# Rohit Dholakia

*1206-1750 Davie Street*
*Vancouver V6G 1W3*
*+1 (778) 714 9229*
*rdholaki@sfu.ca*
*https://github.com/rohitdholakia*


## Education

Sept 2012–April 2014 **MSc in Computing Science**, *Simon Fraser University*, British Columbia, Canada, *3.6/4.33*.
My dissertation, supervised by Dr Anoop Sarkar, resulted in significantly better translation models for four different and very resource-poor languages, using the approach of triangulation. Our results were published as part of a long paper at a leading Machine Translation conference, AMTA.

2006–2010 **B. Tech in Computer Science** , *SASTRA University*, Tamilnadu, India, *7.98/10.0*.
Under the guidance of Dr Sairam, I implemented Collaborative filtering algorithms for the MovieLens dataset

## Publications

Oct 2014 *Pivot-Based Triangulation for Low-Resource Languages*, AMTA 2014

## Work Experience

June 2014 - **Computer Scientist**, *Simba Technologies*, Vancouver.
- Project lead working on a real-time compression protocol for ResultSets which has been submitted as a JIRA to Apache Hive project (HIVE-10438)
- Organizing and leading the efforts to set up clusters on AWS for various internal projects

Feb 2014 - May 2014 **NSERC Engage Internship**, *2hat Security*, Kelowna.
- As part of a NSERC ENGAGE grant, built Machine Learning models to classify noisy chat snippets that could classify whether a chat snippet is abusive or not with 70% accuracy. Also wrote a Language Identification module for the chat snippets. Used in online games for kids.

Fall 2012 - Feb 2014 **Research Assistant**: Natural Language Lab, SFU

July, 2011 - June, 2012 **Software Developer**, *MoveInSync Technologies*, Hyderabad.
- Designed, implemented and deployed an end-to-end automation framework using Selenium 2 to build, test and deploy the product for every weekly build.
- Designed, and implemented a multi-threaded Java application to simulate large workloads for the application with the aim of finding loopholes that only came up at a larger scale.

July, 2010 - May, 2011 **Project Associate**, *SIEL*, IIIT Hyderabad.
- Led a Nokia-funded project to develop and deploy small yet accurate CRF models to enable short messages in 6 major Indian languages, on MeeGo platform using only 128MB of RAM.
- Improved transliteration engine of a cross-lingual search engine for Indian languages, using large-scale CRF models. Brought accuracy from 60% to more than 75% for all the languages.

## Projects

Spring 2012 Mining StackExchange
*Mined complete StackExchange dump( 22G). Found that majority of users stop contributing after attaining a badge, average time to answer is low all across StackExchange.*
*Mentor*: Dr. Jian Pei, Data Mining Lab, SFU
*Language*: Python

Spring 2012 Large-scale Truecasing using Distributed Language Models
*Using Redis, 200 concurrent clients were run to perform truecasing with a gigaword language model (generated using 1200 million words).*
*Mentor*: Dr. Arrvindh Shriraman, Systems Lab, SFU
*Language*: Python

Summer 2013 Implementing core algorithms of Machine Translation
*IBM Models, Decoding, Reranking and Evaluation metrics*
*Mentor*: Dr. Anoop Sarkar, Natural Language Lab, SFU
*Language*: Python

## Awards

Graduate Fellowship Fall 2013

## Technical Skills

Languages C, Java, Python, SQL
Toolkits Moses, Kriya, Eclipse, Redis, Apache Hadoop, Apache Hive

## Coursework

Big Data Mining, Algorithms, Natural Language Processing, Statistical Machine Translation