

Mourad Ouzzani, Research Associate Professor, Purdue University

CONTACT INFORMATION

155 S Grant Street
Cyber Center
Purdue University
West Lafayette, Indiana, 47907-2108 USA

Voice: (765) 494-7917
Fax: (765) 496-2275
E-mail: mourad@cs.purdue.edu
Web: <http://www.cs.purdue.edu/homes/mourad>

RESEARCH INTERESTS

My research interests lie in the fields of data management and service computing and how they enable discovery and innovation. More specifically, I am interested in issues related to data integration (schema matching, schema mapping, and record linkage), data quality (repair and data quality rules), spatio temporal data management (constraints awareness and massively parallel), database systems for scientific data (annotation and provenance), and service oriented computing (efficient querying and composition, access control, and service oriented large scale cyberinfrastructures). I am involved in several multidisciplinary projects including a web-based system to manage experiments and analysis workflows in high throughput ionomics, a pipeline for metabolomic data, a plant science literature social network, a virtual community for global engineering, and public health surveillance.

EDUCATION

Ph.D., Computer Science, **Virginia Tech**, Virginia, USA May 2004
M.S., Computer Science, **USTHB**, Algiers, Algeria April 1995
B.S., Computer Science (*summa cum laude*), **USTHB**, Algiers, Algeria June 1991

EMPLOYMENT

Purdue University
Research Associate Professor, Cyber Center, Discovery Park July 2010 - Present
Research Assistant Professor, Cyber Center, Discovery Park March 2006 - June 2010
Visiting Assistant Professor, Dept. of Computer Science, June 2004 - March 2006

GRANTS

- *The iHUB: A Collaborative International Network for Ionomics* (CoPI) National Science Foundation, \$530,472, 2010-2014.
- *Web-based Monitoring in Mothers of Late-Preterm Infants* (CoPI) International Lactation Consultant Association, \$10,000, 2010-2014.
- *Commigrate – A Community-based Data Integration System* (PI). National Science Foundation, \$498,373, 2009-2012.
- *Mass Informatics of Two Dimensional Gas Chromatography Time-of-flight Mass Spectrometry* (CoPI), National Institutes of Health, \$1,405,380, 2009-2012.
- *Center of Excellence for Command, Control and Interoperability, Visualization Sciences and Education Lead* (CoPI), Department of Homeland Security, \$15M, 2009-2014.
- *Preference- And Context-Aware Query Processing for Location-based Database Servers* (PI), National Science Foundation, \$192,868, 2008-2011.
- *Purdue University Regional Visualization and Analytics Center* (CoPI), Department of Homeland Security, \$1.4M, 2005-2009.
- *The Purdue Biosamples PLM-Hub: Development of a Web Community Resource for Product Lifecycle Management of Life Science Products* (CoPI), New Manufacturing Economy Booster Plan Award, Discovery Park and the Office of the Vice President for Research, Purdue University, \$40,000, 2007-2008.
- *Integration and Analysis of Health Data for Indiana Pandemic Influenza Plans* (CoPI), Indiana State Department of Health, \$51,200, 2006.
- *The Cyber Center, Discovery Park* (CoPI), The Lilly Endowment, \$2.5M, 2005-2008.
- *CBR3 Pilot Grant – Adapting the Pipeline to Analyze Clinical Proteomics Data* (Collaborator), Purdue University, \$50,000, 2007.

HONORS,
SCHOLARSHIPS, AND
AWARDS

- Elevated to Senior Member for ACM, 2009.
- Purdue Seeds of Success, VACCINE: Visual Analytics for Command, Control, and Interoperability Environments, 2009.
- Nominated by the Virginia Tech CS Department for the Outstanding Graduate Student Award 2004-2005, January 2005.
- Awarded the Japan Society for the Promotion of Science Postdoctoral Fellowship, August 2004.
- Nominated by the Virginia Tech CS Department for the Graduate Student of the Month Award, April 2003.
- Outstanding reviewer by the IEEE Internet Computing, November 2002.
- Nominated by the CS Department for the Outstanding Graduate Student Award, 2001.
- USENIX scholar (\$21,000), 2000.
- Three year scholarship for PhD studies, QUT, Brisbane, Australia, 1998.
- Awarded a four year scholarship for graduate studies in France, 1991.
- Outstanding High School Student National Award by the Presidency of Algeria, June 1986.

