

Rushin Shah

E-Mail: rnshah@cs.cmu.edu

Homepage: www.cs.cmu.edu/~rnshah

Phone: 412-475-7953

GHC 6412, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213

Objective

To obtain a full-time Software Development / Engineering position.

Interests

Machine Learning, Information Extraction and Retrieval, Natural Language Processing, Social Media, Recommendation Systems

Education

M.S., Language Technologies (GPA: 3.58 / 4.0)

Carnegie Mellon University, Pittsburgh, USA.

Expected Graduation: August, 2011

B. Tech and M. Tech (with Honors), Computer Science & Engineering

Indian Institute of Technology (IIT) Kharagpur, India.

May, 2008

Selected Coursework

Machine Learning, Information Retrieval, Web Application Development, Natural Language Processing, Analysis of Social Media, Software Engineering, Graduate Algorithms, Artificial Intelligence, Database Systems, Object Oriented Systems.

Work Experience

- **Graduate Research Assistant, Carnegie Mellon University** (Language Technologies Institute), with Dr. Robert Frederking and Dr. Anatole Gershman. September 2009 – Present [*Java, Perl, Stanford and UIUC NLP Tools*]
 - Designed a new approach to NER for resource-scarce languages that uses machine translation systems.
 - Developed a method to produce ensemble NER and co-reference systems by combining existing systems.
 - Created a large scale cross-document co-reference system and visualization tool.
 - Developed a system to extract biographical attributes of named entities from text.
 - **Visiting Research Scholar, University of Pennsylvania** (Linguistic Data Consortium), with Dr. Lyle Ungar and Dr. Mark Liberman. September 2008 – August 2009 [*Perl*]
 - Developed a novel machine learning approach to rank both word-sense and part-of-speech labels produced by a morphological analyzer for Arabic text. Our system achieved superior results to the published state-of-art.
 - System used to speed up manual annotation of Arabic corpora and also as a part of an educational tool.
 - **Summer Intern, University of Pennsylvania**, with Dr. Lyle Ungar. May 2006 – July 2006 [*Java, MALLET*]
 - Implemented a machine learning system to predict species labels for gene mentions in medical text.
-

Graduate Projects

- **Design and Implementation of a Related Entity Finding System.** Term Project, Experimental IR. [*Java*]
 - Created a system that takes a source entity and type of relation as input and finds related entities in a web-scale corpus. This system will be submitted to the TREC 2011 Evaluation.
 - **Topical Clustering and Summarization of Tweets.** Term Project, Analysis of Social Media. [*Java, Eclipse*]
 - Created a system to cluster tweets based on topics and perform summarization by selecting representative tweets from topic clusters
 - **Extensions to Carnegie Mellon's Question Answering system Ephyra.** Term Project, Software Engineering for Language Technologies. [*Java, Eclipse*]
 - **Using Machine Learning Techniques to Predict Sports Statistics.** Term Project, Machine Learning. [*Java, MATLAB*]
 - Implemented regression models to predict a NBA team's season performance based on its record in previous seasons.
-

Recent Publications

- **Topical Clustering of Tweets.** Kevin Dela Rosa, Rushin Shah, Bo Lin, Anatole Gershman, Robert Frederking. Proceedings of the ACM SIGIR 3rd Workshop on Social Web Search and Mining (SIGIR - SWSM 2011).
 - **Improving Cross-Document Co-Reference with Semi-Supervised Information Extraction Models.** Rushin Shah, Bo Lin, Kevin Dela Rosa, Anatole Gershman, Robert Frederking. Symposium on Machine Learning in Speech and Language Processing, 2011.
 - **A New Approach to Lexical Disambiguation of Arabic Text.** R Shah, P S Dhillon, M Liberman, D Foster, M Maamouri, L Ungar. Proceedings of Empirical Methods in Natural Language Processing, 2010.
 - **SYNERGY: A Named Entity Recognition System for Resource-scarce Languages such as Swahili using Online Machine Translation.** R Shah, B Lin, A Gershman, R Frederking. LREC 2010 Workshop on African Language Technology.
-

Skills

- **Languages:** Java, Perl, C/C++, MySQL
- **Machine Learning and Natural Language Processing Tools:** Stanford NLP Tools, MALLET, MATLAB, OpenNLP, Weka, MinorThird
- **Other Software:** Subversion, LaTeX, Eclipse