

# Indrajit Bhattacharya

IBM Research, India  
Plot 4, Block C Institutional Area  
Vasant Kunj  
New Delhi, 110 070, India

(+91) 011 41292152 (Work)  
(+91) 9811137869 (Mobile)  
indrajit@cs.umd.edu, indrajitb@gmail.com  
<http://www.ibhattacharya.net>

---

## Research Interests

My interest lies in reasoning under uncertainty for analysis and prediction over heterogeneous data using probabilistic graphical models and other machine learning techniques. My research includes developing probabilistic and cut-based formulations for collective relational clustering for applications such as entity resolution in structured databases and document collections, and word sense disambiguation from multilingual corpora. More recently, I have been exploring applications of relational clustering in cross domain transfer of learning.

## Education

UNIVERSITY OF MARYLAND, COLLEGE PARK

Ph.D. in Computer Science  
7/2000 - 12/2006

INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR

B.Tech. in Computer Science  
6/1999

## Research and Professional Experience

RESEARCH SCIENTIST  
4/07 - current

IBM Research, India

I am working on pattern mining over large heterogeneous and noisy information sources. I am investigating research challenges around information integration and data cleansing, clustering of heterogeneous data and cross-domain learning.

RESEARCH ASSISTANT  
1/03 - 02/07  
Advisor: Lise Getoor

LINQS, University of Maryland

I worked on models and algorithms for resolving entities from ambiguous references in structured and semi-structured domains, like bibliographic and natural language data. I have proposed novel similarity measures for unsupervised entity resolution using relationships among entity references and have shown significant performance improvements over traditional methods in several real world datasets. I have also developed a query-based framework for resolving entities in real time. Additionally, I have proposed probabilistic models and novel inference algorithms that resolve entities by discovering latent group structures. For the problem of word sense disambiguation using multiple languages, I have developed generative models for bilingual corpora and have shown that they outperform existing sense disambiguation approaches in real datasets.

RESEARCH ASSISTANT  
6/02 - 12/02

Graphics Lab, University of Maryland

I developed faster rendering techniques and compact representations for 3D models using local similarity in datasets.

INTERN  
6/01 - 8/01

Virtio Corporation  
Campbell, CA

My research included design and implementation of a translator for Virtio's virtual prototyping language to SystemC.

PROJECT OFFICER  
6/99 - 5/00

Indian Institute of Technology  
Kharagpur, India

I was involved in the the National Semiconductor Corporation funded “Virtual Silicon” project in the Department of Computer Science and Engineering, working on verification techniques for the SDL-C prototyping language.

## Teaching Experience

TEACHING ASSISTANT  
Fall 2000 - Fall 2002

University of Maryland

I have been a teaching assistant for “Object Oriented Programming in C++” and “Data Structures”. My duties included managing and grading projects and assignments and teaching lab sections.

## Publications

### Journals and Book Chapters

- [1] “Entity Resolution”, Indrajit Bhattacharya and Lise Getoor, *Encyclopedia of Machine Learning*, Springer-Verlag (to be published)
- [2] “Collective Relational Clustering”, Indrajit Bhattacharya and Lise Getoor, *Chapter in: Constrained Clustering: Advances in Algorithms, Theory, and Applications*, (eds Basu, Davidson, Wagstaff), Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, June 2008.
- [3] “Query-time Entity Resolution”, Indrajit Bhattacharya and Lise Getoor, *Journal of Artificial Intelligence Research (JAIR)*, Volume 30, pages 621-657, 2007.
- [4] “Collective Entity Resolution in Relational Data”, Indrajit Bhattacharya and Lise Getoor, *ACM Transactions on Knowledge Discovery from Data (ACM-TKDD)*, Volume 1, Issue 1, March 2007.
- [5] “Collective Entity Resolution in Relational Data”, Indrajit Bhattacharya and Lise Getoor, *IEEE Data Engineering Bulletin, Special Issue on Data Quality*, June 2006.
- [6] “Entity Resolution in Graphs”, Indrajit Bhattacharya and Lise Getoor, Chapter in *Mining Graph Data*, L. Holder and D. Cook (Editors), Wiley 2006.
- [7] “An Object Oriented Fuzzy Data Model for Similarity Detection in Image Databases”, Arun K. Majumdar, Indrajit Bhattacharya and Amit K. Saha, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 14(5), September/October 2002.
- [8] “Quantified Computation Tree Logic”, Anindya Patthak, Indrajit Bhattacharya, Anirban Dasgupta, Pallab Dasgupta and Partha P. Chakrabarti, *Information Processing Letters*, Vol. 82(3), May 2002.

