

# Uma Sawant

uma@cse.iitb.ac.in

---

I am currently a PhD candidate in IIT Bombay, through a PhD co-op program between IIT Bombay and Yahoo!.

## Experience

- **Research engineer at Yahoo! labs, bangalore** (July 2008 - July 2011)

I worked on Yahoo! search projects which involved understanding user intent and feedback, converting data to features and building ranking models. I have worked on Yahoo! web search and image search ranking models, which have been applied in production.

## Publications

- 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.

## Educational Record

- 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

- Recipient of the **Google India Women in Engineering Award**, 2008.
- Part of the 4-member team of IIT Bombay which secured **second prize in NLP AI-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).

## Research Projects

- **Learning to rank for non-smooth ranking losses** (Masters thesis, under the guidance of 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.
- **Supervised learning of web-pages using visual layout as features**, (with Prof. Soumen Chakrabarti, IITB)  
Our objective was to extract the names and values of entity attributes from product web pages. A web page was interpreted as an undirected graph with various visual and structural properties. We created an extractor, by defining an undirected graphical model on the web page.

## Selected Course Projects and Seminar

- **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.
- **Sort order optimization using functional dependencies in the Volcano optimizer**, (with Prof. S. Sudarshan, IITB)  
A query containing “order by” or “group by” clauses imposes a sort order on the output given by the attributes in these clauses. We extended the Volcano optimizer to exploit the functional dependencies in reducing sort orders. In turn, this enabled the optimizer to choose the less expensive plans corresponding to the smaller sort orders.
- **Seminar on 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, Natural language processing, Implementation techniques in relational databases, Advanced databases

## Technical Skills

- Programming Skills: Java, C, C++, SQL, Matlab, Scilab, Shell scripting
- Operating Systems: Linux, Windows
- Database Systems: PGSQL, 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, literature, swimming.

## References

Will be provided on request.