PUBLICATIONS

Journals

1. Catherine P. Riley, Erik S. Gough, Jing He, Shrinivas S. Jandhyala, Brad Kennedy, Seza Orcun, Mourad Ouzzani, Charles Buck, Ali M. Roumani and Xiang Zhang. The Proteome Discovery Pipeline - A Data Analysis Pipeline for Mass Spectrometry-Based Differential Proteomics Discovery. *The Open Proteomics Journal*, 2010, 3, 8-19
2. Qi Yu, Manjeet Rege, Athman Bouguettaya, Brahim Medjahed, and Mourad Ouzzani. A two-phase framework for quality-aware Web service selection. To appear in *Service Oriented Computing and Applications*.
3. R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. Cleveland, S. Grannis, M. Wade, and D Ebert. "A Visual Analytics Approach to Understanding Spatiotemporal Hotspots" Accepted for publication in the *IEEE Transactions on Visualization & Computer Graphics*, 2009.
4. R. Hafen, D.E Anderson, W.S. Cleveland, R. Maciejewski, D. Ebert, A. Abusalah, M. Yakout, M. Ouzzani, and S.J. Grannis. Syndromic surveillance: STL for modeling, visualizing, and monitoring disease counts *BMC Medical Informatics and Decision Making*, 9(21), April 2009.
5. A. Elmagarmid, A. Samuel, and M. Ouzzani. Community Cyber Infrastructure Enabled Discovery in Science and Engineering. *Computing in Science & Engineering*, August/September 2008.
6. I. Baxter, M. Ouzzani, S. Orcun, B. Kennedy, S.S. Jandhyala, and D. Salt. Purdue Ionomics Information Management System (PIIMS). An Integrated Functional Genomics Platform. *Plant Physiology*, 143(2), February 2007.
7. X. Zhang, M.J. Asara, J. Adamec, M. Ouzzani, and A.K. Elmagarmid. Pre-Processing in Liquid Chromatography-Mass Spectrometry based Proteomics. *Bioinformatics*, 21(21), Oxford University Press, November 2005.
8. M. Ouzzani and A. Bouguettaya. Query Processing and Optimization on the Web. *Distributed and Parallel Databases, an Int. Journal*, 15(3), May 2004.
9. M. Ouzzani and A. Bouguettaya. Efficient Access to Web Services. *IEEE Internet Computing*, 37(3), March 2004.
10. B. Medjahed, A. Rezgui, A. Bouguettaya, and M. Ouzzani. Infrastructure for E-Government Web Services. *IEEE Internet Computing*, 7(1), January/February 2003.
11. A. Bouguettaya, M. Ouzzani, B. Medjahed, and J. Cameron. Managing Government Databases. *IEEE Computer*, 34(2), February 2001.
12. A. Bouguettaya, B. Benatallah, L. Hendra, M. Ouzzani, and J. Beard. Supporting Dynamic Interactions among Web-Based Information Sources. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 12(5), September/October 2000.

13. M. Ouzzani, B. Benatallah, and A. Bouguettaya. Ontological Approach for Information Discovery in Internet Databases. *Distributed and Parallel Databases, an Int. Journal*, 8(3), July 2000.
14. A. Bouguettaya, B. Benatallah, M. Ouzzani, and L. Hendra. WebFINDIT - An Architecture and System for Querying Web-accessible Databases. *IEEE Internet Computing*, 3(4), July/August 1999.

Conferences & Workshops

15. Mohamed Yakout, Ahmed K. Elmagarmid, Hazem Elmeleegy, Mourad Ouzzani, Alan Qi Behavior Based Record Linkage *PVLDB, Singapore, September 2010*
16. Hazem Elmeleegy, Mourad Ouzzani, Ahmed Elmagarmid, and Ahmad Abusalah Preserving Privacy and Fairness in Peer-to-Peer Data Integration *ACM SIGMOD International Conference on Management of Data, SIGMOD 2010, Indianapolis, IN, USA, June 2010*
17. Mohamed Yakout, Ahmed K. Elmagarmid, Jennifer Neville, Mourad Ouzzani GDR: A System for Guided Data Repair (demo) *SIGMOD 2010, Indianapolis, Indiana, USA*
18. M. Y. Eltabakh, W. G. Aref, A.K. Elmagarmid, Y. Laura-Silva, and M. Ouzzani. Supporting Real-world Activities in Database Management Systems *26th International Conference on Data Engineering (ICDE 2010)*, March 1-6, 2010, Long Beach, CA, USA.
19. Nilothpal Talukder, Mourad Ouzzani, Ahmed Elmagarmid, Hazem Elmeleegy, and Mohamed Yakout Privometer: Privacy Protection in Social Networks *2nd International Workshop on Modeling, Managing and Mining of Evolving Social Networks (M3SN 2010) in conjunction with ICDE 2010, Long Beach, CA, 2010*
20. M. Y. Eltabakh, W. G. Aref, A.K. Elmagarmid, M. Ouzzani, and Y. Laura-Silva. Supporting Annotations on Relations *12th International Conference on Extending Database Technology (EDBT 2009)*, March 23-26 2009, Saint-Petersburg, Russia.
21. M. Ouzzani, J.H. Bohn, D. Datta, E.A. Groll, E.D. Hirlleman, and J. Lucena. GlobalHUB: A Virtual Community for Global Engineering Education, Research, and Collaboration *7th Annual ASEE Global Colloquium on Engineering Education, Cape Town, South Africa, October 19-23, 2008*.
22. R. Maciejewski, Stephen Rudolph, R. Hafen, Ahmad Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, Shaun J. Grannis, Michael Wade, D. S. Ebert. Understanding Syndromic Hotspots - A Visual Analytics Approach. *IEEE Symposium on Visual Analytics Science and Technology (VAST 2008)*, pp. 35-42, October 19-24, 2008.
23. F. Paci, M. Ouzzani, and M. Mecella. Verification of Access Control Requirements in Web Services Choreography. *2008 IEEE International Conference on Services Computing (SCC 2008)*, Honolulu, Hawaii, USA, July 8-11, 2008.
24. W. Aref and M. Ouzzani Data Management Challenges for Computational Transportation (invited) *First International Workshop on Computational Transportation Science, July 21, 2008 - Trinity College Dublin, Ireland*.
25. H. Elmeleegy, M. Ouzzani, and A.K. Elmagarmid Usage-based Schema Matching. *Proc. of the 24th Int. Conf. on Data Engineering (ICDE)*, Cancun, Mexico, April 2008.
26. M.Y. Eltabakh, M. Ouzzani, W.G. Aref, A.K. Elmagarmid, Y. Laura-Silva, D. Salt, and I. Baxter Managing Biological Data using bdbms (demo). *Proc. of the 24th Int. Conf. on Data Engineering (ICDE)*, Cancun, Mexico, April 2008.
27. R. Maciejewski, Benjamin Tyner, Y. Jang, C. Zheng, R. Nehme, D. Ebert, M. Ouzzani, L. Glickman, and S. Grannis LAHVA: Linked Animal-Human Health Visual Analytics. *IEEE Symposium on Visual Analytics Science and Technology 2007*, Sacramento, CA, October/November, 2007.

