

Ramakrishna R. Varadarajan

Home Address:

1308 Spring Street, Apt #101,
Madison, WI 53715.
Phone: (608) 770 7628.
Email: ramkris@cs.wisc.edu.
Web page: www.cs.wisc.edu/~ramkris

University Address:

Department of Computer Sciences,
1210 West Dayton Street,
University of Wisconsin-Madison,
Madison, WI 53706-1685.
Phone: (608) 890 0015.

RESEARCH INTERESTS

Databases – Ranked search on graph-structured databases including various domains like bibliographic, healthcare/clinical, biological and intellectual property (IP) databases and XML Search evaluation.

Information Retrieval – Web Search, XML Search, Text summarization.

Data Mining – Frequent Graph Pattern Mining and Significant Substructure Mining.

TEACHING INTERESTS

Database Management Systems, Data Structures, Information Retrieval, Software Engineering, Programming, Graph Theory, Natural Language Processing, Design and Analysis of Algorithms.

EDUCATION

Ph.D. in Computer Science, Florida International University (Aug 04 – April 09)

School of Computing & Information Sciences, Miami

Dissertation: Ranked Search on Data Graphs (*Advisor*: Professor Vagelis Hristidis)

Awards: Dissertation Year Fellowship, Presidential Fellowship, Outstanding Graduate Research award & Excellence award.

CGPA: 3.97/4.0

M.S. in Computer Science, Florida International University (Aug 04 – April 06)

School of Computing & Information Sciences, Miami

CGPA: 3.95/4.0

B.E. in Computer Science & Engineering, University of Madras, India (June 2000 – June 2004)

First Class Honors (Ranked 5th – University Gold Medalist)

RESEARCH EXPERIENCE

Post-Doctoral Research Associate, University of Wisconsin-Madison (May 09 – Present)

Database Research Group, Department of Computer Sciences

University of Wisconsin-Madison.

Advisor: Professor Jignesh M. Patel

Job Description: During my employment as a post-doctoral research associate (since May 2009) in the Database Research Group at the Department of Computer Sciences, University of Wisconsin-Madison, I worked on various problems related to frequent pattern mining in large biological & chemical graph-structured databases and developed solutions for effective and

efficient, exact and approximate frequent pattern mining in large graph transaction databases. Several prototype graph mining systems were designed, implemented and tested.

I developed cutting-edge bioinformatics tools and worked closely with biomedical scientists to solve actual problems. I worked closely with a leading nephrology research group at the University of Michigan (where I my current supervisor Prof. Jignesh Patel was a tenured faculty member till 2008). This group observed that classification of patients based on conventional criteria such as histology and laboratory values in comprehensive datasets often show a significant discrepancy with patterns in gene expression. To address this problem, the group performed a molecular classification of a comprehensive gene expression dataset of 226 patients with 11 kidney related disease states and linked the results back to clinical and histological data. In detail, they selected the regulated genes from each patient by comparison to a pool of healthy controls. To control for noise in the data and redundancy in gene function, they extended the gene lists to networks by adding edges representing co-citations of genes in PubMed abstracts. This produced genetic networks of the kidney disease patients. My task was to compare these genetic networks with an approximate graph matching tool (TALE) and subsequently cluster the networks by similarity. Later, I developed a pattern mining tool that analyzed each cluster for patterns either specific to the cluster or shared across clusters, and the group later tested the results for homogeneity - by appearance of patterns in the patients. Since patterns are hypothesized to indicate biological processes active in a subset of patients, the group investigated the genes for interactions and strived to assign functional annotation to the patterns. Based on function assignment and knowledge of connections between biological processes and phenotype, the group hypothesized about phenotypic effects and tested on the clinical data. Embedding the cluster specific patterns in cross-cluster patterns enabled the integration into the common biological context.

