

Rares Vernica

Department of Computer Science, University of California, Irvine, USA

E-mail: rares@ics.uci.edu Homepage: <http://www.ics.uci.edu/~rares/>

Education

- 2005 - present University of California, Irvine, USA
PhD Candidate in Computer Science, PhD expected in Spring 2011
Advisers: Prof. Michael J. Carey and Prof. Chen Li
- 2004 Politehnica University of Bucharest, Romania
BSc in Computer Science and Engineering

Work Experience and Projects

- 06/2005 - present University of California, Irvine, USA
Graduate Student Researcher, Department of Computer Science
Research in the area of large-scale data management, in particular, efficient indexing and query processing when the data and the query conditions do not match exactly.
Projects:
- ASTERIX: developing a highly scalable parallel platform for managing semi-structured data (<http://asterix.ics.uci.edu/>)
 - Developed parallel techniques for set-similarity joins (fuzzy joins) on large-scale data.
 - Implemented fuzzy-join techniques in ASTERIX and Hadoop.
 - Conducted performance studies and implemented optimizations for improving performance, speedup, and scaleup.
 - Released source code of the techniques.
 - Family Reunification: developing data integration, indexing, and search techniques to help finding people during and after a disaster (<http://fr.ics.uci.edu/>)
 - Developed the core architecture and techniques for approximate search for large-scale data.
 - Conducted performance studies and implemented optimizations for reducing response time.
 - Deployed publicly-accessible Web services for helping people find loved ones during Hurricane Katrina, and Haiti and Chile Earthquakes.
 - Flamingo: developing data cleaning techniques to deal with errors and inconsistencies in information systems (<http://flamingo.ics.uci.edu/>)
 - Developed selectivity estimation and ranking algorithms for approximate search for large-scale data.
 - Conducted performance studies and implemented optimizations for reducing response time.
 - Released open-source library and toolkit for approximate string matching.
 - Deployed a publicly-accessible Web directory search service for UC Irvine.

- 06/2010 - IBM Almaden Research Center, San Jose, CA, USA
- 12/2010 Research Intern, Information Management Group, Manager: Eugene Shekita
 - Developed adaptive techniques for large-scale data processing.
 - Implemented an adaptive framework in Hadoop and ZooKeeper.
 - Conducted performance studies and implemented optimizations for reducing response time.
 - Released code to be integrated in an IBM's analytics platform.
- 06/2007 - Microsoft Research, Redmond, CA, USA
- 09/2007 Research Intern, Data Management, Exploration and Mining Group, Manager: Surajit Chaudhuri
 - Developed entity categorization techniques over large document collections.
 - Conducted performance studies and implemented optimizations for improving result quality and reducing response time.
- 06/2006 - Yahoo!, Sunnyvale, CA, USA
- 09/2006 Technical Intern, Strategic Data Solutions Department, Manager: Kazi Zaman
 - Developed data-pipeline testing framework for shared-nothing distributed architectures.
 - Implemented a testing framework using the TETware library.
- 02/2005 - Transylvania University of Brasov, Romania
- 06/2005 Lecturer, Department of Applied Informatics
 - Prepared syllabus, class projects, and homework assignments for Internet Technologies class.
- 03/2004 - Learning Lab Lower Saxony (L3S), Hannover, Germany
- 08/2004 Research Assistant, Manager: Wolfgang Nejdl
 - Developed publish/subscribe framework for peer-to-peer distributed systems.
 - Implemented a publish/subscribe framework using the JXTA library.
- 02/2003 - Freelance Developer (self-employed)
- 06/2005 · Developed on-line games, Web-browser application, and modules for on-line scheduling, domain registration, and payment systems in Java, Python, and PHP.
- 07/2002 - Finsiel, Bucharest, Romania
- 02/2003 Web Developer
 - Developed internal web portals for the European Environment Agency in Python and Zope.
- 10/2001 - Websoft, Bucharest, Romania
- 07/2002 Web Developer
 - Developed e-commerce applications in PHP.

Honors and Awards

- 09/2010 Yahoo! Key Scientific Challenges Winner in the Web Information Management Area
- 07/2009 Microsoft Student Travel Award for KEYS 2009, Workshop on Keyword Search on Structured Data, collocated with SIGMOD 2009
- 09/2005 - 06/2006 Chair's Fellowship, Department of Computer Science, UC Irvine

09/2005 California Institute for Telecommunications and Information Technology (Calit²)
Fellowship
02/2005 Second place, Pentalog Programming Contest, Brasov, Romania
10/1999 - 02/2004 Merit-Based Scholarship, Politehnica University of Bucharest, Romania

Publications

1. *CIRCUMFLEX: A Scheduling Optimizer for MapReduce Workloads Involving Shared Scans*. Joel Wolf, Deepak Rajan, Kirsten Hildrum, Rohit Khandekar, Sujay Parekh, Kun-Lung Wu, Andrey Balmin, **Rares Vernica**. LADIS 2011, Workshop on Large Scale Distributed Systems and Middleware, collocated with VLDB 2011.
2. *Adaptive MapReduce using Situation-Aware Mappers*. **Rares Vernica**, Andrey Balmin, Kevin S. Beyer, Vuk Ercegovic. Submitted for publication.
3. *ASTERIX: Towards a Scalable, Semistructured Data Platform for Evolving World Models*. Alexander Behm, Vinayak R. Borkar, Michael J. Carey, Chen Li, Nicola Onose, **Rares Vernica**, Alin Deutsch, Yannis Papakonstantinou, Vassilis J. Tsotras. Journal of Distributed and Parallel Databases 2011.
4. *Hyracks: A Flexible and Extensible Foundation for Data-Intensive Computing*. Vinayak R. Borkar, Michael J. Carey, Raman Grover, Nicola Onose, **Rares Vernica**. ICDE 2011.
5. *AKYRA: Efficient Keyword-Query Cleaning in Relational Databases*. **Rares Vernica**, Chen Li. Technical Report, University of California, Irvine 2010.
6. *Efficient Parallel Set-Similarity Joins Using MapReduce*. **Rares Vernica**, Michael J. Carey, Chen Li. SIGMOD 2010.
7. *Efficient Top-k Algorithms for Fuzzy Search in String Collections*. **Rares Vernica**, Chen Li. KEYS 2009, Workshop on Keyword Search on Structured Data, collocated with SIGMOD 2009.
8. *Entity Categorization Over Large Document Collections*. Venkatesh Ganti, Arnd Christian König, **Rares Vernica**. KDD 2008.
9. *SEPIA: Estimating Selectivities of Approximate String Predicates in Large Databases*. Liang Jin, Chen Li, **Rares Vernica**. VLDB J. 2008.
10. *Relaxing Join and Selection Queries*. Nick Koudas, Chen Li, Anthony K. H. Tung, **Rares Vernica**. VLDB 2006.

Professional Services

2011 External reviewer: VLDB
2010 External reviewer: ICDE, AAI
2009 External reviewer: EDBT, ICDM, IIWeb
2008 External reviewer: VLDB, ICDE, WWW, EDBT, CIKM, ICDL, CoopIS
2007 External reviewer: SIGMOD, KDD
2006 External reviewer: KDD, WAIM, APWeb

Hobbies

Kendo, All United States Kendo Federation, Rank: 2-dan (black belt).