

Matthew S. Perry

<http://knoesis.wright.edu/students/mperry>
msperry@gmail.com

RESEARCH INTERESTS

I have broad research interests in database systems, geographic information systems, and the Semantic Web. My specific interests include: management of spatial and temporal data on the Semantic Web, design and implementation of analytical query operators for RDF graphs, and ontology-based querying of relational data.

EDUCATION

Wright State University – Dayton, OH August 2008

- Ph.D. in Computer Science and Engineering
Dissertation: *A Framework for Spatial, Temporal and Thematic Analytics over Semantic Web Data*
Advisor: Amit P. Sheth

The University of Georgia – Athens, GA August 2002

- B.S. in Computer Science
Graduated with Summa Cum Laude honors

INDUSTRY EXPERIENCE

Oracle USA, Inc. – Nashua, NH August 2008 – Present
Senior Member of Technical Staff – Semantic Technologies Group

Oracle USA, Inc. – Nashua, NH May 2006 – August 2006
Research Intern – Semantic Technologies Group

- Researched operators for ontology-based querying of keyword columns in relational tables
- Contributed code to core Oracle database product

Semagix, Inc. – Athens, GA May 2004 – August 2004
Intern Java Developer May 2003 – August 2003

- Refactored modules, created new modules and corrected bugs in Semagix Freedom content management software
- Tuned JDBC code to increase performance of database operations

ACADEMIC EXPERIENCE

Wright State University – Dayton, OH January 2007 - August 2008
Graduate Research Assistant – Kno.e.sis Center

- Researched new query operators involving spatial, temporal and thematic relationships in RDF graphs

- Researched techniques for RDF query processing

The University of Georgia – Athens, GA August 2006 – December 2006
Graduate Research Assistant – LSDIS Lab August 2004 – May 2006

- Researched algorithms for evaluation of analytical queries over RDF graphs including semantic association and subgraph extraction queries
- Developed tools to create synthetic data sets for testing Semantic Web applications

The University of Georgia – Athens, GA August 2003 – May 2004
Graduate Teaching Assistant – Computer Science Department

- Clarified lecture concepts, graded programming assignments for introductory and intermediate level programming courses
- Explained and graded lab assignments for an introductory programming course

The University of Georgia – Athens, GA August 2002 – May 2003
Graduate Research Assistant – Computer Science Department

- Researched data mining techniques using Genetic Algorithms for discovering classification rules

SKILLS

Programming Languages: Java, C, C++
Web: HTML, Java Servlets, JSP, XML
Semantic Web: RDF, OWL, SPARQL, Oracle Semantic Data Store
Database: SQL, PL/SQL, JDBC
Operating Systems: Windows, UNIX, LINUX
GIS: ESRI ArcGIS, TransCAD

PUBLICATIONS

CONFERENCE PUBLICATIONS

- C1** Matthew Perry, Amit Sheth, Farshad Hakimpour, Prateek Jain. Supporting Complex Thematic, Spatial and Temporal Queries over Semantic Web Data, Second International Conference on Geospatial Semantics (GEOS '07), Mexico City, MX, November 29 – 30, 2007
- C2** Matthew Perry, Farshad Hakimpour, Amit Sheth. Analyzing Theme, Space and Time: An Ontology-based Approach, Fourteenth International Symposium on Advances in Geographic Information Systems (ACM-GIS '06), Arlington, VA, November 10 – 11, 2006

JOURNAL PUBLICATIONS

- J1** Amit Sheth and Matthew Perry. Traveling the Semantic Web through Space, Time and Theme, IEEE Internet Computing, Volume 12, Issue 2, February/March 2008, pp. 81-86
- J2** Cartic Ramakrishnan, William Milnor, Matthew Perry, Amit Sheth. Discovering Informative Connection Subgraphs in Multi-relational Graphs, SIGKDD Explorations Special Issue on Link Mining, Volume 7, Issue 2, December 2005, pp. 56-63

WORKSHOP PUBLICATIONS

- W1** Farshad Hakimpour, Boanerges Aleman-Meza, Matthew Perry, Amit Sheth. Data Processing in Space, Time, and Semantics Dimensions, Terra Cognita 2006 - Directions to the Geospatial Semantic Web, in conjunction with the Fifth International Semantic Web Conference (ISWC '06), Athens, GA, November 6, 2006
- W2** Matthew Perry, Maciej Janik, Cartic Ramakrishnan, Conrad Ibanez, Ismailcem Budak Arpinar, Amit Sheth. Peer-to-Peer Discovery of Semantic Associations, Second International Workshop on Peer-to-Peer Knowledge Management (P2PKM '05), San Diego, CA, July 17, 2005

BOOK CHAPTERS

- B1** Matthew Perry, Amit Sheth, Ismailcem Budak Arpinar. Geospatial and Temporal Semantic Analytics, To appear in Encyclopedia of Geoinformatics, Hassan A. Karimi (Ed), Idea-Group Inc., 2008
- B2** Farshad Hakimpour, Boanerges Aleman-Meza, Matthew Perry, Amit Sheth. Spatiotemporal-Thematic Data Processing in Semantic Web, The Geospatial Web, Springer-Verlag, May, 2007

POSTERS & SHORT PAPERS

- P1** Matthew Perry, Farshad Hakimpour, Amit Sheth. Analyzing Theme, Space and Time: An Ontology-based Approach, Fifth International Semantic Web Conference (ISWC '06), Poster Paper, Athens, GA, November 5 – 9, 2006
- P2** Matthew Perry. TOntoGen: A Synthetic Data Set Generator for Semantic Web Applications, AIS SIGSEMIS Bulletin, Volume 2, Issue 2, (April – June) 2005, pp. 46-48
- P3** Matthew Perry and Eric Stiles. SEMPL: A Semantic Portal, Thirteenth International World Wide Web Conference (WWW '04), Poster Paper, New York, NY, May 17-22, 2004

RESEARCH & DEVELOPMENT SYNOPSIS

RESEARCH CONTRIBUTIONS

The majority of my research, including my dissertation research, has been done in the context of an NSF funded project titled SemDis: Discovering Complex Relationships in the Semantic Web.

