain information that is spatial, temporal, or pertains to objects that have a spatial or temporal nature. These sources are owned and maintained by different organizations, cover different parts of the world, and are stored in different types of systems in different formats. The challenge is how to efficiently integrate these sources together as if they reside in a centralized system.

In a previous demonstration [1], we showed our information mediator and wrapper technology to retrieve and integrate data from several web and database sources. In this demonstration, we focus on the integration of spatial and temporal data sources and demonstrate *The WorldInfo Assistant*, an application that extracts spatial, temporal and other information about different countries in the world from different web and database sources and interactively integrates the data using a dynamic, hierarchical constraint engine (termed Heracles [2]). New challenges are introduced as a result of the limited accessibility to spatio-temporal web sources. In particular, we discuss alternative approaches we use in the WorldInfo Assistant to efficiently query moving objects with predefined paths and schedules.

2 The WorldInfo Assistant

The WorldInfo Assistant integrates data coming from several web and remote/local database sources in real time as if the data is retrieved from one centralized source. The WorldInfo Assistant provides a GUI for user interaction and presents the data in different formats such as text, satellite images, maps, etc. The users specify the region of the world they are interested in and the WorldInfo Assistant provides a variety of information for the selected region of the world from pre-specified web sources. For example, latest news information is obtained online from *cnn.com* and up-to-date weather information is fetched from *weather.yahoo.com*. Maps for the WorldInfo Assistant are retrieved off-line from National Imagery and Mapping Agency (NIMA) and stored in our own databases. Similarly, The satellite images are obtained from both US Geographical Systems and NIMA. Important strategic points in the area as well as the feature vector data are obtained from NIMA.

---

*This research has been funded in part by NSF grants EEC-9529152 (IMSC ERC) and ITR-0082826, NASA/JPL contract nr. 961518, DARPA and USAF under agreement numbers F30602-00-01-0504 and F30602-99-1-0524, and unrestricted cash/equipment gifts from NCR, IBM, Intel and SUN.

Proceedings of the 27th VLDB Conference, Roma, Italy, 2001

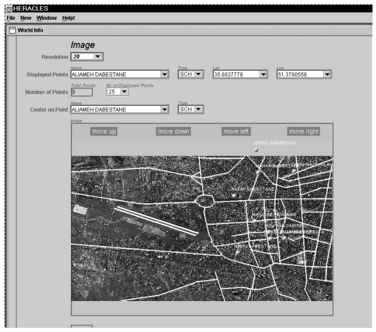

Figure 1: Satellite image of Tehran, Iran with major streets and runways superimposed on it

An interesting feature of the WorldInfo Assistant is its capability to provide and present detailed spatial information about a region. More specifically, it provides satellite images and maps with different resolutions, strategic points, and over 80 detailed vector data features such as railroads, runways, etc. for a region. Vector data and strategic points are superimposed as data layers on satellite images per user request. The detailed information about the points and features are also provided in a separate window. Figure 1 depicts a snapshot of the application that shows a satellite image of part of Tehran, Iran with major streets and two runways superimposed on the satellite images.

3 Efficiently Querying Moving Objects

In this demonstration, we consider an environment where the content of the moving object database does not need to be modified to reflect the movement of the objects. We term this environment as moving objects with predefined paths and schedules. An example application that we demonstrate is to query the location of trains moving on a railroad network. By storing the schedules of trains departures and arrivals, the locations of the stations and the vector data corresponding to the railroad network, we have enough information to query the location of any moving object (i.e., train) at any given time. The challenge is that queries of the type of finding the location of a train in a given time interval are time consuming. This is due to expensive spatial functions (e.g., the shortest path function) that need to be performed on large vector data as well as the temporal intersection function that need to be applied on large sets of time intervals.

One solution to reduce the query processing time of moving objects with predefined paths and schedules is to pre-compute the required information and materialize it using a moving object data model such as the 3D trajectory model. This is a feasible approach if we assume that different schedules, railroads and stations information are all local and over which we have full control. However, with our assumed distributed environment, the sources of information that we would like to access are autonomous and dynamic. That is, we do not have administrative control over them, cannot modify their structure, or write data to them. The sources can change their information without warning. Different sources may contain overlapping information or only fragments of desired data. Therefore, we investigated alternative distributed query plans to realize the integration of spatial and temporal information (e.g., network of the railroads and schedules of the trains) from distributed, heterogeneous sources.

We investigated traditional filter+semi-join plans by either applying the temporal filter first and then perform the spatial semi-join or vice versa. However, there are two main drawbacks with pure filter+semi-join plans. First, spatial filters (e.g., identifying all railroad segments that overlap with a given point) are computationally complex resulting in long local or remote query processing time. Second, temporal filters (identifying all intervals that overlap with a given interval) usually have poor selectivity due to the large range of intervals covered by each instance in the temporal source (e.g., schedule table). That is, many schedules usually intersect with any query interval. Thus, temporal filters cannot effectively reduce the amount of data transferred over the network. We can overcome the first obstacle by either performing a pre-computation step and then applying a less expensive function/filter, or delaying the spatial filter until we reduce the size of the spatial data (e.g., railroad vector data) significantly. We address the second obstacle by proposing a spatio-temporal filter (termed deviation filter), instead of temporal-only filters, which can also exploit the spatial characteristics of the data to improve the selectivity.

We integrated the path deviation query plan into the WorldInfo Assistant to support efficiently querying moving objects with predefined paths and schedules.

4 Conclusion

Information integration is becoming the core technology to extract and correlate data from web sources and new approaches to integrate sources with different data types are required. In this demonstration, we present the WorldInfo Assistant, an example of the technology that provides access to spatio-temporal information. We also present research issues in efficiently querying moving objects in a distributed environment within the WorldInfo Assistant framework.

References

[1] G. Barish, Y. Chen, D. DiPasquo, C. Knoblock, S. Minton, I. Muslea, and C. Shahabi, *Theater-loc: Using information integration technology to build virtual applications*, Proceedings of the International Conference on Data Engineering (ICDE) (2000).

[2] C. Knoblock, S. Minton, J.L. Ambite, M. Muslea, J. Oh, and M. Frank, *Mixed-initiative, multi-source information assistants*, The Tenth International World Wide Web Conference (WWW10), Hong Kong (2001).