28. D. E. Anderson, C. Zheng, R. Maciejewski, R. Hafen, W. S. Cleveland, D. S. Ebert, M. Ouzzani, and S.J. Grannis. STL and Local Regression for Modeling Disease Surveillance Counts. *2007 International Society for Disease Surveillance Sixth Annual Conference*, Indianapolis, IN, October 2007.
29. M.Y. Eltabakh, M. Ouzzani, and W.G. Aref Duplicate Elimination in Space-partitioning Tree Indexes. *Proc. of the 19th International Conference on Scientific and Statistical Database Management (SSDBM 2007)*, July 2007, Banff, Canada.
30. M.Y. Eltabakh, M. Ouzzani, and W.G. Aref bdbms - A Database Management System for Biological Data. *Proc. of the Third Biennial Conference on Innovative Data Systems Research (CIDR 2007)*, Asilomar, CA, USA, January 2007.
31. M.Y. Eltabakh, W.G. Aref, M. Ouzzani, and M.H. Ali Discovering Consensus Patterns in Biological Databases. *Proc. of the VLDB Workshop on Data Mining on Bioinformatics*, Seoul, Korea, September 2006.
32. H. Elmongui, M. Ouzzani, and W.G. Aref Challenges in Spatio-temporal Stream Query Optimization. *Proc. of the 5th ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE'06)*, Chicago, IL, USA, June 2006.
33. M. Mecella, M. Ouzzani, F. Paci, and E. Bertino Access Control Enforcement for Conversation-based Web Services. *Proc. of the 15th International World Wide Web Conference*, Edinburgh, UK, May 2006.
34. D. Ebert, Y.-H. Lu, E. Delp, W. Cleveland, A. Chaturvedi, A.K. Elmagarmid, and M. Ouzzani. Resource- and Task-Driven Visualization Adaptation. *HI'06 Workshop: Information Visualization and Interaction Techniques for Collaboration across Multiple Displays*, April 2006.
35. A. Bouguettaya, B. Medjahed, A. Rezgui, M. Ouzzani, X. Liu, and Q. Yu WebDG - A Platform for E-Government Web Services. *Proc. of the ER 2004 Workshop on Digital Government: Systems and Technologies*, Shanghai, China, November 2004.
36. A. Bouguettaya, A. K. Elmagarmid, M. Ouzzani, and B. Medjahed. Database Middleware for Distributed Ontologies in State and Federal Family & Social Services. *DG.O 2004*, Seattle, Washington, USA.
37. M. Ouzzani and A. Bouguettaya. A Query Paradigm for Web Services. *Proc. of the 1st Int. Conf. on Web Services*, Las Vegas, June 2003.
38. M. Ouzzani, A. Bouguettaya, and B. Medjahed. Optimized Querying of E-Government Services. *Proc. of the 2003 Nat. Conf. on Digital Government Research*, Boston, May 2003.
39. B. Medjahed, A. Bouguettaya, and M. Ouzzani. Semantic Web Enabled E-Government Services. *Proc. of the 2003 Nat. Conf. on Digital Government Research*, Boston, May 2003.
40. A. Rezgui, M. Ouzzani, A. Bouguettaya, and B. Medjahed. Preserving Privacy in Web Services. *Proc. of the 4th Int. ACM Workshop on Web Information and Data Management*, Fairfax, November 2002.
41. A. Bouguettaya, B. Medjahed, A. Rezgui, M. Ouzzani, and Z. Wen. Privacy Preserving Composition of Government Web Services (Demo Paper). *Proc. of the 2002 Nat. Conf. on Digital Government Research*, Los Angeles, May 2002.
42. B. Medjahed, M. Ouzzani, and A. Bouguettaya. Using Web Services in E-Government Applications. *Proc. of the 2002 Nat. Conf. on Digital Government Research*, May 2002.
43. A. Bouguettaya, A. K. Elmagarmid, B. Medjahed, and M. Ouzzani. Ontology-based Support for Digital Government. *Proc. of 27th Int. Conf. on Very Large Data Bases (VLDB)*, Rome, Italy, September 2001.
44. A. Bouguettaya, A. Elmagarmid, B. Medjahed, and M. Ouzzani. A Web-based Architecture for Government Databases and Services. *Proc. of the 2001 Nat. Conf. on Digital Government Research*, Los Angeles, May 2001.

