

Rajendra Prasath, Ph.D

Senior Researcher (CSE)

rajendra@cse.iitkgp.ernet.in

drprasath@gmail.com

<http://cse.iitkgp.ac.in/~rpr>

DoB: 29/05/1974

Dept of Computer Science and Engineering

Indian Institute of Technology

Kharagpur – 721 302, West Bengal, India

+91 97344 29269 / +91 76020 80978

EDUCATION:

Ph.D [Computer Science and Engineering]

Indian Institute of Technology, Kharagpur

Topic : The Ordering of Contents in Multilingual Information Retrieval

M.Tech* [Computer Science and Engineering]

Indian Institute of Technology

Kharagpur, India

Ph.D [Mathematics – Computer Science]

University of Madras, Chennai

Thesis : Design of algorithms for shared resource allocation and certain selected problems in distributed systems

PGDOR [Operations Research]

Pondicherry University, Pondicherry, India

M.Sc [Mathematics]

Ramanujan Institute for Advanced Study in Mathematics (RIASM),

University of Madras, Chennai

* Thesis was nominated for The Best Application Thesis Award at IIT Kharagpur;

FIELD OF RESEARCH:

- Information Retrieval and evaluations, Content Filtering, Information Extraction, Web Ranking; Machine Learning - Text Categorization and Clustering
- Mining the Text - Textual Case Based Reasoning, Analysis of Textual content
- Unstructured text processing, web content extraction into structured data
- Distributed Algorithms [Message Passing Systems] and Complex Networks

AWARDS:

- **MHRD Research Fellowship**, Indian Institute of Technology, Kharagpur
- **Visiting Fellow**, The Artificial Intelligence Research Institute (IIIA), Spanish National Research Council (CSIC), Barcelona, Spain
- **Visiting Fellow**, Swedish Institute of Computer Science (SICS), Kista, Sweden
- **ERCIM Allan Bensoussan Post Doctoral Fellow**, Norwegian University of Science and Technology (NTNU), Norway [2009-2010; again in 2011-2012]
- **AICTE QIP Fellowship** by Govt of India during July 2006 – May 2008
- **Reviewer** for Journals: IEEE/ACM Transactions on Networking, Information Sciences, Journal of Convergence Information Technology and Several International Conferences
- **Editorial Member**, Journal of Data Mining and Knowledge Discovery, Bioinfo Publishers
- **Reviewer** for Springer - Verlag, Heidelberg, Germany and Society for Industrial and Applied Mathematics (SIAM), Philadelphia, USA.
- **Listed** in the 6th edition of Marquis' Who is Who in science and engineering, USA
- **University Research Fellowship (URF)**, University of Madras, Nov 2001 – Apr 2003
- **Copyright**: three of my tools were copyrighted in **CLIA** Project, Govt of India
- **Organizer**: Seminar on "Algorithms & Artificial Systems" at Univ. of Madras

LIST OF PUBLICATIONS:

BOOKS:

1. **R.Rajendra Prasath**, Message Passing Approaches in Interconnection Networks - Towards Distributed Applications, VDM-Verlag, Germany, September 2010, ISBN: 978-3-639-26732-7.

WORK IN PROGRESS:

R.Rajendra Prasath and Sudeshna Sarkar

1. Diversification of Search Results through the discovery of the numerous specific topics from overlay graphs (in Progress)
2. Exploring Numerous topics to assist Web Documents Retrieval towards comprehensive content search, (in Progress)

UNDER REVIEW:

1. **R.Rajendra Prasath** and Sudeshna Sarkar, A pseudo-relevance feedback based method to find comprehensive web documents, (Submitted)

PUBLISHED:

2012:

1. **R.Rajendra Prasath**, Vijai Kumar and Sudeshna Sarkar, Assisting Web Documents Retrieval with Topic Identification in Tourism Domain, (to appear)

2011:

1. **Rajendra Prasath**, Implicit Semantic Relations Identification through Distributed Representations for Effective Text Retrieval, DESIDOC Journal of Library & Information Technology, Vol. 31, No. 4, pp. 242-246, July 2011.
2. **Rajendra Prasath**, Shared resource allocation using token based control strategy in augmented ring networks, in: A. Abraham et al. (Eds.): ACC 2011, Part II, Communications in Computer and Information Science (CCIS) - Springer, Vol. 191, pp. 555-567, June 2011.
3. **Rajendra Prasath** and Pinar Öztürk, Finding Potential Seeds through Rank Aggregation of Web Searches, in: 4th Int. Conf. on Pattern Recognition and Machine Intelligence (PREMI 2011), Lecture Notes in Computer Science, Vol. 6744, pp. 227-234, June 2011.

