

Maciej Janik

107 Barrington Dr.
Apt #3.
Athens, GA 30605

Phone: (706) 254-4123
e-mail: mjanik@uga.edu
URL: <http://lsdis.cs.uga.edu/~mjanik>

Education

PhD in Computer Science

01/2004 – Present [University of Georgia, LSDIS Lab](#)
Athens, Georgia
advisor: [Prof. Krzysztof Kochut](#)
focus: *Text classification using ontology without the training set; Large scale Semantic Web storage and querying systems.*
Planned completion: Summer 2008.

M.S. in Computer Science

09/1994 - 07/1999 [AGH - University of Science and Technology](#)
Krakow, Poland
advisor: [Prof. Krzysztof Zielinski](#)
focus: *Fault tolerance middleware systems for CORBA applications.*

M.S. in Marketing and Business Management

specialization: Human Resource Management
09/1997 - 07/2001 [AGH - University of Science and Technology](#)
Krakow, Poland
advisor: Prof. Boguslaw Wiernek

Experience

Graduate Research Assistant

01/2004 - Present [LSDIS Lab](#), UGA
Athens, Georgia

- Research on **text classification using ontologies** – classifying text with the exclusive use of ontology (no training set used) and classification to a semantic context defined in ontology.
- Member of **SemDis**, NSF-ITR funded project "[Semantic Discovery: Discovering Complex Relationships in the Semantic Web](#)". Research focus on defining and representing meaningful and interesting relationships (called: semantic associations) between entities in RDF graphs. It includes creation of algorithms for discovering informative paths, indexing and querying of complex semantic relationships, as well as ranking association results in terms of relevance to the context of the query.
 - [Brahms](#) - research on high-performance RDF/S storage and querying system for large ontologies implemented in C++. The system has been created as OpenSource and is being already used outside of UGA.
 - **SPARQLer** - extension of [SPARQL](#) query language for RDF to support flexible-length path queries with specific semantics defined with the use of Regular Expressions on the properties included in the result path.
- **GlycoVault** (part of [Glycomics](#), an NHI funded project "Integrated Technology Resource for Biomedical Glycomics: Technological Research and Development Project IV") - storage and semantic search engine for heterogeneous data produced in Glycomics experiments performed on genes and glycans at [CCRC](#).

Summer Intern – programmer / research

5/2006 - 8/2006

Amazon.com

development center in Seattle, WA

- [Amapedia](#) - tests on category assignments based on description text.
- Experiments on product categorization from one sub-tree of Amazon's catalog and extraction of attributes (like price, length, color, etc) defined specifically for this class.
- [AskVille](#) - initial tests on categorizing questions asked to the system to suggest appropriate topics.

Programmer / Analyst

11/2000 - 12/2003

[Sabre Holdings](#)

development center in Krakow, Poland

- [Webservices for B2B](#) - set of webservices for querying, browsing and booking of airline tickets, hotel rooms and car rentals.
- Integration of [TAM airlines](#) (Brazil) booking engine with Sabre framework.
- Airfare management for large travel agencies to manage negotiated airfares used by booking applications (support: STADAF, Consolidator, ATPCO)
- Hotel module of Sabre dotRes platform - responsible for whole module (display, search and booking of hotel rooms)

Programmer / Researcher

09/1998 - 07/1999

[DSRG](#) (Distributed Software Research Group)

AGH - University of Science and Technology

Krakow, Poland

Research on security and fault-tolerance in object distributed systems (CORBA, C++), extended to Master thesis supervised by Prof. Krzysztof Zielinski. Design and implementation of fault-tolerance framework for systems that use CORBA.

System Administrator / Web Developer

08/1996 - 05/1997

[Interkom](#)

Krakow, Poland

Part-time position of system administrator and web developer with a Krakow-based internet service provider.

