

# SumanSaurabh

Suman Saurabh

LNM Institute of Information Technology, Jaipur, India

Proposal Link: <https://wiki.apache.org/nutch/SumanSaurabh/GSoC2015Nutch>

## About me:

I am final year B. Tech student at LNM Institute of Information Technology.

I have been involved with Apache community and have successfully completed GSoC 2014. I built an OSGi module called Speech to Text Enhancement Engine for Apache Stanbol.

In my sophomore I did my intern in CVIT(Centre for Visual Information Technology) Lab at IIIT Hyderabad, India. Project titled "**Probabilistic Representation and Recognition**" for optimizing traditional recognition techniques used in Scene Text Recognition.

## Motivation: Why Apache Nutch though I have contributed to Apache Stanbol?

Before the start of my B Tech thesis on Big Data Analytics, I have done far and wide exploration of data and information. World wide web has brought enormous and ever-growing amount of data and information.

I have been building MapReduce application on Hadoop Framework for analyzing agricultural data. I developed theoretical understanding of Hadoop Framework and coded MapReduce programs for data mining and exploration but most techniques resembled boiler-plate solution of MapReduce design patterns.

Working on this project would allow me a deeper and practical knowledge of Hadoop Framework and its applicability with scalable Enterprise Applications.

I have not been involved with Nutch but I have deep fascination on Web Mining particularly "**How Google became Google?**". The amount of data web has made it an important source of research. Furthermore I want to do my higher studies in Web Mining and contributing to Nutch community would help me on it.

This opportunity would enhance my knowledge of web exploration and analysis. It would also allow me to build a relationship with Nutch community and their contributors at a professional level. Their guidance would expose me to broader ideas.

Email: <<MailTo(ss DOT sumansaurabh92 AT SPAMFREE gmail DOT com)>>

LinkedIn: <http://in.linkedin.com/in/ssumansaurabh>

Github: <https://github.com/sumansaurabh>

Portfolio Website: <http://sumansaurabh.tk/>

Blog: <http://ssumansaurabh.blogspot.in/>

...

---

[CategoryHomepage](#)