I also developed a new graph mining tool and algorithm for mining patterns in very large graph datasets. The general problem is to find frequent subgraphs in a large set of graphs. There are various applications for this problem including in social networking and in biology. While there has been a lot of previous research on frequent (sub) graph pattern mining methods, in practice using these methods is challenging. First, existing graph mining methods require an input parameter, such as minimum support, which the user has to guess upfront. Guessing the right parameter value is hard as picking a low value makes the mining algorithm run extremely slow, while picking a high value might miss important patterns. Second, existing graph-mining methods are not progressive; consequently, the user has to wait for the algorithm to complete before seeing even the first result. What is desired in practice is an efficient, parameter-free, online frequent graph mining tool that the user can simply point the data to, and quickly start seeing results presented progressively in decreasing support order. I developed a Tool for Online Graph Analytics (TOGA) that achieves this goal. Extensive empirical evaluation demonstrates the efficiency and usability of our new method for frequent graph mining.

Graduate Research Assistant, Florida International University (Jan 05 – May 09)
Databases & Systems Research Laboratory (DSRL)
School of Computing & Information Sciences, Miami
Advisor: Professor Vagelis Hristidis

Job Description: During my graduate study from 2004-2009, I worked as a Research Assistant in the Databases & Systems Research (DSRL) Lab at the School of Computer Sciences, Florida International University. As a research assistant, I worked on various problems related to searching (information discovery) graph-structured databases. In particular, I worked on developing effective, efficient and user-friendly keyword-based information retrieval techniques for various domains including the web, bibliographic, biological and clinical databases. Limitations of current web search engines were pinpointed and novel methods proposed. A

prototype web search system was designed, implemented and tested. Query-specific web document summarization techniques were proposed and user surveys were conducted to measure user satisfaction. Limitations of authority-flow based graph search techniques are identified and novel techniques to overcome them were proposed, implemented and tested. Novel user-friendly information retrieval framework and techniques to query hyperlinked data sources in various domains (including bibliographical, biological and clinical domains) were prototyped and tested.

Intern at IBM India Research Lab (May – August 2008)

Supervisors: Raghuram Krishnapuram and Prasad M Deshpande

Job Description: During my research internship at IBM Research Lab in India from May 2008 - Aug 2008, I worked on implementing new techniques to facilitate automatic web information extraction using visual cues & regions. I studied & developed novel solutions for the problem. Information in web pages usually exhibits some visual pattern, which can be exploited for effective information extraction. A general framework that allows declarative specification of information extraction rules based on spatial layout was proposed. This framework is complementary to traditional text based rules framework and allows a seamless combination of spatial layout based rules with traditional text based rules. An algebra that enables such a system and its efficient implementation using standard relational and text indexing features of a relational database were developed. Finally, we demonstrate the simplicity and efficiency of this system for a task involving the extraction of software system requirements from software product pages.

TEACHING EXPERIENCE

Undergraduate Course Instructor, Florida International University (Spring 2008)

School of Computing & Information Sciences, Miami

Course Title: Introduction to Programming in Java

Job Description: Taught a batch of forty undergraduate students, the preliminary course in programming, intended especially for IT majors. Offered weekly lectures, constructed and maintained the class web page and graded bi-weekly assignments. Prepared and graded the midterm and final exams.

Teaching Assistant, Florida International University (2006 – 2008)

School of Computing & Information Sciences, Miami

Courses: Data Structures and Principles of DBMS

Job Description: Offered review lectures, prepared and graded bi-weekly assignments, graded the midterm and final exams in Data Structures class. Assisted in preparing and grading midterm and final exams and offered review lectures in the DBMS class.

Lab Assistant, Florida International University (2004 – 2006)

School of Computing & Information Sciences, Miami

Labs: Operating Systems, Computer Data Analysis, Introduction to Micro-computers and Computer Applications for Business.

Job Description: Conducted weekly lab sessions, prepared and graded the lab assignments.

PUBLICATIONS

Book

1. **Ramakrishna Varadarajan** : “*Ranked Keyword Search on Graph-Structured Databases* : Techniques for User-friendly, High Quality and Efficient Information Discovery on Data Graphs”. Publisher: VDM Verlag (February 24, 2010). ISBN-10: 3639237269. ISBN-13: 978-3639237269. ([Barnes & Noble](#) , [Amazon](#)).

