

Maria Grineva

ETH Zurich

Institut f. Informationssysteme CAB E 73.2, Universitätstrasse 6, 8092 Zürich

+41 44 633 91 26 maria.grineva@inf.ethz.ch

My research interests lie in the fields of database system internals, data integration, XML technologies, content management, information retrieval and text mining. I have 8+ years of experience in software engineering, bringing my research results into industrial strength systems.

EDUCATION

Ph.D. in Computer Science, Institute for System Programming of Russian Academy of Sciences, Russia, 2007

Thesis: Methods to Support Active Behavior in XML Database Management Systems

Advisor: Sergey Kuznetsov, a leading Russian specialist in the area of Database and Operating Systems

M.S. in Computer Science, Moscow State University, Russia, 2003

Thesis: Comparative Analysis of XML-schema Specification Languages and XML-document Validation Methods

EXPERIENCE

February 2010 – present

Senior Researcher at Systems Group, Swiss Federal Institute of Technology Zurich (ETHZ). Working with Prof. Donald Kossmann.

Selected projects in this position:

Filtering Noise in Topical Microblogging Streams

Real-time Product Recommendations Based On User's Updates in Social Networks

December 2001 – February 2010

Researcher at the Institute for System Programming of Russian Academy of Sciences

Texterra – Semantic Text Mining and Search (Feb. 2008 – Feb. 2010)

Texterra (<http://modis.ispras.ru/texterra>) takes a novel approach to semantic text mining and search by utilizing semantics extracted from user-generated knowledge bases (such as Wikipedia, etc). It mines Wikipedia to build a semantic network with more than 6 millions nodes (for English Wikipedia). The network is then used to interpret meanings and relationships of terms in arbitrary text documents. We have developed new methods that utilize knowledge extracted from Wikipedia to solve classical NLP tasks: keywords extraction, text summarization, topic inference, text categorization and query expansion.

There are two products based on the Texterra technology: (1) ad targeting system for Web pages; (2) search and navigation engine for Blogosphere (<http://blognoon.com>)

Responsibilities in this project:

- Lead scientist
- Software architect
- Design and implementation of Wikipedia Knowledge Base subsystem, Semantic Text Processing pipeline and database backend

TweetSieve – news search engine for Twitter (Feb. - Apr. 2008)

TweetSieve - a system that allows obtaining news on any given subject by sifting the Twitter stream. Our work is related to frequency-based analysis that has been applied to blogs, but higher latency and lower coverage in blogs makes the analysis less effective than in case of micro-blogs (Twitter)

Sedna – Native XML Database System (2003 – Feb. 2010)

Sedna is a leading open-source native XML database. It provides a full range of core database services - persistent storage, ACID transactions, security, indices, hot backup. Support for W3C XQuery implementation, tight integration of XQuery with full-text search facilities and a node-level update language. More information:

<http://modis.ispras.ru/sedna/>

Responsibilities in this project:

- Design and implementation of:
 - System architecture and inter-process interaction
 - Trigger subsystem
 - Database Users&Privileges subsystem
 - Client-Server Protocol, low level API

Content Management Solution for “Great Russian Encyclopedia” Publishing Company www.greatbook.ru (2006-2007)

Full lifecycle content and knowledge management solution for production of encyclopedias and references. The solution provides a single-sourcing publishing with content reuse, various editorial services (version management, content validation), publishing on various media (paper, Web, DVD), superior search & navigation. It is an XML-centric solution based on Sedna XML database.

Responsibilities in this project:

- Design and implementation of editorial services

BizQuery – XML-based Virtual Data Integration System (Dec. 2001 – Sept. 2003)

BizQuery is a data integration system for heterogeneous data with support for XML and relational databases. It provides a uniform query interface over a mediated schema implementing "Global As View" (GAV) approach. Global schema is defined in XML and queried in XQuery. BizQuery supports virtual data integration: a query to the global schema is executed on-the-fly by transforming it into specialized queries over the original databases that guaranties up-to-data information.

Responsibilities in this project:

- Design and implementation of DTD Schema mapper tool

September 2008 – present

Co-founder of Semantic Dimension, Inc. (venture-funded startup)

Main Product:

The Tweeted Times – a real-time personalized newspaper generated from your Twitter account

<http://www.tweetedtimes.com>

RESEARCH GRANTS AND AWARDS

I participated in several research projects supported by various grants and organizations.

March 2002 – January 2010

10 projects in the area of XML data management and information retrieval supported by grants from RFBR (Russian Foundation for Basic Research).

April 2007 – January 2010

Research Project “Natural Language Processing (NLP) Tools for Taming Information Explosion”. Supported by HP Labs Innovation Research Program.

In **2006** I was awarded with the grant “The best Ph.D. students of Russian Academy of Sciences” by Russian Science Support Foundation.

PROFESSIONAL SERVICES

- Program Committee member at the following conferences:
SYRCoDIS 2008, SYRCoDIS 2007, SYRCoDIS 2006
- Program Committee Co-Chair : Web As A Stream – FQAS 2009
- Reviewer: World Wide Web Journal, ADBIS 2006

TEACHING AND STUDENT ADVISING

Spring 2011

Teaching course “The Structure of Social And Information Networks” at ETH Zurich

Spring 2010

Assisting course “Introduction to Natural Language Processing” given by Keith Hall (Google), at ETH Zurich

2003 – 2009

Co-leading research student seminar on Data Management, scientific adviser of 4 M.S. students at Moscow State University.

SELECTED PUBLICATIONS

- **Blognoon: Exploring A Topic In The Blogosphere**
M. Grineva, M. Grinev, D. Lizorkin, A. Boldakov
WWW2011: 20th International World Wide Web Conference, Hyderabad, India
- **Sedna: Native XML Database Management System (Internals Overview)**

D. Lizorkin, M. Grineva, M. Grinev, I. Taranov, A. Kalinin, I. Shcheklein, L. Novak, A. Fomichev

To appear in proceedings of ACM SIGMOD/PODS 2010, Indianapolis, Indiana, USA

- **Sifting Micro-blogging Stream for Events of User Interest (Demo)**

M. Grineva, M. Grinev, Alexander Boldakov, Leonid Novak, Andrey Syssoev, D. Lizorkin

ACM SIGIR 2009, Boston, Massachusetts, USA

- **Extracting Key Terms From Noisy and Multitheme Documents**

M. Grineva, M. Grinev, D. Lizorkin

WWW2009: 18th International World Wide Web Conference, Madrid, Spain

- **Analysis of Community Structure in Wikipedia (Poster)**

M. Grineva, O. Medelyan, D. Lizorkin

WWW2009: 18th International World Wide Web Conference, Madrid, Spain

- **Effective Extraction of Thematically Grouped Key Terms From Text**

M. Grineva, M. Grinev, D. Lizorkin

AAAI-SSS-09: Social Semantic Web: Where Web 2.0 Meets Web 3.0. Stanford University, USA

- **Harnessing Wikipedia for Smart Tags Clustering**

M. Grineva, M. Grinev, D. Turdakov, P. Velikhov, A. Boldakov

KASW 2008: International Workshop on "Knowledge Acquisition from the Social Web", Graz, Austria

- **Query Triggers for XML DBMS: Efficient implementation based on shadow mechanism**

Maria Grineva and Maxim Grinev

Programming and Computer Software Journal, Vol. 33, No. 4, pp. 204-213, 2007