45. A. Bouguettaya, B. Benatallah, L. Hendra, J. Beard, K. Smith, and M. Ouzzani. World Wide Database - Integrating the Web, CORBA, and Databases (Demo). *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, Philadelphia, June 1999.
46. A. Bouguettaya, B. Benatallah, M. Ouzzani, and L. Hendra. Using Java and CORBA for Implementing Internet Databases. *Proc. of the 15th Int. Conf. on Data Engineering (ICDE)*, Sydney, Australia, March 1999.
47. M. Ouzzani and M. Atroun. Towards more Concurrency in Multidatabases. *Proc. of 7th Rencontres Francophones du Parallélisme*, Mons, Belgium, May 1995.
48. M. Ouzzani, M. Atroun, and N. Belkhodja. A Top-Down Approach to Two-Level Serializability. *Proc. of the 20th Int. Conf. on Very Large Data Bases (VLDB)*, Santiago. Chile, September 1994.
49. M. Ouzzani, M.A. Atroun, , and N.L. Belkhodja. Transaction Management in Multidatabase Systems. *Proc. of the Journées Int. des Systèmes Informatiques (JISI'94)*, Tunis, May 1994.

Book Chapters

50. M. Ouzzani, A. Bouguettaya, and A. Elmagarmid. WS-Query – A Framework to Efficiently Query Semantic Web Service *In Emergent Web Intelligence (editors)*, Springer-Verlag, to appear.
51. B. Medjahed, M. Ouzzani, and A. Elmagarmid. Generalization of ACID Properties. *Encyclopedia of Database Systems*, L. Liu and T. Oszu (Eds), ISBN 978-0-387-35544-3 Springer-Verlag, April 2009.
52. M. Ouzzani, B. Medjahed, A. Elmagarmid. Correctness Criteria beyond Serializability. *Encyclopedia of Database Systems*, L. Liu and T. Oszu (Eds), ISBN 978-0-387-35544-3 Springer-Verlag, April 2009.
53. X. Zhang, S. Orcun, M. Ouzzani, and C. Oh. Mass Informatics in Differential Proteomics. *Encyclopedia of Data Warehousing and Mining - 2nd Edition*, August 2008.
54. A. Bouguettaya, A. Rezgui, B. Medjahed, and M. Ouzzani. Internet Computing Support for Digital Government, Chapter 3, *in The Practical Handbook of Internet Computing*, CRC Press, 2004.
55. A. Bouguettaya, M. Ouzzani, B. Medjahed, and A. Elmagarmid. Supporting Data and Services Access in Digital Government Environments. *Kluwer Academic Publishers*, 2002.
56. A. Bouguettaya, B. Benatallah, B. Medjahed, M. Ouzzani, and L. Hendra. Adaptive Web-based Communities. *IDEA Group Publishing*, 2002.
57. A. Bouguettaya, B. Medjahed, M. Ouzzani, and Y. Meng. Ubiquitous Access to Web Databases, chapter 9. *IDEA Group Publishing*, 2002.

Posters

58. M. Ouzzani, Ivan Baxter, Gemez Marshal, Maged Zereba, and D. Salt. Cyber-enabled Discovery - The Case of the Purdue Ionomics Information Management System. 2008 Microsoft eScience Workshop, Indianapolis, IN, December 7-9th, 2008.
59. R. Maciejewski, Y. Jang, D. S. Ebert, W. S. Cleveland, M. Ouzzani, S. J. Grannis, L. T. Glickman. LAHVA: Linked Animal-Human Health Visual Analytics. *2007 International Society for Disease Surveillance, Sixth Annual Conference*, Indianapolis, IN, October 2007.
60. X. Zhang, S. Orcun, S. Jandhyla, M. Ouzzani, I. Baxter, D. Salt, A.K. Elmagarmid, F.E. Rgnier, V. Jo Davisson, and J. Pekny. A Cyber Infrastructure for Biomarker Discovery in Mass Spectrometry-Based Proteomics. *54th ASMS Conference on Mass Spectrometry- Bioinformatics III*, Seattle, WA, May 28 - June 1, 2006.
61. M. Eltabakh, M. Ouzzani, W. G. Aref, and A.K. Elmagarmid. bdbms: A Database Engine for Biological Data. *Third Annual Indiana Bioinformatics Conference*, Indianapolis, IN, May 2006.

62. A. Abusalah, M. Ouzzani, and A. K. Elmagarmid. Public Health Surveillance Systems: Overview, Limitations and Challenges. *Midwest Database Research Symposium*, Urbana Champaign, IL, April 2006.
63. M. Mecella, M. Ouzzani, F. Paci, and E. Bertino. Access Control Enforcement for Conversation-based Web Services. *Midwest Database Research Symposium*, Urbana Champaign, IL, April 2006.
64. M. Ouzzani, W.G. Aref, A.K. Elmagarmid, H. Elmeleegy, M. Eltabakh, S. Prabhakar, M. Zhang, X. Zhang. Online Data System for Comparative Proteomics. *Midwest Database Research Symposium*, Chicago, IL, April 2005.
65. X. Zhang, M.J. Asara, J. Adamec, and M. Ouzzani. Data Pre-Processing in LC-MS based Proteomics. *USHUPO 2005 Symposium*, Washington D.C. March, 2005.

Others

66. M. Ouzzani, W.G. Aref, E. Bertino, A.C. Catlin, C.W. Clifton, W. Hon, A.K. Elmagarmid, A. Ghafoor, S.E. Hambrusch, S. Prabhakar, J.S. Vitter, and X. Zhang. The Indiana Center for Database Systems at Purdue University. *SIGMOD Record*, 34(2), June 2005.