Book Chapters

2. **Ramakrishna Varadarajan**, Vagelis Hristidis and Fernando Farfan: “*Searching Electronic Health Records*”. Book Details: “**Information Discovery on Electronic Health Records**”. CRC - Taylor & Francis, December 2009. Editor: Vagelis Hristidis.
3. Fernando Farfan, **Ramakrishna Varadarajan** and Vagelis Hristidis: “*Electronic Health Records*”. Book Details: “**Information Discovery on Electronic Health Records**”. CRC - Taylor & Francis, December 2009. Editor: Vagelis Hristidis.

Journal papers

4. Vagelis Hristidis, **Ramakrishna Varadarajan**, Paul Biondich, Redmond Burke and Michael Weiner: “*Information Discovery on Electronic Medical Records Using Authority-Flow Techniques*”. In BMC Medical Informatics and Decision Making, 2010.
5. Vagelis Hristidis, Yannis Papakonstantinou and **Ramakrishna Varadarajan**: “*Using Proximity Search to Estimate Authority Flow*”, IEEE Transactions on Knowledge and Data Engineering (TKDE) 2010.
6. **Ramakrishna Varadarajan**, Vagelis Hristidis and Tao Li: “*Beyond Single-Page Web Search Results*”, IEEE Transactions on Knowledge and Data Engineering (TKDE) 2008.

Conference papers

7. **Ramakrishna Varadarajan**, Felix Eichinger, Jignesh Patel, Matthias Kretzler: “*Molecular Re-Classification of Renal Disease using Approximate Graph Matching, Clustering and Pattern Mining*” (poster paper), ISMB 2010, Boston.
8. Felix Eichinger, **Ramakrishna Varadarajan**, Jignesh Patel, Matthias Kretzler: “*Towards a Molecular Classification of Kidney Diseases Based on Network Analysis*” (presentation paper), Rocky 2010 (8th international Rocky Mountain Bioinformatics Conference), Colorado.
9. Vagelis Hristidis, Eduardo Ruiz, Alejandro Hernandez, Fernando Farfan, **Ramakrishna Varadarajan**: “*PatentsSearcher: A Novel Portal to Search and Explore Patents*” (<http://www.patentssearcher.com/>). In 3rd International Workshop on Patent Information Retrieval (PaIR 2010), ACM CIKM 2010.
10. **Ramakrishna Varadarajan**, Vagelis Hristidis, Louiqa Raschid, Maria-Esther Vidal, Luis Ibanez and Hector Rodriguez-Drumond: “*Flexible and Efficient Querying and Ranking on Hyperlinked Data Sources*” (full paper), Extending Database Technology (EDBT) 2009, Saint-Petersburg, Russia. (Acceptance rate – 32.50% Impact factor – 0.90).
11. **Ramakrishna Varadarajan**, Vagelis Hristidis and Louiqa Raschid: “*Explaining and Reformulating Authority Flow Queries*” (full paper), IEEE 24th International Conference on Data Engineering (ICDE) 2008, Cancun, Mexico. (Acceptance rate – 19% Impact factor – 0.97).

12. **Ramakrishna Varadarajan** and Vagelis Hristidis: “A System for Query-specific Document Summarization” (full paper), ACM 15th Conference on Information and Knowledge Management (CIKM) 2006, Arlington, VA, pages 622-631. (Acceptance rate – 15% Impact factor – 0.90).
13. **Ramakrishna Varadarajan**, Vagelis Hristidis and Tao Li: “Searching the Web using Composed Pages” (poster paper), ACM SIGIR Conference on Research and Development on Information Retrieval 2006, Seattle, WA, pages 713-714. (Acceptance rate – 37% Impact factor – 0.94).
14. **Ramakrishna Varadarajan** and Vagelis Hristidis: “Structure-Based Query Specific Document Summarization” (poster paper), ACM 14th Conference on Information and Knowledge Management (CIKM) 2005, Bremen, Germany.

Current/Ongoing Research Work

15. **Ramakrishna Varadarajan**, Jignesh Patel: “Practical and Efficient Online Frequent Graph Mining”. (Under review), in ICDE 2012.
16. **Ramakrishna Varadarajan**, Vagelis Hristidis, Fernando Farfan: “Comparing Top-k XML Lists” (Under review), in Information Systems Journal.
17. Vijil Chenthamarakshan, **Ramakrishna Varadarajan**, Prasad Deshpande and Raghuram Krishnapuram: “WYSIWYE: An Algebra for Expressing Spatial and Textual Rules for Information Extraction”. (Under review).

