

INTEGRATED MEDIA SYSTEMS CENTER

A National Science Foundation
Engineering Research Center at the
UNIVERSITY OF SOUTHERN CALIFORNIA

PRINCIPAL INVESTIGATORS

Prof. Craig Knoblock
knoblock@isi.edu (213) 740-8600

Prof. Cyrus Shahabi
cshahabi@usc.edu (213) 740-8600

Prof. Dennis McLeod
McLeod@usc.edu (213) 740-7285

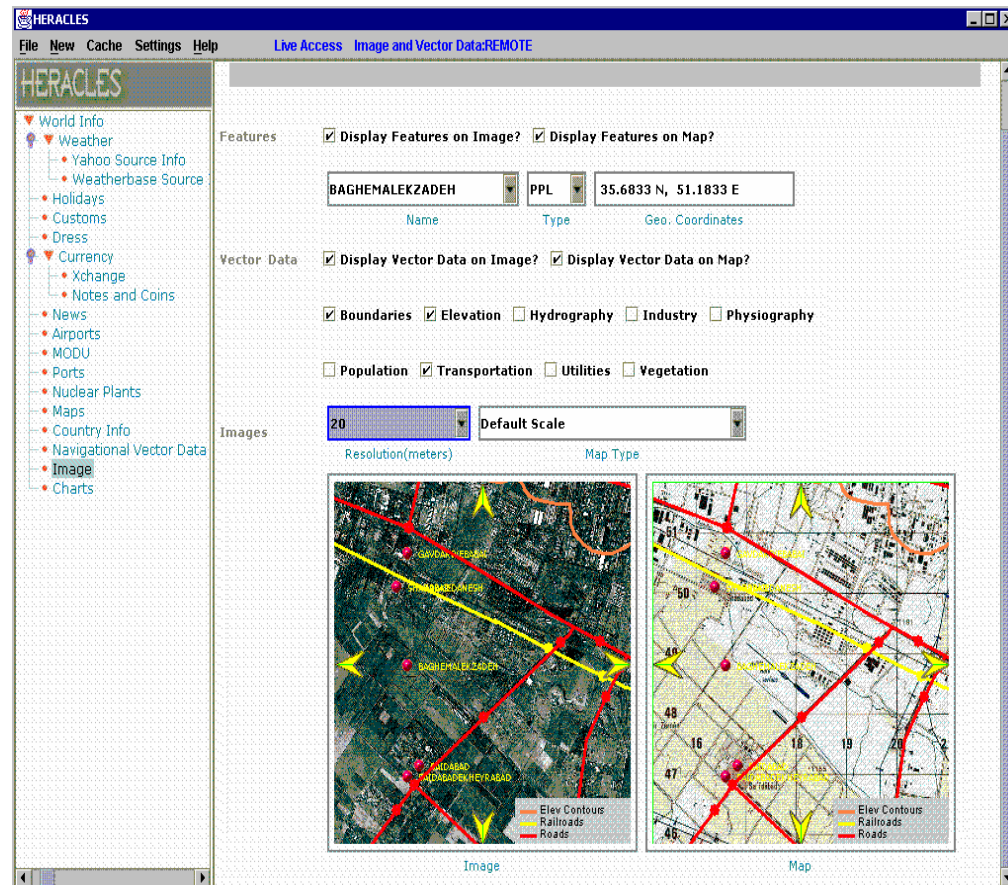
ISI STAFF

Jose Luis Ambite, Maria Muslea, Jean Oh

USC STUDENTS (EXPECTED DEGREES):

Ching-Chien Chen, Mohammad Kolahdouzan,
Mehdi Sharifzadeh, Snehal Thakkar (Ph.D.)

Data Management: The WorldInfo Assistant, Integration of Spatio-Temporal Information on the Internet



BRIEF DESCRIPTION OF TECHNOLOGY DEMONSTRATION

The WorldInfo Assistant demonstrates the seamless integration of heterogeneous data from multiple distributed sources, both traditional and web-based. Our research is built upon several important technologies: automatic information extraction, mediator-based planning and execution of queries, local data materialization, spatial data conflation, storage and retrieval of complex multimedia types, and autonomous plan execution and optimization. Specifically, we have built and deployed the WorldInfo Assistant demonstration, which illustrates how widely distributed and diversely formatted information can be easily queried, integrated, and represented. This information includes web-based data (such as news and weather from multiple websites) and traditional data (such as different resolution satellite imagery and maps, digital elevation information and vector data that represent various features like railroads, streets, runways, coastlines, power lines, buildings and populated areas).

UNIQUE OR DISTINGUISHING CHARACTERISTICS RELATIVE TO STATE-OF-THE-ART

- Information integration. Multimedia, spatio-temporal/vector data and digital elevation information storage/query/delivery. Information extraction. Local data materialization. Optimized query planning and execution. Spatio-temporal query support. K-Nearest Neighbor search. Geo-spatial data integration and conflation.

APPLICATIONS

- Integration of any semi-structured heterogeneous data sources.
- Accessibility of web-based sources similar to traditional databases.
- Integration of spatio-temporal information.
- Integration of various geo-spatial datasets.
- Any GIS application on the Web.

RECENT HIGHLIGHTS, LEVEL OF DEVELOPMENT, UPCOMING MILESTONES

- The WorldInfo Assistant: Java based application demonstrating underlying technologies.
- Investigating new distributed architecture to provide support for large amount of spatial and vector data for the World Info Assistant.
- Developing an information integration approach that utilizes conflation techniques to alleviate the spatial inconsistencies between diverse geo-spatial datasets.