RESEARCH
EXPERIENCE

Purdue University

June 2004 - present

Data Integration and Data Quality

The number of information sources used in science, engineering, and industry is sharply increasing, as well as the need for large-scale data integration solutions to enable effective access to these sources. Along with several collaborators and students we are looking at different issues of data integration and data quality including:

- **Schema matching and mapping** – While traditional schema matching and mapping techniques rely either on schema-based or instance-based information, we propose *usage-based* schema matching and mapping techniques. These techniques go beyond the traditional information sources, and bases their decisions on information extracted from the data usage information like query logs.
- **Record linkage** – Linking records across different sources has many applications and is key to data integration. We propose to link records based on on the *behavior* of these entities as reflected in the transaction logs they generate. For example, analyzing the purchase histories of two a priori different customers can indicate that they actually represent the same person.
- **Data Cleaning and Repair** – Data to be integrated is often dirty and contains many errors. We are working on an integrated adaptive feedback system to improve the quality of the data by making the best use of users feedback. Most existing techniques rely on simple duplicate elimination and do not involve users until the end of the entire data cleaning process.

Web Services and Mashups

Web services are becoming a fundamental building block for Web applications. I am involved in research related to different aspects of web services and service oriented computing, in particular:

- **Access Control** – We are looking at authorization control mechanisms that take into account the conversational aspect of Web services. We proposed a novel model for access control in Web services that is suitable when interactions between the client and the Web service are conversational and long-running. Our conversation-based access control model allows service providers to limit how much knowledge clients have on their access control policies in terms of the credentials specified in these policies.
- **Social Network-based Mashups** – The widespread adoption of Web 2.0 technologies have created unique opportunities to leverage the wisdom of the crowd in enabling more meaningful analysis tool interactions and knowledge extraction on community cyberinfrastructures. We are working on a social network-like cyberinfrastructure where scientists share their experiences and expertise in using different analysis tools with the goal of enabling a more effective use

and integration (mashups) of these tools. This social-control based cyberinfrastructure will act as a hub that contains information on different analysis tools and their discovery options, as well as their corresponding social information in terms of quality, trust, and usage information and statistics to help scientists select and integrate (mashup) the right tools for their complex analyses.

- **Knowledge-Based Dynamic Maintenance System** – The Navy’s next generation of vessels will be characterized by communication and power systems of significantly increased complexity and capability. It is crucial that the extensive shore-based expertise associated with these new ships is available to support crews as they perform their missions at sea. The project aims at helping the sailor faced with operational failure and maintenance troubleshooting. I was in charge of enabling the integration with other Navy projects based on the use of Web services and enterprise integration standards as well supervising graduate students and reporting to our sponsors. The system produced from this project has been handed over to the Navy.

Database System Support for Scientific Databases

- **bdbms** – Biological databases are essential to biological experimentation and analysis and biologists are increasingly using databases for storing and managing their data. Life sciences are a case in point where biological databases have become essential to experimentation and analysis. Database technology has not kept pace with the proliferation and specific requirements of biological databases. The goal of this project is to design and prototype bdbms, an extensible database management system that fits the needs of biological databases. More specifically, (1) Annotation and provenance management including storage, indexing, and querying of annotation and provenance as first class objects in bdbms, (2) Local dependency tracking to track the dependencies and derivations among data items, and (3) Update authorization to support data curation via content-based authorization, in contrast to identity-based authorization.
- **Supporting Real world Activities** – Databases are integral to many application domains in which the cycle of processing the data is complex and may involve real-world activities that are external to the database like wet-lab experiments and collecting instrument readings. Thus, an database update may render dependent data items invalid or inconsistent until the real-world activities involved in deriving these items are re-performed and the output results are reflected back into the database. The presence of these long delays between the updates along with the need for the intermediate results to be instantly available for querying makes integrating these real-world activities within the database systems a challenging task. We are addressing these challenges through techniques that reflect updates immediately into the database while keeping track of the dependent and potentially invalid data items until they are re-validated and reflecting their status in query results.
- **Intelligent Data System for Mass Spectrometry** – In this project, we proposed a novel strategy for analyzing data from mass spectrometers, called Intelligent Data Systems for Mass Spectrometry (IDSMS). IDSMS acquires data only for interesting analytes and allows the online analysis of this data. The IDSMS provides feedbacks to the instruments at different stages to change its mode and modify its parameters on the fly. This continuous re-tuning of the instrument through real time data analysis and feedback assures that there is no wasted data collection or wasted samples. It also shortens the time to achieve a meaningful answer for the analysis by eliminating most of the re-runs typically required in existing techniques. I was in charge of designing the architecture of the IDSMS and the coordination of the project tasks and progress.

Spatio Temporal Database Systems

- **Preference- And Context-Aware Query Processing for Location-based Database Servers** – The processing of location-based queries is usually isolated from various concepts of user preferences and/or context. For example, when a user looks for either a clinic, a restaurant, or an evacuation route, the user actually wants to find the best one according to the user preferences and context. Existing location-based query processors reduce the meaning

of best to be only the closest clinic, restaurant, or evacuation route in terms of pre-computed distances. In this project, we aim to enable practical realization of location-based services by embedding various forms of preferences and context into the core processing of location-based queries. We do not define new location-based queries, instead, we want to redefine the answer of existing traditional location-based queries by taking into consideration user preferences, user context, environmental context, and database-specific context.

- **Massively Parallel Spatio Temporal Data Management** – The ubiquity of spatiotemporal data, location detection devices, and location services warrant the use of massively parallel architectures to process the vast amounts of moving objects and location-based services and queries. We are looking at new issues in terms of architecture, computing model, query processing and optimization, and service-based interface for dealing with large amount of streaming data and continuous queries. We are looking at ways to leverage Map-Reduce to support such environment.