### Conferences and Workshops

- [1] “Cross-Guided Clustering: Transfer of Relevant Supervision across Domains for Improved Clustering”, Indrajit Bhattacharya, Shantanu Godbole, Sachindra Joshi and Ashish Verma, *IEEE International Conference on Data Mining (ICDM)*, Miami, December, 2009.
- [2] “Enabling Analysts in Managed Services for CRM Analytics”, Indrajit Bhattacharya, Shantanu Godbole, Ajay Gupta, Ashish Verma, Jeff Achtermann and Kevin English, *Industrial Track in The ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, Paris, July, 2009.
- [3] “Structured entity identification and document categorization: two tasks with one joint model”, Indrajit Bhattacharya, Shantanu Godbole, and Sachindra Joshi, *ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, Las Vegas, August, 2008.
- [4] “Online Collective Entity Resolution”, Indrajit Bhattacharya and Lise Getoor, *NECTAR Track, Conference on Artificial Intelligence (AAAI)*, Vancouver, July 2007.
- [5] “Query-Time Entity Resolution”, Indrajit Bhattacharya, Louis Licamele and Lise Getoor, *The 12th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, Philadelphia, August 2006.
- [6] “Relational Clustering for Entity Resolution Queries”, Indrajit Bhattacharya, Louis Licamele and Lise Getoor, *ICML 2006 Workshop on Statistical Relational Learning (SRL)*, Pittsburgh, June 2006.

- [7] “A Latent Dirichlet Model for Unsupervised Entity Resolution”, Indrajit Bhattacharya and Lise Getoor, *6th SIAM Conference on Data Mining (SDM)*, Bethesda, Maryland, April 2006. (**Best Research Paper Award**)
- [8] “Relational Clustering for Multi-type Entity Resolution”, Indrajit Bhattacharya and Lise Getoor, *11th ACM SIGKDD Workshop on Multi Relational Data Mining (MRDM)*, Chicago, August 2005.
- [9] “Similarity Searches in Peer-to-peer Databases”, Indrajit Bhattacharya, Srinivas Kashyap and Srinivas Parthasarathy, *25th International Conference on Distributed Computing Systems (ICDCS)*, Columbus, Ohio, June 2005.
- [10] “Deduplication and Group Detection using Links”, Indrajit Bhattacharya and Lise Getoor, *10th ACM SIGKDD Workshop on Link Analysis and Group Detection (LinkKDD)*, Seattle, August 2004.
- [11] “The University of Maryland Senseval-3 system descriptions”, Clara Cabezas, Indrajit Bhattacharya and Philip Resnik, *3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3)*, Barcelona, July 2004.
- [12] “Unsupervised Sense Disambiguation using Bilingual Probabilistic Models”, Indrajit Bhattacharya, Lise Getoor and Yoshua Bengio, *42nd Meeting of the Association for Computational Linguistics (ACL)*, Barcelona, July 2004.
- [13] “Iterative Record Linkage for Cleaning and Integration”, Indrajit Bhattacharya and Lise Getoor, *9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD)*, Paris, June 2004.

## Professional Activities

- Program Committee Memberships
  - *IEEE International Conference on Data Mining (ICDM)*, 2009
  - *International Joint Conference on Artificial Intelligence (IJCAI)*, 2009
  - *International Conference on Machine Learning (ICML)*, 2009
  - *ACM WebKDD Workshop*, 2008
  - *Annual Meeting of the Association for Computational Linguistics (ACL-HLT)*, 2008
  - *ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, 2008
  - *International Conference on Database Systems for Advanced Applications (DASFAA) (Demo)*, 2008
  - *IEEE International Conference on Data Mining (ICDM)*, 2007
  - *Twenty-Second Conference on Artificial Intelligence (AAAI)*, 2007
- Reviewer for Journals such as *IEEE Transactions on Knowledge and Data Engineering*, *ACM Transactions on Database Systems*, *IEEE Transactions on Neural Networks*, *ACM Journal on Data and Information Quality*, *Pattern Analysis and Applications*

## References

References available on request.