

Deepa Paranjpe

Contact adeepa_paranjpe@yahoo.com

Home Page http://www.geocities.com/adeepa_paranjpe/index.html

Professional Interests To build systems, experiment with real data and devise solutions for data mining, machine learning, information retrieval and natural language processing problems.

Education **MTech, Computer Science and Engineering** 2002 – 2004

Indian Institute Of Technology, Bombay, India

CPI : 9.3

Thesis Title : A Machine Learning Approach for Factual Question Answering

Advisor : Prof. Soumen Chakrabarti

B.E., Computer Science and Engineering 1997 – 2001

Visvesvaraya Regional College of Engg., Nagpur, India

Ranked third in the university

Industry Experience **IBM India Research Lab, Delhi** August 2004 – July 2005

I worked in the Knowledge Management group managed by Dr. Raghuram Krishnapuram. My work involved developing algorithms for matching and integrating “text” data from different sources for the purpose of schema discovery for large scale information integration. In an exploratory project dealing with understanding of text, extracting entities and semantic relationships between them, we explored the possibility of a machine learning based approach for identifying entities and relationships in text documents.

Internship at Yahoo! Inc., Bangalore April 2004 – June 2004

Improved the accuracy of the automatic product categorizer used for categorizing products related data. This work was done in collaboration with Ganesh Ramakrishnan under the guidance of Dr. Byron Dom. We developed an algorithm that uses the natural language phrase structure of the titles of the products to identify “keywords” in the title. We also developed a structure-sensitive framework that takes into account the layout and natural language structure of documents to be categorized. I was responsible for the learning algorithm used to identify “keywords” in phrased product titles.

Motorola India Electronics Limited, Hyderabad July 2001 – July 2002

I worked in the research group of the Semi-conductor Products’ Sector of Motorola. I developed a format-conversion tool for a simulator used for verification of digital systems. I also worked on a project related to VoIP which involved designing a gateway for voice communication between the IP network and Motorola’s proprietary iDEN(Integrated Digital Enhanced Network).

Semi-supervised clustering using relative feedback

December 2004

Developed an algorithm for semi-supervised clustering that uses relative supervision in the form of triplets of data points (x,y,z) indicating that x is more similar to y than z. This work was done at IBM in collaboration with Nimit Kumar and Krishna Kummamuru.

Query Interface for structured Data Sources

August 2004

Implemented an “easy-to-use” query interface for a DB2 database containing relational data or XML documents. The interface can handle both natural language queries as well as keyword (typed) queries.

Learning to tag with named entities

August 2004

Implemented a classifier-based entity tagger that learns, from annotated training data, the “rules” for tagging the named entities. This was done as part of an information integration project at IBM.

A Machine Learning Approach for Factual Question Answering

June 2003 – Dec 2004

Designed and implemented a question answering (QA) system which “gets better with practice” by applying machine learning techniques to logs of past questions and their answers (QA pairs). I was responsible for developing and implementing the learning algorithms for identifying the “key” elements in the natural language question. This work was done in collaboration with Ganesh Ramakrishnan under the guidance of Prof. Soumen Chakrabarti and Prof. Pushpak Bhattacharya.

Fine grained PageRank Algorithm with Text Bias

Jan – May 2003

Designed and implemented the PageRank algorithm operational at the micronodes of the DOM representation of every webpage. Topic sensitivity, in the form of text bias surrounding an out-going hyper-link, was used to bias the random walk. The results obtained were comparable with a text biased implementation of the HITS algorithm. This work was done under the guidance of Prof. Soumen Chakrabarti.

Learning Similarity Measures for Clustering

Jan – May 2003

Designed and implemented a supervised as well as semi-supervised learning algorithm for learning the similarity measure required to do clustering. This work was done under the guidance of Prof. Sunita Sarawagi.

Using the Aspect Model for Question Answering

Jan – May 2003

Investigated using the Aspect model in automatic question answering. The aspect model was used to cluster the various questions, depending upon the cue phrases, *who, when, which etc..* The cue words were discarded by WordNet in our earlier Question Answering model using Bayesian Inferencing. This work was done under the guidance of Prof. Soumen Chakrabarti.

Structural Risk Minimization and Support Vector Machines

July – Dec 2003

Presented a seminar on Structural Risk Minimization and Support Vector Machines (SVMs) under the guidance of Prof. Soumen Chakrabarti.

Finding Aggregates in Streaming Data

Nov – Dec 2002

Implemented a system that identifies aggregate information from data streams using wavelet transforms. This work was done under the guidance of Prof. Krithi Ramamritham.

Undeletion in Second Extended FileSystem

2000 – 2001

Implemented undelete in the ext2 filesystem, a thorough study of the Second Extended FileSystem was done and system calls were implemented to support undeletion of files.

Languages: C, C++, Java, Perl, Python, SQL, XQuery, Pascal, Fortran, Prolog, Scheme, Haskell

Operating Systems: Linux, UNIX, FreeBSD, Windows, Solaris

Database Systems: DB2, Oracle, BerkeleyDB

Publications

Semi-Supervised Clustering with Metric Learning using Relative Comparisons

Nimit Kumar, Krishna Kummamuru, Deepa Paranjpe
ICDM, 2006

A Structure Sensitive Framework for Text Categorization

Ganesh Ramakrishnan, Deepa Paranjpe, Byron Dom
ACM Fourteenth Conference on Information and Knowledge Management (CIKM) 2005, Bremen, Germany

Is Question Answering an acquired skill ?

Ganesh Ramakrishnan, Soumen Chakrabarti, Deepa Paranjpe, Pushpak Bhattacharya
WWW 2004

Soft Word Sense Disambiguation

Ganesh Ramakrishnan, Pritviraj, Deepa Paranjpe, Pushpak Bhattacharyya, Soumen Chakrabarti
International Conference on Global Wordnet 2004

Passage Scoring for Question Answering via Bayesian Inference on Lexical Relations

Deepa Paranjpe, Ganesh Ramakrishnan, Sumana Srinivasan
TREC 2003

Generic Text Summarization Using Wordnet for Novelty and Hard

Ganesh Ramakrishnan, Kedar Bellare, Chirag Shah, Deepa Paranjpe
TREC 2003

Competitions

Question Answering

Ganesh Ramakrishnan, Deepa Paranjpe, Soumen Chakrabarti
Text Retrieval Conference, 2003 (The Question Answering Track)

High Accuracy Document Retrieval

Ganesh Ramakrishnan, Deepa Paranjpe, Soumen Chakrabarti
Text Retrieval Conference, 2003 (The High Accuracy Document Retrieval (HARD) Track)

Retrieval of novel sentences

Ganesh Ramakrishnan, Deepa Paranjpe, Soumen Chakrabarti
Text Retrieval Conference, 2003 (The Novelty Track)

Scholarships

Department fellowship at IIT, Bombay

Teaching Assistant for

First course on Artificial Intelligence taken by Prof. Pushpak Bhattacharya.
Advanced course on Neural Networks taken by Prof. Pushpak Bhattacharya.
Programming Languages taken by Prof. Soumen Chakrabarti.

Courses Attended

Data Mining, Hypertext Retrieval and Mining, Operating Systems, Compiler construction, Theory of Computation, Algorithms and Complexity, Design and Implementation of Functional programming languages, Artificial Intelligence, Concepts in Programming Languages, Database Design and Implementation, Asynchronous Systems, Linear And Non-linear Optimization, Language and the Mind.

Other Interests

Hindustani classical music, reading fiction, painting, cinema, philosophy and trekking.