Graph Traversal Algorithms for RDF Data

The SemDis project investigated a new querying paradigm that focused on querying about relationships between resources in RDF graphs. Semantic association paths and connection subgraphs were defined as a means to qualify the relatedness of resources. I worked with colleagues to develop algorithms for semantic association discovery over distributed RDF datasets in a peer-to-peer setting [W2]. We also developed algorithms for connection subgraph discovery that centered around the idea of modeling an RDF graph as an electrical circuit [J2].

Semantic Discovery over Spatial and Temporal Data

My dissertation research focuses on a framework for utilizing spatial and temporal data within the SemDis project. I proposed a modeling approach for spatial, temporal and thematic data that used an upper-level ontology to incorporate spatial objects into RDF and utilized temporal RDF graphs for temporal information [C1].

I formalized a querying approach using a notion of context and presented an efficient implementation involving an extension of Oracle DBMS [C2]. The final aspect of my dissertation is an extension of the SPARQL query language to enable this querying approach.

Semantics-based Querying of Relational Data

During my internship at Oracle, I worked with the semantic technologies group on extensions of the Oracle Database to allow querying traditional relational data using ontologies. Our extension allowed users to assert that a keyword column references terms from an ontology (e.g., a diagnosis column in a patient records table that references terms from NCI Cancer Ontology). We implemented relational operators that allowed querying such relational tables based on semantic relatedness and semantic similarity computed over the referenced ontology.

DEVELOPMENT EXPERIENCE

I have significant hands-on experience working with large software projects in both industrial and academic settings.

Industry Development

During my internships, I was able to contribute production-level code that became a part of each company's commercial product. At Semagix, I made modifications to an existing Java codebase to correct bugs and refactor modules for better maintainability. At Oracle, I worked with the group that brought support for Semantic Web data to Oracle database. I used Oracle's extensibility framework to implement a new relational operator called *sem_related*. The operator was implemented using PL/SQL and C and shipped as a part of Oracle 11g.

Academic Development

Two representative software projects completed during my academic research include a tool for generating synthetic RDF datasets to test various Semantic Web applications and a series of database extensions completed during my dissertation research. The dataset generator, called TOnToGen [P2], was released as a plugin for the Protege ontology editor and was used by other researchers and myself in evaluations of work presented at ISWC, ESWC, WWW and GEOS conferences and in SIGKDD Explorations journal. I extended Oracle database with a variety of user-defined functions that combined aspects of Oracle Spatial and Oracle Semantic Data Store. The implementation showed good scalability during extensive testing with synthetic and real-world RDF datasets of over 25 million triples.

PROFESSIONAL SERVICE

Program Committee Member

- International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE), 2008
- International Workshop on Semantic Metadata Management and Applications (SeMMA), 2008
- International Conference on Geospatial Semantics (GEOS), 2007
- Terra Cognita – Directions to the Geospatial Semantic Web, International Workshop, 2007

External Reviewer

- Knowledge and Information Systems, International Journal (KAIS), 2008

- International Conference on Applications of Natural Language to Information Systems (NLDB), 2008
- International Conference on Web Reasoning and Rule Systems (RR), 2007
- International Workshop on Cooperative Information Agents (CIA), 2007
- Indian International Conference on Artificial Intelligence (IICAI), 2007
- GeoWeb, International Conference, 2007
- Asian Semantic Web Conference (ASWC), 2006
- International Conference on Geospatial Semantics (GEOS), 2005
- International Journal on Semantic Web and Information Systems (IJSWIS), 2005, 2006, 2007
- International Conference on Data Engineering (ICDE), 2005
- International Conference on Intelligence and Security Informatics (ISI), 2005
- International Semantic Web Conference (ISWC), 2005
- International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE), 2005
- International Conference on Web Information Systems Engineering (WISE), 2005

SIGNIFICANT ASSISTANCE IN GRANT PROPOSALS (*with intellectual contributions*)

Spatio-Temporal-Thematic Queries of Semantic Web Data: a Study of Expressivity and Efficiency

Investigators: Amit Sheth (PI), T. K. Prasad

Status: Funded

SemGrid: Semantic Discovery on Adaptive Services Grid

Investigators: Amit Sheth (PI), I. Budak Arpinar (UGA), Krys Kochut (UGA),
John A. Miller (UGA)

Status: Funded

GSAV: Geospatial Semantic Analytics and Visualization

Investigators: Amit Sheth (PI), Ramesh Jain (UC Irvine), I. Budak Arpinar (UGA),
Leon Deligiannidis (UGA), E. Lynn Uery (UGA),
Philip Burns (Computer Technology Associates)

Status: Not Funded