2010:

4. **R. Rajendra Prasath** and Sudeshna Sarkar, Unsupervised Feature Generation using Knowledge Repositories for Effective Text Categorization (ECAI 2010), in: Proc. of the 19th European Conference on Artificial Intelligence, Lisbon, Portugal, pp. 1101 - 1102, August 2010.
5. **Rajendra Prasath**, Algorithms for distributed sorting and prefix computation in static ad hoc mobile networks, in: Proc. of the 2010 Int. Conf. on Electronics and Information Engineering (ICEIE 2010), Vol. 2, pp. 144 - 148, August 2010.
6. **Rajendra Prasath** and Pinar Öztürk, Similarity Assessment through blocking and affordance assignment in Textual CBR, in: Proc of the Reasoning from Experiences on the Web (WebCBR 2010), pp. 151-160, July 2010.
7. Pinar Öztürk and **Rajendra Prasath**, Recognition of higher-order relations among features in textual cases using random indexing, in: Proc. of the 18th Int. Conf. on Case-Based Reasoning (ICCBR 2010), Lecture Notes in Computer Science, Vol. 6176, pp. 272-286, July, 2010.
8. **Rajendra Prasath**, Learning Age and Gender using Co-occurrence of Non-Dictionary Words from Stylistic Variations, in: Proc. of the 7th Int. Conf. on Rough Sets and Current Trends in Computing, Lecture Notes in Computer Science, Vol. 6086, pp. 544-550, June 2010.
9. Pinar Öztürk, **Rajendra Prasath** and Hans Moen, Distributed Representations to detect Higher Order Term Correlations in Textual Content, Proc. of the 7th Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC 2010), Lecture Notes in Computer Science, Vol. 6086, pp. 740-750, June 2010.
10. **R. Rajendra Prasath**, An alternative time - optimal distributed sorting algorithm on a line network, in: Proc. of 6th Int. Conf. on Networked Computing (INC2010), pp. 64-69, May 2010.

2009:

11. M.Rustagi, **R. Rajendra Prasath**, Sumit Goswami and Sudeshna Sarkar, Learning Age and Gender of Blogger from Stylistic Variation, in: 3rd Int. Conf. on Pattern Recognition and Machine Intelligence (PReMI 2009), Lecture Notes in Computer Science, Vol. 5909, pp. 205-212, December 2009.

12. **R. Rajendra Prasath** and Sudeshna Sarkar, Improving text categorization using hyperlinks in external knowledge repository, in: Proc. of the First Norwegian Artificial Intelligence Symposium (NAIS 2009), pp. 79 – 90, November 2009.

13. **R. Rajendra Prasath**, *Token based control with fault tolerance in augmented rings network*, College Sadhana (ISSN: 0974 6838), Vol. 2, No.1, pp. 1-10, August 2009.

2008:

14. **R. Rajendra Prasath** and Sudeshna Sarkar, *Augmenting Texts for Improving Categorization using External Knowledge Repository*, College Sadhana (ISSN: 0974 6838), 1(1), pp. 1-8, August 2008.

2004:

15. **R. Rajendra Prasath** and P.Thangavel, Token based control algorithms for shared resource allocation in general networks, in: Proc. of 3rd Int. Conf. on Information, held at Hosei University, Tokyo, Japan, pp. 531-534, November 2004.

2003:

16. **R. Rajendra Prasath** and P.Thangavel, *Shared resource allocation using token passing strategy in interconnected networks*, Information – An International Journal, Vol. 6, No. 2, pp. 197-206, April 2003.

