

Arkanath Pathak

Curriculum Vitae (generated on 21 Jan, 2017)

contact **Interests**

Address

Unit 312, 50 Murray St
Pyrmont NSW 2009
Australia

Phone

(+61) 478171051

Email

pathak.arkanath@gmail.com

Website

http://arkanath.com

Machine Learning and related fields in AI

Education

2012-2016

B.Tech. Computer Science
Cumulative GPA : 9.40/10

Indian Institute of Technology (IIT) Kharagpur

2012

Class XII, CBSE
Aggregate Percentage: 88.8

St. John's Senior Secondary School

2010

Class X, CBSE
Cumulative Grade: 9.6/10

Lord Mahavira School

languages

bilingual Hindi/English

Achievements

citizenship

Indian

2013, 2014

Our team qualified the online round of ACM-ICPC Amritapuri
Onsite Rank : 50 in 2013, 32 in 2014

2009, 2010

Selected in Regional Mathematical Olympiad
Secured rank 8 (Delhi Region) in 2010

coding handles

Codeforces : arkanath
(top rating: **1932**)

2010

Awarded with the KVPY Scholarship
Around 200 students selected from all over India

TopCoder : arkanath
(top rating: **1329**)

2011

Selected in National Standard Examination in Physics
Around 300 students selected from all over India

CodeChef : arkanath
(top rating: **5784**)

2012

Secured a rank of 313 in IIT-JEE
Around 5,00,000 students appeared from all over India

Experience

Software Engineer, Google, Sydney

September 2016 - Present

Working on Google Cloud platform.

Projects

Improving YouTube personalization using clustering of videos (Internship at Google, Bangalore)

People Involved: Sumit Sanghai (Internship Mentor), Vivek Sahasranaman, Ajai Tirumali

Duration: May - July, 2015

Description: We explored new ways to improve YouTube user profiles by trying to find ways of modelling interests that are not well represented by Knowledge Graph entities (e.g. "70's music"). Towards this goal, we used clusters of videos as users' features. The input data for the clustering was derived from video correlations due to user co-watches. We tried k-means, HAC and LDA to generate video clusters. We built a simple video recommendation system using clusters as features. We had to deal with large amounts of input data, and thus, had to use compute clusters for distributing the tasks. Consequently, the project also involved heavy usage of distributed frameworks, like MapReduce. We also tried ways to generate cluster descriptors using n-gram language models and video entities. My work at the internship was well appreciated and I was offered a full-time Software Engineering position at Google HQ as a result.

Clustering of mixed data by integrating fuzzy, probabilistic and collaborative clustering framework (Research project carried at Indian Statistical Institute, Kolkata)

People Involved: Prof. Nikhil R. Pal (Project Guide)

Duration: May - June, 2014

Description: A new algorithm for clustering data with both numerical and categorical attributes was developed. Our work was published in International Journal of Fuzzy Systems.

A two-phase approach towards identifying argument structure in natural language (B.Tech. Project)

People Involved: Dr. Pawan Goyal and Dr. Plaban Bhowmick

Duration: February 2015 - April 2016

Description: The project is related to a relatively new research problem called Argumentation Mining. We present a two-stage strategy for extracting argument structure in a natural language text with an underlying argument. The task of automated identification of argument structure is difficult since it involves the problem of natural language inference. Furthermore, in the case of arguments the relationship is much more complex. Our work was presented as a full paper at NLPTEA 2016, held in conjunction with COLING 2016.

Constructing character profiles and networks in Gladiator (2000 film) using an unsupervised approach (Term project for Natural Language Processing (CS60057 - IIT Kharagpur))

People Involved: Dr. Pawan Goyal (Project Guide), Akshay Gupta, Buddha Prakash, Gaurav Sharma, Sanyam Agarwal, Utpal

Duration: August - November, 2015

Description: Although there have been approaches to analyze the characters and their relationships in other contexts like novels, there have been almost no successful attempt to analyze the characters in a movie based on the transcript. In an unsupervised manner, our goal was to construct and study the character profiles, relationship structure, sentiment dynamics and other interesting aspects of the academy award winning movie Gladiator (2000) using its transcript and subtitles. The mid-report and final presentation for the project can be found at http://arkanath.com/projects/gladiator_nlp.html

Publications

- *Arkanath Pathak and Nikhil R. Pal, Clustering of Mixed Data by Integrating Fuzzy, Probabilistic, and Collaborative Clustering Framework*, International Journal of Fuzzy Systems, 2016, Volume 18, Issue 3, pp 339–348, DOI 10.1007/s40815-016-0168-y
- *Arkanath Pathak, Pawan Goyal and Plaban Bhowmick, Automated Identification of Argument Structure in Natural Language*, presented as a full paper at NLPTEA 2016, held in conjunction with COLING 2016.

Creations

Abwid : Artificial assistant which interacts with the public using a chat interface.

Music Walker : Web application to explore and listen to new songs using the likes of user.

MwBot : Bot which periodically recommends songs for the user.

FreshBackMac : Mac application for fresh desktop backgrounds.

SleekArchive, GuessWho15, NowPlaying for iTunes, AutoLyrica for iTunes, JhonnyM

Courses / Workshops

- Machine Learning, 2013 (at Coursera)
- Machine Learning, 2014 (at IIT Kharagpur)
- Advanced Machine Learning, 2016 (at IIT Kharagpur)
- Artificial Intelligence, 2015 (at IIT Kharagpur)
- Natural Language Processing, 2015 (at IIT Kharagpur)
- Information Retrieval, 2015 (at IIT Kharagpur)
- Bay Area Deep Learning School, 2016 (at Stanford)

Full list of courses available at <http://arkanath.com/kgpcourses.html>