UNDERLYING TECHNOLOGIES

- Automatic information extraction: facilitates the gathering of data from semi-structured sources (such as web pages).
- Mediator-based planning and execution: the primary mechanism for both optimized query planning and information integration among multiple data sources.
- Local materialization: dynamically materializing data at the mediator by analyzing users' query distribution, source structure, and update frequency.
- Query optimization: responsible for runtime optimization of information retrieval.
- Multimedia storage and retrieval: provides accessibility to unstructured data (such as video).
- Spatial query optimization: responsible for executing distributed spatial queries in an optimal manner.
- Spatio-temporal information storage/query/integration.
- Efficient vector data query/delivery.
- Autonomous optimized plan execution: enables the automatic maintenance of integrated information.
- K-Nearest Neighbor search: utilizing space embedding techniques to reduce the cost of computation of K-Nearest Neighbor algorithms.
- Geo-spatial data conflation: automatically and accurately integrating different geo-spatial datasets that cover the overlapping regions.

LIST OF PUBLICATIONS, REFERENCES, URLs

- Cyrus Shahabi, Ching-Chien Chen, Snehal Thakkar and Craig A. Knoblock, Automatically Annotating and Integrating Spatial Datasets. *Submitted to the 8th International Symposium on Spatial and Temporal Databases, Santorini island, Greece, July 24-27, 2003*
- Cyrus Shahabi, Mohammad R. Kolahehdouzan and Mehdi Sharifzadeh, A Road Network Embedding Technique for K-Nearest Neighbor Search in Moving Object Databases, The 10th ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS), McLean, VA, U.S.A. , November 2002
- Cyrus Shahabi, Mohammad R. Kolahehdouzan, Snehal Thakkar, Jose Luis Ambite and Craig A. Knoblock, Efficiently Querying Moving Objects with Pre-defined Paths in a Distributed Environment, *The Ninth ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS)*, Atlanta, Georgia, U.S.A. , November 2001
- Jose Luis Ambite, Craig A. Knoblock, Mohammad R. Kolahehdouzan, Maria Muslea, Cyrus Shahabi and Snehal Thakkar, The WorldInfo Assistant: Spatio-Temporal Information Integration on the Web , *27th International Conference on Very Large Data Bases (VLDB2001)* , Rome, Italy , September 2001
- Craig A. Knoblock, Jose Luis Ambite, Steven Minton, Cyrus Shahabi, Mohammad Kolahehdouzan, Maria Muslea, Jean Oh, and Snehal Thakkar, Integrating the World: The WorldInfo Assistant , *International Conference on Artificial Intelligence (IC-AI)*, Las Vegas, June 2001
- Greg Barish, Yi-Shin Chen, Dan DiPasquo, Craig A. Knoblock, Steven Minton, Ion Muslea and Cyrus Shahabi, *TheaterLoc: Using Information Integration Technology to Rapidly Build Virtual Applications*, in Proceedings of the 16th International Conference on Data Engineering, 2000.
- Greg Barish, Craig A. Knoblock, Yi-Shin Chen, Steven Minton, Andrew Philpot, Cyrus Shahabi, *The TheaterLoc Virtual Application*, Twelfth Innovative Applications of AI Conference, 2000.
- Greg Barish, Craig Knoblock, Yi-Shin Chen, Steven Minton, Andrew Philpot and Cyrus Shahabi, *TheaterLoc: A Case Study in Building An Information Integration Application*, To appear in IJCAI Workshop on Intelligent Information Integration, 1999
- Cyrus Shahabi and Maytham Safar. *Efficient Retrieval and Spatial Querying of 2-D Objects*". To appear in the IEEE Multimedia Systems 99, Florence, Italy, June 1999.
- Cyrus Shahabi, Latifur Khan, Dennis Mcleod and Vishal Shah. *Run-Time Optimization of Join Queries for Distributed Databases over the Internet*. In Proceedings of the Communication Networks and Distributed Systems Modeling and Simulation (CNDS '99), San Francisco, CA, January 99.
- Ion Muslea, Steve Minton, Craig Knoblock. *A Hierarchical Approach to Wrapper Induction*. 3rd Conference on Autonomous Agents, Seattle, WA, 1999.
- Craig A. Knoblock, Steven Minton, Jose Luis Ambite, Naveen Ashish, Pragnesh Jay Modi, Ion Muslea, Andrew Philpot, and Sheila Tejada. *Modeling Web Sources for Information Integration*. Proceedings of the Fifteenth National Conference on Artificial Intelligence, Madison, WI, 1998.
- Ion Muslea, Steve Minton, Craig Knoblock. *Wrapper Induction for Semistructured, Web-based Information Sources*. Proceedings of the Conference on Automatic Learning and Discovery CONALD-98, Pittsburgh, PA, 1998.
- Naveen Ashish, Craig A. Knoblock and Cyrus Shahabi. *Semantic Caching for Information Integration*. AAAI Workshop on AI and Data Integration, July 1998, Madison, WI
- <http://infolab.usc.edu/>
- <http://www.isi.edu/info-agents/>

For additional information, please contact the Principal Investigator listed above via email, or contact
Isaac Maya, Ph.D., P.E.
Director, Industry and Technology Transfer Programs imaya@imsc.usc.edu
213-740-2592

Integrated Media Systems Center
3740 McClintock Avenue, Suite 131
Los Angeles, CA 90089-2561
213-740-8931 (fax)

For additional information on the Integrated Media Systems Center (IMSC), please visit our Web site at <http://imsc.usc.edu>

IM-WorldInfo Assistant