17. **R. Rajendra Prasath** and P.Thangavel, Token based control algorithm with central coordinators, in: Proc. of the National Seminar on Algorithms and Artificial Systems, Allied Publishers, Chennai, pp. 25-40, January 2003.

2002:

18. P.Thangavel and **R. Rajendra Prasath**, *A note on token based control in rings and linear arrays*, The Journal of CSI, 32(3), pp. 62-65, March 2002.

2001:

19. **R. Rajendra Prasath** and P.Thangavel, Shared resource allocation in Chordal ring networks using token based control mechanism, in: Proc. of the International Conference on Industrial Mathematics (ICIM'2000), held at Indian Institute of Technology, Madras during August 2001.

2000:

20. **R. Rajendra Prasath** and P.Thangavel, *Token based message passing in bidirectional ring extensions*, Journal of Madras University (WMY-2000 Special Issue) - Section B: Sciences, Vol. 52, pp. 145-159, 2000.

21. **R. Rajendra Prasath** and P.Thangavel, Shared Resource Allocation using Token based control strategy in ring extension topologies, in: Recent Trends in Mathematical Sciences, (Eds.) J.C.Misra and S.B.Sinha, Narosa Publishing House, New Delhi, pp. 53-63, December 2000.

22. **R. Rajendra Prasath**, Shared resource allocation using baked potato routing, in: Proc. of International Conference Asia-Pacific Telecom 2000 (APT – 2000) on Advances in Telecommunication and Information Technology, pp. 144-152, December 2000.

DEVELOPER ARTICLES / MONOGRAPHS:

1. **R. Rajendra Prasath** and Sumit Goswami, Graclus: A Fast Graph Clustering Tool, PC Quest, pp. 70-73, April 2012.

2. **R. Rajendra Prasath** and Sumit Goswami, Embedding Open Source Search Engine, PC Quest, pp. 87-88, November 2008.

3. **R. Rajendra Prasath**, The power of Parallel Computation – Basic Design Issues, in: "EXPLORA: Knowledge Unleashed", pp. 115-121, September 2002.

TECHNICAL REPORTS:

1. TamilNews -The Search Engine for Regional News, April 2007.

2. Routing Issues in Sensor Network, April 2007.

3. Interaction Styles using Haptic Devices - The Designer Approach, May 2007.

4. Modeling Epidemics on Complex Networks, March 2007.

5. Garbage Collection – Handling Dynamic Memory Management, November 2006.

6. Delay Tolerant Routing, November 2006.

PROFESSIONAL EXPERIENCE:

Organization : IIT Kharagpur *January 2009 – Till Date*
Title : Senior Research Fellow
Funded by : Ministry of Information Technology, Govt. of India
Responsibilities: **Research and Development of CLIA system**

Organization : NTNU, Norway *August 2010 – September 2010*
Title : Research Scientist, IDI, NTNU
Responsibilities: **Research and Development of TCBR system**

Organization : NTNU, Norway *August 2009 – July 2010*
Title : Post Doctoral Fellow (ERCIM)
Funded by : European Research Consortium for Informatics and Mathematics
Responsibilities: **Research and Development of TCBR system**

Organization : IIT Kharagpur *May 2008 – Jul 2009*
Title : Senior Project Officer / Researcher
Funded by : Ministry of Communications & Information Technology, Govt. of India
Responsibilities: **Research and Development of CLIA system**

Organization : MNM Jain Engineering College *July 2004 – July 2006*
Job Title : Assistant Professor
Students : Under graduates [B.E / B.Tech]
Responsibilities: **Teaching, Research & Administrative responsibilities**

Organization : University of Madras *Nov 1998 – June 2004*
Job Title : Research Scholar / Guest Lecturer
Students : Post graduate Students
Responsibilities: **Research, Teaching and Project**

VISITING POSITIONS / POST DOCS:

COUNTRIES VISITED: Gyeongju, South Korea (INC2010), Barcelona, Spain (ERCIM-1), Stockholm, Sweden (ERCIM-2), Warsaw, Poland (RSCTC 2010), Alessandria, Italy (ICCBR 2010 / WebCBR 2010) and Lisbon, Portugal (ECAI 2010)