AWARDS, HONORS AND FELLOWSHIPS

- **Dissertation Year Fellowship**, Florida International University, 2008-2009.
- **Presidential Fellowship**, School of Computing and Information Sciences, Florida International University, 2008-2009.
- **Outstanding Graduate Research Award**, School of Computing and Information Sciences, Florida International University, Miami, 2007.
- **Excellence Award**, School of Computing and Information Sciences (SCIS), Florida International University (FIU), Miami, 2006-2007.
- **Student & New Researcher Travel Support**, SIGIR 2006.
- **Graduate Committee Travel award** for IEEE ICDE 2008, School of Computing and Information Sciences, Florida International University.
- **Graduate Committee Travel award** for ACM CIKM 2006, School of Computing and Information Sciences, Florida International University.
- **Travel award** for IEEE ICDE 2008, Graduate Student Association - Florida international University.
- **Travel award** for SIGIR 2006, Graduate Student Association- Florida international University.
- **University Gold Medal and Shield** from University of Madras, Chennai, India. Program: B.E Computer Science and Engineering.
Rank: 5th in the University (First Class Honors) - Percentage Score: 90%.

PRESENTATIONS

- “Searching & Mining Graph-structured Databases”, invited talk, University of Connecticut, Storrs, CT, 2011.
- “Searching & Mining Clinical/HealthCare Databases”, invited talk, University of Notre Dame, South Bend, IN, 2011.
- “Searching, Extracting and Mining E-Commerce Data”, invited talk, Sears Holdings Corporation, Chicago, IL, 2011.

- “*Searching & Mining Graph-structured Databases*”, invited talk, Virginia State University, St. Petersburg, VA, 2011.
- “*Introduction to Graph Theory*”, invited talk, Virginia State University, St. Petersburg, VA, 2011.
- “*Searching & Mining Graph-structured Databases*”, invited talk, HP Labs, Palo Alto, CA, 2011.
- “*Searching & Mining Graph-structured Databases*”, invited talk, University of Michigan-Dearborn, Dearborn, MI, 2011.
- “*Searching & Mining Graph-structured Databases*”, invited talk, IBM Almaden Research Center, San Jose, CA, 2011.
- “*Searching & Mining Graph-structured Databases*”, invited talk, Greenplum Inc, San Mateo, CA, 2010.
- “*Web Information Extraction using Visual Regions*”, summer internship at IBM India Research lab, 2008.
- “*Explaining and Reformulating Authority Flow Queries*”, IEEE ICDE 2008, Cancun, Mexico (short presentation).
- “*Ranked Search on Data Graphs*”, PhD Proposal/Dissertation, Florida International University, School of Computing & Information Sciences, Miami.
- “*A System for Query-specific Document Summarization*”, ACM CIKM 2006, Arlington, VA.
- “*Searching the Web Using Composed Pages*”, Poster Presentation, SIGIR 2006, Seattle, WA.

TECHNICAL SKILLS

- Operating Systems – Windows, Unix, Linux.
- Programming Languages – C, C++, Java, Visual Basic, SQL, PL-SQL.
- Web Development and Software tools – HTML, JavaScript, Java SDK, NetBeans, Java Servlets, JSP, ASP, PHP, XML, Macromedia Dreamweaver, MS FrontPage, MS Visio, Rational Rose, Apache Tomcat Web Server (deployment & maintenance web applications).
- Database Systems – Oracle, IBM DB2, MSSQL Server, MySql, MS Access.
- Networking – Java TCP/IP, UDP & C++ TCP/IP networking.

