

Siddhartha Reddy Kothakapu

<http://www.grok.in/>
<http://www.linkedin.com/in/sidsr>

sids@grok.in
+91 998 028 7437

EDUCATION	M.Tech in Information Technology (3.71/4.00 CGPA) International Institute of Information Technology, Bangalore	Jun 2006
EXPERIENCE	Senior/ Software Engineer , Ziva Software Pvt. Ltd., Bangalore <i>Ziva Software</i> is a Venture Capital funded startup based out of Bangalore, providing products and services in the <i>Mobile Search</i> domain. Established in late 2005, Ziva has provided search solutions to different players in the mobile industry including several operators and content owners. In 2007, Ziva launched <i>Zook</i> , it's consumer facing mobile search system. Was the first non-founding employee to join the company. Have been a core member of the team, involved in most of the projects, and actively participated in defining the development roadmap for them.	Jul 2006 - Dec 2008
INTERESTS	Information Engineering—Retrieval, Extraction & Management, Scalable Architectures, Distributed & Cloud Computing, Usability & User Experience	
SKILLS	Languages: Perl, Java, PHP, Ruby, Javascript, C/C++, C# Frameworks/Libraries/Softwares: Hadoop, Nutch, Lucene, Solr, Sphinx, WEKA, Ruby on Rails, Prototype Others: HTML, CSS, JSP, Bash Shell Scripting	
SELECTED PROJECTS	Buddhi Online , Ziva Software A system to fetch <i>Yahoo! BOSS</i> Web search results in response to a user query, fetch the result web pages and extract <i>snippets</i> that can possibly provide an answer to the query right away. Developed a prototype of the system as part of a feasibility analysis. Working on building out the complete system in order to deploy it in production.	TEAM SIZE: 1 Nov 2007 - Present
	Buddhi , Ziva Software A dream to crawl the Web and automatically extract <i>information content entities</i> off the crawled pages. Have been involved in the project right from the conception of the idea. <ul style="list-style-type: none">◇ Designed the <i>Map-Reduce</i> based architecture for the system.◇ Implemented a <i>focused crawler</i> based on the <i>Nutch</i> open source crawler. The crawler has been deployed and regularly crawls major portions of the Web.◇ Introduced the concept of—and implemented—a <i>HTML reducer</i> that transforms complex HTML pages into minimal HTML for easier analysis. This is the first step of the analysis performed by the system.◇ Defined a <i>page abstractor</i> as a component that discerns logical structures from unstructured HTML. Implemented a <i>table abstractor</i> that identifies tabular structure in web pages—even when the HTML does not contain any tables per se.◇ Set up the <i>Hadoop</i> based infrastructure for the crawling and extraction. Initially set it up on a cluster of <i>Amazon EC2</i> machines with <i>Amazon S3</i> as a data store. Later moved it to a set of dedicated machines.	TEAM SIZE: 2-3 Apr 2007 - Present

Zook, Ziva Software TEAM SIZE: 2-3 Jan 2007 - Present

Ziva Software's flagship mobile search engine. The main promise is to provide exact answers to questions in the least number of clicks; implying that the search engine provides information right away rather than directing the user to another website that *might* have the information.

Have been involved in all aspects of the functioning of the search engine.

- ◊ Set up the search system infrastructure.
- ◊ Designed the *User Experience* and *User Interface* of the WAP and SMS interfaces.
- ◊ Responsible for deploying releases.
- ◊ Created a HTTP based API for accessing the search system.

Zivabot, Ziva Software TEAM SIZE: 1 Aug 2006 - Oct 2006

A *scalable, highly-configurable, generic scraping system* for crawling specific websites containing useful information and then extracting that information.

Was fully responsible for the design and implementation of the system. Configured the system and set it up to work with an initial set of websites; this provided the base content for the Zook mobile search engine.

Better Web Buys TEAM SIZE: 2 Feb 2006 - May 2006

A *faceted price comparison shopping search engine* with over 20 million items across more than 100 merchants.

Was completely responsible for the whole project right from fetching the product feeds, parsing and storing them in a database, indexing, searching, user experience and setting up the entire system on a cluster of machines. Used *Sphinx* as the backend full-text indexing system.

yadb – Yet Another DataBase, IIIT, Bangalore TEAM SIZE: 3 Aug 2005 - Dec 2005

A reasonably *complete RDBMS* with a buffered memory manager, B+ tree indexing and a complete database management module including support for basic SQL, table management, paged storage etc. Part of the *Advanced Databases* course.

Top News, IIIT, Bangalore TEAM SIZE: 1 Sep 2005 - Oct 2005

A system to bring out the top news of the hour by *clustering the news items* obtained from the feeds provided by several news sources. Any news being discussed by multiple sources within a given (variable) timeframe is marked as top news. A bi-partite graph based clustering algorithm was used for clustering based on just the title and summary of the news item. Project for the *Web Information Retrieval* course.

SiDuSE, IIIT, Bangalore TEAM SIZE: 2 Aug 2004 - Sep 2004

A *Simple Desktop Search Engine* for Linux/UNIX systems. Creates a Trie-based full-text index of the files on the system. Apart from allowing the basic *AND*, *OR* & *NOT* operators, the system supports complex querying through nested queries, phrase matching and prefix queries. Part of the *Mathematical Foundations of Computing* course.

SOS – S Operating System TEAM SIZE: 2 Sep 2002 - Oct 2003

A simple monolithic kernel with kernel level threads, kernel level IPC and advanced memory management.

PUBLICATION **Measures of “Ignorance” on the Web** Dec 2006

International Conference on Management of Data, Delhi, India

HONORS

- ◊ *Valedictorian* for the class for 2006 at IIIT, Bangalore.
- ◊ Was awarded the *Honeywell Scholarship* for academic excellence at IIIT, Bangalore.
- ◊ Was chosen as the *employee of the year* for 2007-08 at Ziva Software.