	Position	Institution	Period	Nature of Work
1	Visiting Fellow	Norwegian University of Science and Technology (NTNU), Norway	Sept 30, 2011 to October 14, 2011	Talks, Research Collaboration
2	Post Doctoral Fellow(ERCIM)	Norwegian University of Science and Technology (NTNU), Norway	August 1, 2009 to July 31, 2010	Research and Developments
3	Visiting Fellow	IIIA - CSIC, Universitat Autònoma de Barcelona, Spain	May 23, 2010 to June 1, 2010	Research, Talks, Discussions
4	Visiting Fellow	Swedish Institute of Computer Science (SICS), Kista, Sweden	June 09, 2010 to June 16, 2010	Research, Discussions

INVITED TALKS:

1. **Invited Talk** at: CLIA and NER workshop held at DAIICT – Gandhi Nagar, Gujarat during September 28-30, 2012
2. **Invited Talk** at: Winter School on IR techniques at Indian Statistical Institute (ISI), Bangalore during December 21-23, 2010
3. **Invited Talk** at: VHNSN College, Virudhungan-626 001 on August 7, 2010
4. **Invited Talk** at: IIIA - Institut d'Investigació en Intel·ligència Artificial (Artificial Intelligence Research Laboratory), CSIC - Spanish National Research Council, Universitat Autònoma de Barcelona, Spain on May 25, 2010
5. **Invited Talk** at: Swedish Institute of Computer Science(SICS), Kista, Sweden on June 14, 2010

PROFESSIONAL ACTIVITIES:

1. **Program Committee Member**, The India-Norway Workshop on Web Concepts and Technologies 2011 (INWWCT 2011) held at IDI - NTNU, Norway during October 2-9, 2011
2. **Adhoc Track Coordinator [Tamil] in FIRE 2011**, The Third Workshop on Forum for Information Retrieval Evaluation (FIRE) held at IIT Bombay during December 2-4, 2011
3. **Organizing Committee Member**, The 8th International Conference on Natural Language Processing (ICON 2010) will be held at IIT Kharagpur, India, during December 8-11, 2010
4. **Member**, Self Organizing Systems (SOS) group at Dept of Computing Science, NTNU, Norway from August 2009 to September 2010
5. **Organizing Member**, Third National Conference on Nonlinear Systems and Dynamics, February 06-08, 2006 and VIII Ramanujan Symposium on Recent Developments in Nonlinear Systems, February 14-16, 2001 both at University of Madras, Chennai, India
6. **Convener and Coordinator**, The National Workshop on Network Security, held at MNM JEC [under Anna University], Thoraipakkam, Chennai, September 2004
7. **Organizer**, The National Seminar on "Algorithms & Artificial Systems" at University of Madras, Chepauk, Chennai during Jan 30 - Feb 1, 2003
8. **Resource Person and Expert**, Master's Program at ICE, University of Madras, Chennai, India during February 14-16, 1999-2004

REFERENCES:

1. **Dr. Sudeshna Sarkar**
Professor, Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur – 721 302, West Bengal. India
Phone: +91 03222 283494 (Off) / +91 3222 278985 (Fax) / +91 94340 16956 (Mobile)
Email: sudeshna@cse.iitkgp.ernet.in, shudeshna@gmail.com
Web: www.facweb.iitkgp.ernet.in/~sudeshna
2. **Dr. R.Sahadevan**
Professor, Ramanujan Institute for Advanced Study in Mathematics
University of Madras, Chepauk, Chennai - 600 005, India
Phone: +91-44-25360357 / +91-44-2472 5350 / +91 9444835592 (Mobile)
Email: ramajayamsaha@yahoo.co.in, sahadevan@unom.ac.in
Web: www.unom.ac.in/departments/mathematics/sahadevan_profile.html
3. **Dr. Pinar Ozturk**
Associate Professor, Department of Computer and Information Science (IDI)
Norwegian University of Science and Technology (NTNU)
316, IT - bygget, Sem Sælands Vei 7 – 9, NO-7491 Trondheim, Norway
Office: +47 735 51019 / +47 918 97451 (Mobile)
E-mail: pinar@idi.ntnu.no Web: www.idi.ntnu.no/~pinar
4. **Dr. Pabitra Mitra**
Associate Professor, Department of Computer Science and Engineering
Indian Institute of Technology
Kharagpur – 721 302, West Bengal, India
Phone: +91 3222 28 2356(O), +91 3222 256315(R), +91 3222 278985 (Fax)
Email: pabitra@cse.iitkgp.ernet.in, pabitra@gmail.com Web: <http://cse.iitkgp.ac.in/~pabitra>