Cyberinfrastructure and Databases Support in Life Sciences

- **Purdue Ionomics Information Management System (PiiMS)** – We are developing PiiMS (www.ionomicshub.org) to deal with high throughput elemental profiling (ionomics) data. PiiMS provides integrated workflow control, formalized metadata and data collection, storage and analysis to facilitate high-throughput phenotyping, along with integrated tools for data search, retrieval and visualization for hypothesis development. PiiMS is deployed as a web-enabled system, allowing integration of distributed workflow processes, and the open source data analysis by numerous laboratories. In this project, I am taking a leadership role in defining the architecture, managing a group of professional programmers, and making sure that the scientists' requirements and needs are met. Three different organisms are being now supported, namely Arabidopsis Thaliana, yeast, and rice, through three different instances involving thousands of samples and experiments. This system is currently in full production use in Dr. David Salt's lab at Purdue University and is being used to serve several other labs around the US.
- **Metabolomics Profiling Pipeline** – High-throughput (HT) omics analysis platforms are the technologies of choice for analyzing complex mixtures derived from biological fluids such as blood and urine. Analyzing these complex mixtures has the potential to facilitate the development of preventive, predictive and personalized medicine markets in health and wellness as well as specific disease areas. One such powerful HT approach is two dimensional gas chromatography - mass spectrometry (GCxGC/TOF-MS). We propose to develop a systematic approach for analysis of metabolomic data generated from a GCxGC/TOF-MS instrument, in the form of a Metabolomics Profiling Pipeline (MetPP). We are developing multiple data mining algorithms that will be integrated into a single web-based application with interactive visual data mining. The integrated Informatics tools will not only significantly improve analysis speed but will also reduce human error. This solution will enable routine use of GCxGC/TOF-MS in analyses of clinically relevant biological samples.
- **Purdue Discovery Pipeline (PDP)** – The Purdue discovery pipeline (www.purdue.edu/dp/proteomics) support proteomics and metabolomics research through a web based data analysis pipeline. This pipeline accepts experimental data generated by mass spectrometers; then performs sequential data analysis routines. The web based infrastructure allows researchers to review the output generated by the data analysis routines in the form of reports and visualization through interactive graphs and charts. Laboratory administrators can manage laboratory inputs and the lifecycle of experiments in the pipeline through re-submission of data or alteration of input parameters of analysis routines. I have taken a leadership role in building this system based on a service oriented architecture, managing a group of professional programmers, and making sure that the scientists' requirements and needs are met.
- **OMICS Knowledge Based for Cancer Care Engineering** – The major goal of the

Cancer Care Engineering (CCE) hierarchy of projects is the application of principles and methodologies of systems engineering to the cancer problem. Cancer is viewed as a system that can be mathematically modeled for predictive behavior from the development of cancer to the determination and delivery of cancer care. The OMICS Knowledge Base system is designed to support the building, evaluation and refinement of statistical models based on OMICS data. The system will provide the needed underlying functions to support the storage, access, query, processing, extraction, analysis, modeling, and visualization of the vast quantities and complex representations of OMICS-related data. I worked on the metadata and workflow aspects which are critical to support the above mentioned functionalities.

- **Building the EcoliHub Community** – This project aims at building a one-stop-shopping for the E. coli community information resources, EcoliHub. It uses existing knowledge and enable new discoveries for a deeper understanding of life processes. These tools and the advances they allow will greatly impact human health both through application to pathogenic bacteria, especially enteropathogenic E. coli, Shigella, and Salmonella species, and because many cellular processes are universal. EcoliHub relies heavily on the use of service oriented architecture to integrate all available data from several online databases with ecoli (-related) data. I was initially involved in building a prototype system to access various E. coli databases using web services and in managing the project large team of programmers and graduate students.

Virtual Communities

- **GlobalHUB - Advancing Global Engineering (www.globalhub.org)** – I am the technical director of GlobalHUB, the first virtual organization dedicated to global engineering education and research. Our aim is to provide a premiere platform where students, educators, researchers, and practitioners collaborate on global design projects, access and contribute resources related to global engineering, and run and contribute design and simulations tools. Being an organic organization and using Web 2.0 technologies, GlobalHUB allows its users to actively participate in its evolution. GlobalHUB users can (i) access and contribute content, (ii) connect with other peers, (iii) form global collaborative teams, (iv) run CAD and simulations tools out of their browsers, and (v) participate through a feedback system to the evolution of GlobalHUB. The group feature offers a wiki space, a discussion forum, direct access to group specific resources, and intensive usage statistics. GlobalHUB has been now accessed from more than 140 countries from over all the world with more than 2,000 active users and more than 10,00 overall visitors.
- **LitCloud - A Literature Social Network** – LitCloud is a social network for scientists to share , tag, and comment on literature entries Most scientists will miss a significant number of relevant literature as more than 2000 papers are published every month in plant science alone. LitCloud will allow (1) accessing the latest and most reliable ionomics literature, (2) sharing data between LitCloud and your RSS reader, (3) tracking entries along with their tags, comments, and starring actions, (4) import/Export RSS feeds along with tags, comments, and starring actions, and (5) enabling collaborative & rapid filtering and refinement of literature.

Public Health Monitoring and Management

This is part of the Purdue University Regional Visualization and Analytics Center (PURVAC - www.purvac.org). We are looking at different aspects of public health surveillance and how existing capabilities can be enhanced through better database support, data integration, and visual analytics capabilities. We are also looking on how syndromic surveillance can be improved by looking at different data feeds including animal data and colleges' health centers. I am in charge of coordinating the efforts of the health group within PURVAC.