Publications

Maciej Janik, Krys Kochut. "[Training-less Ontology-based Text Categorization](#)", Workshop on Exploiting Semantic Annotations in Information Retrieval ([ESAIR 2008](#)) at the 30th European Conference on Information Retrieval ([ECIR'08](#)), Glasgow, Scotland, 30 March 2008

Matthew Eavenson, **Maciej Janik**, Shravya Nimmagadda, John A. Miller, Krys J. Kochut, William S. York. "[GlycoBrowser - A Tool for Contextual Visualization of Biological Data and Pathways Using Ontologies](#)", 4-th International Symposium on Bioinformatics Research and Applications ([ISBRA2008](#)), Atlanta, Georgia (May 2008) [to appear]

Shravya Nimmagadda, Amrita Basu, Matthew Eavenson, Jun Han, **Maciej Janik**, Rajesh Narra, Kishore Nimmagadda, Arpan Sharma, Krys J. Kochut, John A. Miller and William S. York, "[GlycoVault: A Bioinformatics Infrastructure for Glycan Pathway Visualization, Analysis and Modeling.](#)" Proceedings of the 5th International Conference on Information Technology: New Generations ([ITNG'08](#)), Las Vegas, Nevada (April 2008) [to appear]

Krys Kochut, **Maciej Janik**. "[SPARQLeR: Extended Sparql for Semantic Association Discovery](#)", [Fourth European Semantic Web Conference](#), Innsbruck, Austria, 3-7 June 2007

Maciej Janik, Krys Kochut. "[BRAHMS: A WorkBench RDF Store And High Performance Memory System for Semantic Association Discovery](#)", [Fourth International Semantic Web Conference](#), Galway, Ireland, 6-10 November 2005

Matthew Perry, **Maciej Janik**, Cartic Ramakrishnan, Conrad Ibanez, Budak Arpinar, Amit Sheth. "[Peer-to-Peer Discovery of Semantic Associations](#)", [Second International Workshop on Peer-to-Peer Knowledge Management](#), San Diego, CA, July 17, 2005

Significant creative contributions

Brahms has been designed as a fast main-memory RDF/S storage, capable of storing, accessing and querying large ontologies. It does not use any relational database backend and all data is kept in main memory. BRAHMS is implemented in C++ for high performance and strict memory control. It has been successfully used in several projects with the LSDIS Lab, as well as outside of UGA (eg. [Knoesis center](#) at [Wright State University](#)). BRAHMS is available for download as OpenSource software at <http://lsdis.cs.uga.edu/projects/semdis/brahms/>.

SPARQLeR has been created as extension and fast implementation of SPARQL query language for RDF. In addition to original language, it supports queries for paths of flexible-length, required in SemDis project. The language extension enables the user to define strict semantics of searched paths with the use of Regular Expressions on relationships included in the result path. Additional constraints can be applied to structural properties or entities. SPARQLeR is intended to be soon available for download as OpenSource software.

Presentations

[Sparqler, Extended Sparql for Semantic Association Discovery](#). Public presentation at 4th European Semantic Web Conference, ESWC 2007, in Innsbruck, Austria.

[BRAHMS – fast main-memory RDF/S storage](#). Public presentation at 4th International Semantic Web Conference, ISWC 2005, in Galway, Ireland.

Technical skills

Programming	C / C++, Java, Visual Basic
Semantic Web	<u>RDF/RDFS</u> , <u>OWL</u> , <u>SPARQL</u> , <u>RDQL</u> , <u>Jena</u> , <u>Sesame</u> , <u>Redland</u> , <u>Brahms</u>
Web technologies	JSP, JavaScript, Servlets, HTML
Databases	SQL, JDBC
Operating systems	Windows, Linux, Unix
Other	UML

Research interests

- Semantics and Semantic Web
- Automatic document classification with the use of Ontology
- RDF/S storages and querying systems for large ontologies
- Graph algorithms
- Machine learning
- Semantics of relationships and its use in search
- Discovery of meaningful semantic association paths and subgraphs