MEMBERSHIPS / ORGANIZATION ASSOCIATED:

- Professional Member, Association of Computing Machinery (ACM)
- Member, Young Researchers Committee, World Federation of Soft Computing
- Member, International Rough set society, Warsaw
- Member, Norwegian Artificial Intelligence Society (NAIS), Norway
- Member, Information Retrieval Facility, Vienna, Austria and IRSS, INDIA

EXTRA CURRICULAR ACTIVITIES:

- | | |
|--------------------|---|
| LYRICS | Writing Lyrics in Tamil and Management articles in Magazines |
| WEB DESIGN | Designing Professional web pages using popular technologies: CSS, AJAX |
| SOCIAL WORK | Participated in National Social Service & Cadet Corps at School and College |
| HOBBIES | Reading, Writing stories / lyrics in Tamil, Watching Tennis, Mathematics |

TECHNICAL SKILLS:


- **Programming** : C, C++, JAVA, Perl, PHP, Servlets, XML, Swing, Javascript, CSS, Ajax, Shell Script
- **Databases** : MySQL, MS SQL, PL - SQL
- **Platforms** : Enterprise Linux, Solaris, Windows 2000, XP, Vista, Ubuntu Servers
- **Open Source** : Nutch, Lucene, Hadoop, Solr, Netbeans IDE, Eclipse IDE, Tomcat, Glassfish
- **Tools** : Matlab, Maple, LaTeX

RESEARCH CUM SOFTWARE DEVELOPMENT

Listed below are some of the projects I have developed / involved in:

1) CROSS LINGUAL INFORMATION ACCESS SYSTEM:

We have developed a Cross Lingual Information Access System with 9 Indian languages support. In this consortium project, we have developed ranking algorithms and its evaluation with various orthogonal queries. Also we have the responsibilities for the complete integration of the system with other developers of the CLIA consortia. User can give the query in his/her preferred language and can get results in other languages of his interest. I have been serving as the chief architect of CLIA system since its inception.

Output : CLIA System – Deliverables with the support of Nutch, Hadoop, Solr
Product:  with the support of 9 Indian Languages

As the Chief Architect, I have designed and developed the Core MODULES:

parse-cml : Main stream html cleaner and noisy text remover
index-cml : Lucene indexer
ranker : Content Specific Scoring
SpiSEEDS : A new spider to collect potential urls as seeds
detectDND : Near duplicate document detector

using

JAVA – Core Module / System development
Hadoop – To process massive data in parallel
Solr – The powerful text indexer and distributed searcher
Tomcat – To host search engine with multilingual support
Ajax, CSS – To capture user clicks and user activity & to present an elegant GUI

2) KEYPHRASE EXTRACTION FROM SINGLE DOCUMENT:

Purpose: Extract key phrases from single web document based on sentence co-occurrence data

JAVA – Core Implementations of the proposed KPE from single document.
Lucene, Solr, Hadoop – To index the extracted Key Phrases

3) NEAR DUPLICATE CONTENT DETECTION (NDD) OF WEB DOCUMENTS:

Purpose: Identify & eliminate near duplicate documents in multilingual web collections

JAVA – Core Implementations of the proposed NDD and IR systems.
Lucene, Solr, Hadoop – User logging info with their detection history

4) TOPIC ASSISTED (TOURISM / HEALTH [MEDHELP]) IR SYSTEM:

Purpose: “3rd Doctor” - To provide suggestions about a specific disease under diagnosis based on user experiences and their knowledge sharing.