Virginia Tech

Efficient Delivery of Web Services (PhD Dissertation)

August 1999 - May 2004

The goal is to design and build a comprehensive query infrastructure for the efficient delivery of Web services. Treating Web services as first class objects is a fundamental step towards achieving

the envisioned Semantic Web. Semantic-aware processing of information requires extensive use of Web services. In this research, I investigated a new query scheme where queries are resolved by combining Web service invocations. To efficiently deploy such a scheme, I proposed an optimization strategy based on aggregating Quality of Web service (QoWS) of different Web services. QoWS is adjusted through a dynamic rating scheme and multilevel matching. Web service rating provides an assessment of their behavior. The multilevel matching enables similar and partial answers.

A Service Platform for Digital Government

The aim of the WebDG project for Digital Government is to enable citizens get timely information from local, state, and federal governments. The disadvantaged citizens would have most of their needs satisfied in one single meeting. We have built a platform to organize, access, integrate, and manage both government databases and applications. Web services have been shown to be an important enabler of digital government. I was in charge of defining the architecture of the system. I also implemented the service query component of WebDG and supervised several Master students who built different pieces of this system.

Uniform Access to Web Databases

I was responsible for extending and maintaining WebFINDIT at Virginia Tech. I migrated the whole WebFINDIT system to a network of several Sun Solaris Workstations. This includes creating 14 databases using five different DBMSs and setting up different CORBA ORBs. I co-supervised different extensions including XML based querying, WebFINDIT access through RMI and DCOM, and XML-based metadata layer.

Queensland University of Technology, Brisbane, Australia **Jan. 1998 - Aug. 1999** *Uniform Access to Web Databases*

I participated in the design and implementation of the WebFINDIT system for the interoperation of autonomous and heterogeneous Web databases. More specifically, we looked at the architectural, language, and query processing issues. The system middleware is based on off-the-shelf products from industry standards, e.g., CORBA, ActiveX/DCOM, EJB, RMI, etc. My work centers on the definition, design, and implementation of query optimization techniques for Web databases.

USTHB, Algiers, Algeria **December 1992 - April 1995** *Concurrency Control for Multidatabase Systems (Thesis Master)*

The topic of my thesis was on concurrency control for multidatabase systems. I proposed novel and efficient concurrency control techniques in such systems based on Two-Level Serializability. This has resulted in a publication in the Very Large Data Base conference (VLDB'94).

LECTURES AND PRESENTATIONS

- *Commigrate Project @ Purdue*, Invited Talk, Yahoo Seminar, UIUC, IL, Fall 2010
- *Community-driven Data Integration*, Invited Seminar, University of Sharjah, Sharjah, UAE, Fall 2008
- *Working with Domain Scientists and Engineers*, Invited Seminar, University of Sharjah, Sharjah, UAE, Fall 2008
- *Community-driven Data Integration*, Invited Seminar, United Arab Emirates University, Al Ain, UAE, Fall 2008
- *The Service Web*, Invited Lecture, CS490W Web Information System, Purdue University, Fall 2007.
- *Duplicate Elimination in Space-partitioning Tree Indexes*, Presentation, 19th International Conference on Scientific and Statistical Database Management, July 2007, Banff, Canada.
- *Semantic Web Services – From Research to Practice*, Presentation, Wright State University, Spring 2007.
- *Semantic Web Services – From Research to Practice*, Presentation, Qatar University, Spring 2007.
- *Efficient Delivery of Web Services*, Presentation, West Michigan University, Spring 2005.
- *The Cyber Center*, several presentations around campus to promote the activities of the Cyber

Center, 2005-2007.

- *A Query Paradigm for Web Services*, Presentation, the 2003 International Conference on Web Services (ICWS'03), June 2003, Nevada.
- *Miscellaneous presentations, seminars, and tutorials*, Virginia Tech and Purdue University.

TEACHING AND
MENTORING
EXPERIENCE

Purdue University

PostDoc

- Brandeis Marshall – Image Retrieval, Aggregation, and Ranking 2007 - 2008
- Eduard Dragut – Data Integration on the Web July 2010 - Pres.

Graduate Students

- Ahmed M. Aly - PhD Student
- Nilothpal Talukder - PhD Student
- Rajasekar Karthik - Master Student

Undergraduate Research Projects

2005 - 2009

- Enabling single sign-on in the ionomics hub project
- Goji – Cyber-enabled Communities for Data Intensive eScience
- Biosamples Database Support and Tracking
- Implementing content-based authorization for bdbms, a biological database management system
- Creating a test bed environment for Web services
- Enabling Access to remote Ecoli Databases
- Using biomoby for accessing Web services

Virginia Tech

M.Sc. Independent Study Projects (Co-Advisor)

2002 - 2003

- A Uniform Interface for Querying Government Databases and Invoking Services.
- Implementation of Distributed Ontologies.
- Deploying Government E-Services using HP E-Speak and DCOM.
- Deploying Government E-Services using HP E-speak and EJB.
- Extending WebFINDIT with XML Repositories.
- Implementing Web Services for Government Applications.

Queensland University of Technology

Graduate Teaching Assistant

July 1999 - August 1999

Undergraduate course – Database Management Systems.

Associate Lecturer

March 1999 - June 1999

Advance graduate course – Current Advances in Database Technology.

Graduate Teaching Assistant

March 1999 - June 1999

Undergraduate course – Object Oriented Analysis and Design Methodology.

