

Diptesh Chatterjee

Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur
Kharagpur – 721302
India

Phone: (+91)9564293052
diptesh.chh.1987@gmail.com
http://cse.iitkgp.ac.in/~dipteshc

Education

DEGREE	UNIVERSITY/BOARD	YEAR	MARKS/GRADE
Dual Degree(Bachelor of Technology and Masters of Technology) in Computer Science and Engineering	Indian Institute of Technology Kharagpur	2006-present	GPA: 9.05/10 after 8 semesters (ranked 4 th out of 32 students in the department)
Senior Secondary	West Bengal council for Higher Secondary Education	2006	94.5% marks
Secondary	Council for Indian School Certificate Examination	2004	97.4 % marks

Major Projects

SUMMER INTERNSHIP

Department of Computing Sciences, Simon Fraser University
Mentor: Dr. Anoop Sarkar

May 2010 – July 2010
Vancouver, Canada

During the course of the internship, I developed a C++ wrapper for implementing the Factored Language Model functions present in the SRILM language modeling toolkit. The internship also involved designing and implementing Factored Language Models containing morphology-based feature functions for languages having complex morphology and check the variations in translation quality.

SUMMER INTERNSHIP

Microsoft Research India
Mentor: Dr. Monojit Choudhury

May 2010 – July 2010
Bangalore, India

My main internship project was based on the Dorogovtsev and Mendes model which explains the degree distribution of word co-occurrence networks. We explored the fact that although this model can explain the degree distribution, it is not sufficient to explain some of the global topological properties of the networks. We conducted spectral analysis on networks to demonstrate this.

M. Tech Thesis: Correction of noisy sentences using a monolingual corpus

July 2010 – present

Guide: Dr. Sudeshna Sarkar
July 2010 – present

My work is based on auto-correcting noisy sentences using a corpus of the same language. Initially, I am modeling this as a decoding problem and selecting the n-best outputs based on the language model score. This work is still in its infancy and no substantial output has yet been obtained. If this approach fails, I plan to use Textual Entailment approaches to solve this problem.

Design Lab Project: A framework for cognitive experiments

July 2010 – present

Guide: Dr. Anupam Basu

This project is based on a number of cognitive tests along the likes present in the CANTAB test platform of Cambridge university. The aim of this work is to develop a framework using which we shall be able to administer a number of cognitive tests. I have developed the feature set in order to parameterize the different tests and currently working on developing the GUIs for the tests. The language being used is Python and for developing the GUIs I am using wxPython.

B. Tech Thesis: Bilingual Dictionary Extraction from Comparable Corpora

July 2009-April 2010

Guide: Dr. Sudeshna Sarkar

In this work we considered an Eigenvector based approach and a word co-occurrence based approach in order to derive bilingual dictionaries from English-Hindi non-parallel corpora. We considered only single-word terms in this work.

Term Project: Comparing the Properties of Query Logs and Natural Language

Feb 2010-April 2010

Guide: Dr. Niloy Ganguly

In this project, we modeled query logs and natural language as a set of word co-occurrence graphs and compared the properties of the networks. Based on the properties of the networks, we derived a parallelism between query logs and natural language sentences. The codes were written in Python and C++ and used the Networkx library in Python to compute the network properties

Multi-Document opinion summarization

May 2007 – July 2007

Guide: Dr. Sudeshna Sarkar

In this project, we have considered a statistical approach to document summarization . We have used the lexrank method of sentence ranking but we have used a different metric/formula for ranking. We have not concentrated on Natural Language Generation , and have extracted sentences from the documents based on the ranking formulae used. The entire program was coded in Python and we used the NLTK extensively.

Course Projects (Implementation-based)

♦ Music Information Retrieval

February – April 2009

Guides:Dr. Pabitra Mitra and Dr. Arun Kumar Majumdar

In this project, we developed a system for retrieving best-match music tracks and related metadata from a database based on an audio input. We used the windowed MFCC coefficients for characterizing the tracks and used a Dynamic Time-Warping Algorithm to retrieve the best-matching tracks based on the query audio signal. The coding was done in C++ and MATLAB. This was done for the course **Database Management Systems**.

- ♦ Developed a hotel management software, a game of snakes and ladders, a college student management system, and a multiple choice test taking system as part of the **Software Engineering** course. All the codes had GUI built using Java Swing. The instructor was Dr. Rajib Mall.

- ♦ Developed a Rule-based English-Hindi transliteration system as part of the **Speech and Natural Language Processing** course. The system outputs had 91.6% match with the top 2 outputs provided by Google Indic Transliteration. I tested it over 3000 words. The instructor was Dr. Sudeshna Sarkar.

- ♦ Developed a 16-bit processor with controller as part of **Computer Organization and Architecture Laboratory** course. The instructor was Dr. Dipanwita Roy Chowdhury.
- ♦ Developed a Compiler for a subset of C language as part of **Compilers Laboratory** course. The instructor was Mr. Goutam Biswas.

Software Skills

- ♦ Programming Languages: C, C++, Python, Java, Perl, Prolog
- ♦ Operating Systems: Linux, Windows, Sun Solaris
- ♦ Web development: PHP, HTML
- ♦ Hardware programming: MIPS Assembly Language, Verilog, SPIM Simulator
- ♦ Database: MySQL server, MS SQL server
- ♦ Tools: MATLAB, Weka, SRILM, NLTK, Orange toolkit, Flex, Bison, SWIG

Extra Curricular

- ♦ Head of the problems team of BITWISE 2010, a worldwide algorithm-intensive online programming contest organized by the Department of Computer Science and Engineering IIT Kharagpur.
- ♦ Head of the ENIGMA team of BITWISE 2010. ENIGMA is an online treasure hunt that acts as a prelude to BITWISE..
- ♦ Recipient of honorable mention certificate in “Overnite ACM-ICPC Multi-Provincial Programming Contest” held during Kshitij 2008 and came 6th in the same event in Kshitij 2009
- ♦ Secured 11th rank overall and the best fresher’s code award in “Mission Mars”, an AI programming contest held during Kshitij, the Techno-management fest of IIT Kharagpur in 2007.
- ♦ Ranked 3rd in freshers’ quiz in IIT Kharagpur in 2006.
- ♦ Captain of OPENSOFTE (inter hostel software development competition) team of Meghnad Saha Hall of Residence.