WEB CRAWLER – For crawling medHelp Forum pages
JAVA – Core Implementations of the proposed CBR and IR systems
JAVAScript, AJAX – User Profile management, User clicks and Recommender Systems
mySQL, Lucene, Solr, Hadoop – User logging info with their search history

5) Web CBR: During my stay at IDI, NTNU, Norway, we have carried out this work 2) and 3).

Smart Tourism: Different parts of a text may not be equally important / relevant to a given query. We propose a method to extract and assign affordances to texts and then use these affordances to retrieve / suggest web pages using case-based representations

NUTCH – For crawling tourism specific web content
JAVA – Core Implementations of the proposed web CBR algorithm
Parse-CML – Tool to perform web document segmentation, content extraction, web page segmentation, noisy segment removal and so on.
AJAX, CSS, HTML 5.0 – To perform Web User Interfaces with Responsive web design to capture User Clicks and Behavior

Accidental Reports: Performed corpus preparation with Canadian Safety Reports (CSR Corpus, hereafter). It basically contains reports of the accident happened in four sectors, namely "aviation", "marine", "pipeline" and "rail". The size of the corpus has been collected with 1228 reports each in both English and French languages. This corpus can also be considered as a parallel corpus since the translation of the lines in English and French reports are mostly aligned. These reports are ordered by year in each group.

Eagle Web Crawler – For crawling Accidental Reports
JAVA – Core Implementations of the proposed TCBR systems

6) DISTRIBUTED REPRESENTATIONS FOR TEXTUAL CBR:

Similarity of cases is based on the similarity between their constituting features. Conventional CBR benefits from employing domain specific knowledge for similarity assessment. Correspondingly, TCBR needs to involve higher-order relationships between features. Here, we developed a Textual CBR system in which features / cases are represented using a distributed representation that captures higher order relations among features.

JAVA – Developed Core modules of Distributed Representations for Textual CBR
LUCENE – used for indexing random index vectors
PERL / Shell script – To automate the process execution

7) UNSUPERVISED FEATURE GENERATION USING EXTERNAL KNOWLEDGE REPOSITORIES:

Feature generation is done through contextual analysis of document text and used to explore the meaning of the text fragments based on the concepts identified and described by the humans. Here we have computed the semantic relatedness of text fragment by extracting hyperlinked structures of Wikipedia using content and structure mining.

JAVA – Extracting knowledge concepts from Wikipedia
Weka, Rainbow, lightSVM – Machine Learning tools to build classifiers.
PERL – to build word co-occurrence data in adjacency list format
Graclus – word clusters using the graph based clustering algorithm

8) TAMILNEWS - THE BI-LINGUAL SEARCH ENGINE FOR REGIONAL NEWS:

In this project, we have developed a bi-lingual search engine for searching the Tamil / English news items. The Parser filters noisy contents from news documents and the indexer indexes the texts documents and classifies them under different categories. Two types of the GUI are provided to give the user input in the regional language with soft keyboards. The searcher follows dictionary based approach for query translation in CLIR.

JAVA – Core System for Crawling, Indexing and Searching
Lucene – Text Indexer to index web content
JSP / Servlet / JavaScript – To provide Search Interface with more functionalities
mySQL – To store user profiles

9) STYLOMETRIC ANALYSIS:

In this work, we have reported stylistic differences in blogging for gender and age group variation. The results are based on two mutually independent features: use of slang words and variation in average length of sentences across various age groups and gender.

Perl – Web Crawler Modules
Shell Script – To perform tasks on text data
Weka – Machine learning tool to build classifiers

10) DISTRIBUTED ALGORITHMS FOR MESSAGE PASSING SYSTEMS:

During my research days at University of Madras, I took up doctoral research on distributed systems using message passing strategy in the presence of limited resources. During this time, I developed a discrete event simulator in C that performs distributed sorting using message passing approach on a line network.

Server – RedHat Enterprise Server 4.x
C – To build the discrete event simulator
gcc 3.2 – The GCC compiler used to compile the source code