Uma Sawant
uma.sawant@gmail.com

+91-9731820917

I am currently a PhD candidate in IIT Bombay, with Prof. Soumen Chakrabarti as my thesis adviser.

Research Topic

Web search views the Web as a collection of documents and returns a list of documents ranked in the order of relevance in response to a user query. Identification of entities or objects in queries and documents (e.g. people, places, cars) allows us to do more; such as providing direct answers to many user queries. I am interested in this kind of "Entity-aware search". I am currently studying entity annotation, query interpretation and ranking components for the same.

Experience

- **Research intern at Yahoo! labs, barcelona** (Sept 2014 - Nov 2014)
  I worked on improving entity search results by considering the interaction between entity annotator and ranker components.

- **Research engineer at Yahoo! labs, bangalore** (July 2008 - July 2011)
  I worked on Yahoo! search projects which involved interpreting user queries and click feedback, understanding corpus and building ranking models. I mainly worked on Yahoo! Web search and Image search ranking models, which have been applied in production.

Publications and technical reports

- Mandar Joshi, Uma Sawant, Soumen Chakrabarti, **Knowledge Graph and Corpus Driven Segmentation and Answer Inference for Telegraphic Entity-seeking Queries**. In Conference on Empirical Methods in Natural Language Processing (EMNLP) 2014, Doha, Quatar.

- Uma Sawant, Soumen Chakrabarti, **Learning Joint Query Interpretation and Response Ranking**. In 22nd international World Wide Web Conference (WWW 2013), Rio, Brazil.

- Uma Sawant, Soumen Chakrabarti, **Features and Aggregators for Web-scale Entity Search**. Technical report, 2013.

- Soumen Chakrabarti, Rajiv Khanna, Uma Sawant, Chiru Bhattacharyya, **Structured Learning for Non-Smooth Ranking Losses**. In proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2008), Las Vegas.

- Aniket Dalal, Kumar N., Uma Sawant, Sandeep Shelke and Pushpak Bhattacharyya, **Building Feature Rich POS Tagger for Morphologically Rich Languages: Experiences in Hindi**. In proceedings of 5th International Conference on Natural Language Processing (ICON 2007), IIIT, Hyderabad.

Talks

- An overview of query interpretation and ranking for entity-aware search. Seminar at IIT Kanpur, August 2014

- Joint query interpretation and response ranking for entity-aware search. Presentation at Yahoo summer school, June 2013

Education

- Ongoing Ph.D. in Computer Science from IIT Bombay, Powai with **CPI 8.64/10.0** (2011 - to date)

- M. Tech. in Computer Science from IIT Bombay, Powai with **CPI 9.71/10.0** (2008).

- B. E. in Information Technology from Sardar Patel College of Engineering, Mumbai with **69%** (2005).

- H.S.C. from Sathaye College, Mumbai with **93.50%** (2001).

- S.S.C. from Paranjape Vidyalaya, Mumbai with **88.93%** (1999).

Achievements

- **Yahoo Superstar Award (Team)** to Image Search team, 2010.
- Recipient of the **Google India Women in Engineering Award**, 2008.
- Part of the 4-member team of IIT Bombay which secured **second prize in NLPAI-ML 2006** contest for shallow parsing of Indian languages, a national level contest in the field of natural language processing.
- Secured 99.63 percentile with All India Rank **96** in the **Graduate Aptitude Test in Engineering** (the entrance examination for the IITs), Test of Computer Science (2005).
- Ranked **second in all streams** of the Sardar Patel College of Engineering in first year of engineering (2001-2002).
- Ranked **15th** in the merit list by Maharashtra Board in H.S.C. (2001).
- Recipient of the **National Talent Search Scholarship** awarded by the Government of India (1999).

Masters thesis

- **Learning to rank for non-smooth ranking losses** (with Prof. Soumen Chakrabarti, IITB)
  Our objective was to learn a real-valued ranking function from the given labeled training data. Using structured learning paradigm, we directly optimize for non-smooth ranking losses like Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (NDCG) etc. The challenge here is to avoid the complexity in training structured learner by intelligently choosing training instances.

Selected Course Projects and Seminars

- **Entity Ranking and Evidence Aggregation**. (with Prof. Soumen Chakrabarti, IITB)
  In this seminar, I studied various challenges involved in Entity Ranking over unstructured corpus. The supporting evidence for a potential answer entity can come from different documents across the corpus. I considered multiple methods to evaluate an evidence snippet based on signals like term-entity proximity and term rarity and finally aggregate the evidence to induce ranking over entities.
- **Part-of-speech tagger for Hindi**, (with Prof. Pushpak Bhattacharyya, IITB)
  We built a statistical part-of-speech (POS) tagger for Hindi based on the Maximum Entropy Markov Model. The tagger used language independent as well as language specific features.
- **Statistical natural language modeling**, (with Prof. Soumen Chakrabarti, IITB)
  A statistical natural language model captures the statistical view of the natural language generation process. In this seminar, I studied in detail various popular natural language models, followed by an overview of the discriminative approach to language modeling.

Courses

Algorithms and complexity, Statistical foundations of machine learning, Web search and mining, Data mining, Graphical models, Convex Optimization, Natural language processing, Implementation techniques in relational databases, Advanced databases, Topics in machine learning

Technical Skills

- Java, C, C++, SQL, Matlab, Scilab, Perl, Shell scripting, Oracle, Mysql

Extra-Curricular Activities

- Technical Secretary for hostel 11, IITB, in year 2007-2008.
- Member of the organizing committee of SPACE, the cultural event and Nirmaan, the technical event of Sardar Patel College of Engineering in 2003.
- Interests include music, books, swimming.