Graduate Teaching Assistant

July 1998 - November 1998

Undergraduate course – Database Management Systems.

Ecole National d'Informatique, Algiers

Graduate Teaching Assistant

September 1992 - June 1993

Undergraduate course – Language Theory.

ACADEMIC
ACTIVITIES AND
SERVICES

Research Track Area Coordinator

- The 6th International Conference on Service Oriented Computing (ICSOC'08) Sydney, Australia, December 1-5, 2008.

Session Chairing

- The 26th Int. Conf. on Data Engineering (ICDE), Long Beach, CA, April 2010.
- The 19th International Conference on Scientific and Statistical Database Management (SSDBM 2007), Banff, Canada, July 2007.

Program Committee

- The 27th Int. Conf. on Data Engineering (ICDE), Hanover, Germany, April 2011.
- The ACM SIGMOD International Conference on Management of Data, SIGMOD 2010, Indianapolis, IN, USA, June 2010.
- The 21st Australasian Database Conference (ADC 2010), Brisbane, Australia, January 2010.
- 2009 IEEE Asia-Pacific Services Computing Conference, Singapore, December 2009.
- 14th International Conference on Information Quality (ICIQ), Potsdam, Germany, November 2009.
- 2009 Middleware for Web Services (MWS 2009) Workshop, held at the EDOC 2009 conference in Auckland, New Zealand, Aug-Sep 2009.
- The IEEE 23rd International Conference on Advanced Information Networking and Applications (AINA-09), Bradford, UK, May 2009.
- The 7th ACS/IEEE International Conference on Computer Systems and Applications, May 2009.
- The First Workshop on Web Services in information systems (WWS'09), Algiers, Algeria, February 2009.
- The 20th edition of the Australasian Database Conference (ADC'09), Wellington, New Zealand, Jan 2009.
- The Third IEEE Asia-Pacific Services Computing Conference (APSCC 2008), December 2008.
- The 6th International Conference on Service Oriented Computing (ICSOC'08), December 2008.
- International Workshop on Semantic Extensions to Middleware: Enabling Large Scale Knowledge Applications (SEMELS '08), November 2008.
- Middleware for Web Services (MWS 2008) Workshop in conjunction with EDOC'08, September 2008.
- International Workshop on Context enabled Source and Service Selection, Integration and Adaptation, April 2008.
- The 8th International Conference on Web Information Systems Engineering (WISE'07), December 2007.
- CollaborateCom2007 / The 3rd International Conference on Collaborative Computing: Networking, Applications and Worksharing, November 2007.
- The third International Conference on Signal-Image Technology & Internet-based Systems (SITIS'2007), December 2007.
- Middleware for Web Services (MWS 2007) Workshop in conjunction with EDOC'07, October 2007.
- The 27th International Conference on Distributed Computing Systems (ICDCS'07), June 2007.
- The International Conference on Mobile Data Management (MDM'07), May 2007.
- The Eighth International Symposium on Programming and Systems (ISPS'2007), May 2007.
- 14th International Conference on Cooperative Information Systems (CoopIS 2006), November 2006
- Middleware for Web Services (MWS 2006) Workshop in conjunction with EDOC'06, October 2006.
- Mobile and Wireless Learning Workshop, in Conjunction with IWCMC 2006, July 2006.
- The 6th International Conference on Web Information Systems Engineering (WISE'05), 2005.
- Middleware for Web Services (MWS 2005) Workshop in conjunction with EDOC'05, 2005.
- The 6th International Conference on Mobile Data Management (MDM'05), 2005.
- The Seventh International Symposium on Programming and Systems, 2005.
- The IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE-05), 2005.
- The Second International Workshop on Ubiquitous Computing, 2005.

Journal Refereeing

- The VLDB Journal
- IEEE Transactions on Services Computing
- Distributed and Parallel Databases
- Journal of Data and Information Quality
- IEEE Trans. on Systems, Man, and Cybernetics
- IEEE Internet Computing
- IEEE Computer
- Data & Knowledge Engineering
- The Computer Journal
- International Journal of Cooperative Information Systems
- International Journal of Web Services Research
- Software Practice and Experience
- Information Systems
- The Ubiquitous Computing and Communication Journal
- Journal of Software Maintenance

Technical Societies Memberships

- IEEE, IEEE Computer Society, and ACM since 1998.
- IEEE Technical Committee on Data Engineering and ACM SIGMOD since 2001.

Miscellaneous Items

- Attended an equal access/equal opportunity briefing that provides participants with information on discrimination, harassment, sexual harassment, disabilities, retaliation, and Purdue's process for resolving complaints.
- Fall 2001 - Spring 2002 – Coordinator of the Research Seminar Series in the CS Department.
- Attended workshops on grant writing and career management (VA Tech, Purdue University, and Computing Research Association).
- Attended an intensive course for Informix DBAs organized and funded by Informix, June 2000.
- Attended a workshop on HP E-speak e-service platform, one of the first Web service platform, organized and funded by HP, July 2000.

PROFESSIONAL EXPERIENCE

System Administrator (Virginia Tech), **January, 2000 - December, 2002**

- Installed a network of several Sun/Solaris Workstations using NIS+.
- Installed major databases and other off-the-shelf application in a client-server type mode.

Independent IT Consultant and Trainer **January, 1991 - December, 1997**

Information Technology Consultant and Trainer for several companies. I was involved in the development of several software packages related to inventory management (pharmacy, bookstores) and educational games.