PROFESSIONAL ACTIVITIES

- Program Committee Member in PIKM (4th Workshop for Ph.D. Students in Information and Knowledge Management) 2011.
- Program Committee Member in IWGD (The First International Workshop on Graph Database) 2010 & 2011.
- Program Committee Member in GraphQ (First International Workshop on Querying Graph Structured Data) 2010.
- Reviewer for Very Large Databases (VLDB) Journal, 2011.
- Reviewer for International Journal of Web Information Systems (IJWIS), 2011.
- Reviewer for Elsevier Data and Knowledge Engineering (DKE) Journal, 2010.
- Reviewer for Elsevier Information Sciences (INS), 2010.
- Reviewer for IEEE Transactions on Knowledge and Data Engineering (TKDE), 2009.
- Reviewer for Knowledge and Information Systems (KAIS), 2009.
- Reviewer for Elsevier Information Processing Letters (IPL), 2009.
- External Reviewer for Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2009
- Reviewer for Distributed and Parallel Databases (DAPD), 2008.

- Reviewer for Journal of Super Computing, 2008.
- Reviewer for Conference on Management of Data (COMAD), 2008.
- External Reviewer for International Conference on Data Mining (ICDM), 2008.
- External Reviewer for Workshop for Ph.D. Students in Information and Knowledge Management (PIKM), 2008.
- External Reviewer for Elsevier Data and Knowledge Engineering (DKE) Journal 2008.
- Student Reviewer for ACM Southeast Conference (ACMSE) 2006.

RELEVANT GRADUATE COURSEWORK

Principles of Database Management Systems, Principles of Data Mining, Introduction to Algorithms, Advanced Database Systems, Advanced Topics in Information Retrieval, Advanced Operating Systems, Introduction to Bioinformatics, Compiler Construction, Expert Systems and Advanced Software Engineering.

PROJECTS

Graduate research projects

- PatentsSearcher (<http://www.patentssearcher.com/>) search engine for patent documents exploits domain semantics to improve the quality of discovery and ranking. PatentsSearcher also offers other novel functionalities to help users locate and navigate relevant and important patents or applications.
- A web based demo of the Document Summarization paper available at <http://dbir.cis.fiu.edu/summarization>.
- A web based demo of the Composed Pages paper available at <http://dbir.cis.fiu.edu/WebSearch>.
- A web based demo of the Explaining ObjectRank paper available at <http://dbir.cis.fiu.edu/ObjectRankReformulation>.

Graduate coursework projects

- DBMS Principles – Internet Book Store project using MySql and Php.
- Data Mining – A web based Project in Data Mining for the Approximate Distance classification method.
- Software Engineering – Banking System project.
- Compiler construction – Developed a compiler for a subset of Java, called MiniJava.

Undergraduate coursework projects

- Term Paper Project: Design of a Database Management System using Java.
- Automation of banking using Visual Basic 6.0 and MS access.
- Automated Testing of the Product “Adrenalinnet” using WinRunner 6.0.

CURRENT VISA STATUS – H1-B (academic)

COUNTRY OF CITIZENSHIP - India

PROFESSIONAL REFERENCES

Professor Jignesh M. Patel, Professor,
Department of Computer Sciences, 1210 West Dayton Street,
University of Wisconsin-Madison, Madison, WI 53706-1685.
Phone: 608-263-7308 (W)
Email: jignesh@cs.wisc.edu

Professor Vagelis Hristidis, Assistant Professor,
School of Computing & Information Sciences,
Florida International University, University Park,
ECS 384, 11200 SW 8th Street, Miami, 33199.
Phone: 305-348-6500 (W)
Email: vagelis@cis.fiu.edu

Professor Louiqa Raschid, Professor,
Decision and Information Technology (DIT), Smith School of Business,
Affiliate Professor, Department of Computer Science,
University of Maryland at College Park, MD 20742.
Phone: 301-405-6747(W), 301-405-2228(W).
Email: louiqa@umiacs.umd.edu

Dr. Raghuram Krishnapuram, Manager, Knowledge Management,
IBM India Research Lab, Block D, Embassy Golf Links Business Park,
Bangalore 560071. India.
Phone: +91 80 4177 4597 (W) Fax: +91 80 4177 6279
Email: kraghura@in.ibm.com

Professor Tao Li, Assistant Professor,
School of Computing & Information Sciences,
Florida International University, University Park,
ECS 318, 11200 SW 8th Street, Miami, 33199.
Phone: 305-348-6036 (W)
Email: taoli@cis.fiu.edu
