# Spoken Query Processing

Jeevan Joishi MT13036

Kunal Gupta MT13040

# Problem Statement

- In traditional Information System ,we require to type a keyword for retrieving the documents for this you have to transform your information need to query representation

- In Spoken Query Processing you are required to speak and that spoken query will go under various phase like Query Reduction and Selection of best sub query so that document according verbose query can be retrieved by help of top level sub query

- We select best sub query through Query Quality Predictor as specified in one of the research paper "Reducing  Long Verbose Query"

# Concept In Project

- Noun Detection: Detecting Nouns and noun phrases with the help of Apache Open NLP tools helps in getting the key concepts from the query.
  - POS Taggers-Apache Open NLP tools

- Quality Predictors: To obtain the best sub-query , two quality predictors were implemented viz.
    - IDF based Features
    - Query Scope

- Besides, a third Quality predictor was also studied and tried namely Simplified Clarity Score but was found to be in-efficient.
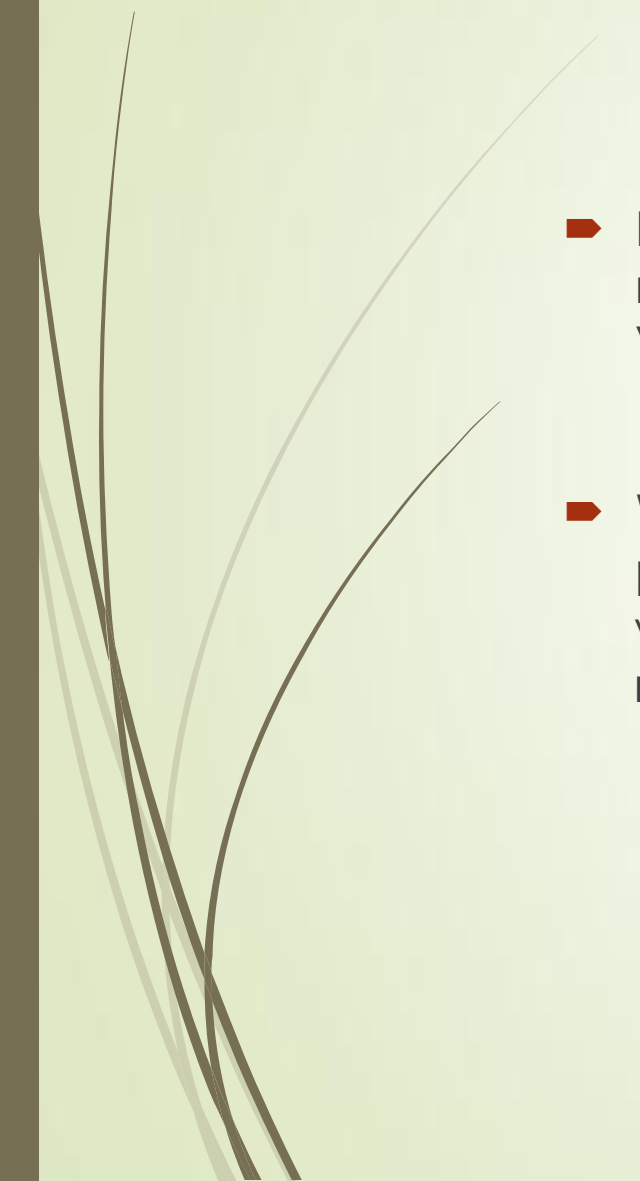
# Concept In Project

- Speech-to-Text: An android app was developed that helps in sending data to the system. Also a server was created that continuously listens to this app.

- Query Expansion and reweighting is based on Global Method of relevance feedback (Query Expansion with WordNet)

  - WordNet is based on Lexical English Database that contain network in which word act as concept and ISA relation is based on Synonym .

  - It may be the case that Corpus contain only automobile documents and user is searching about car so we can give such result by giving expansion in query.

  - To use WordNet we use JAWS API that create instance of WordNet and from that instance we can get Synonym from it of particular word.

# Concept In Project

- Ranking: To best study the efficiency of the search system, vector-space model was also implemented, Comparison between Dirichlet Model and Vector-Space Model was also studied.

- Web Interface: To make the whole app easy to use, the existing web pages was modified to a great deal to include support for textual search as well as spoken queries. Also options to change the quality predictors and ranking measures are also considered.

# ScreenShots
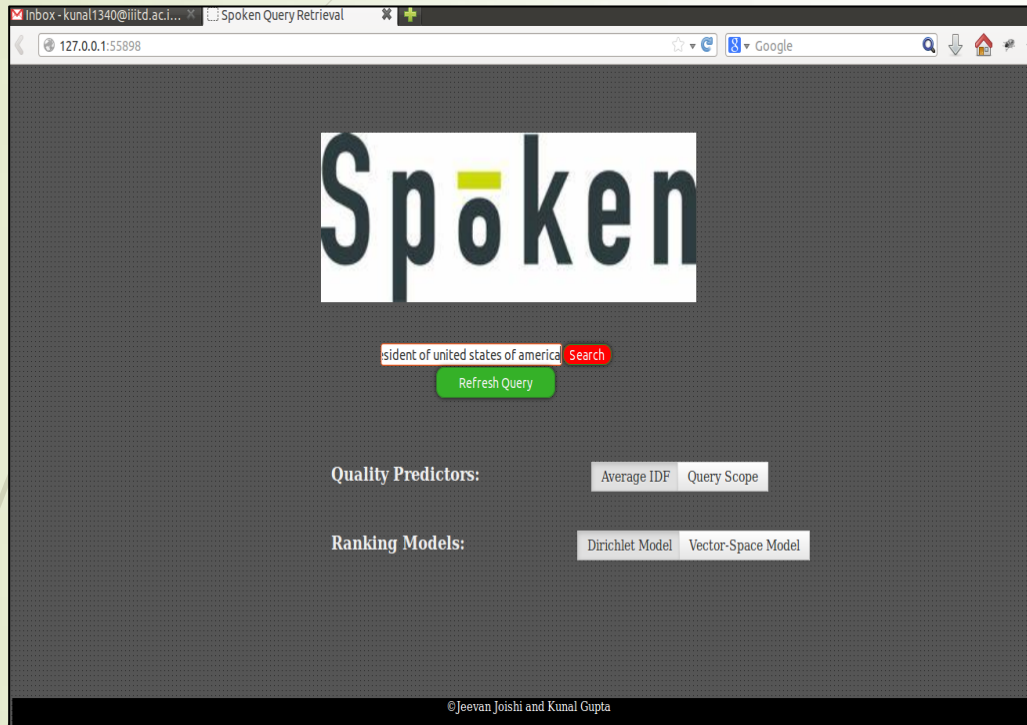


Home Page
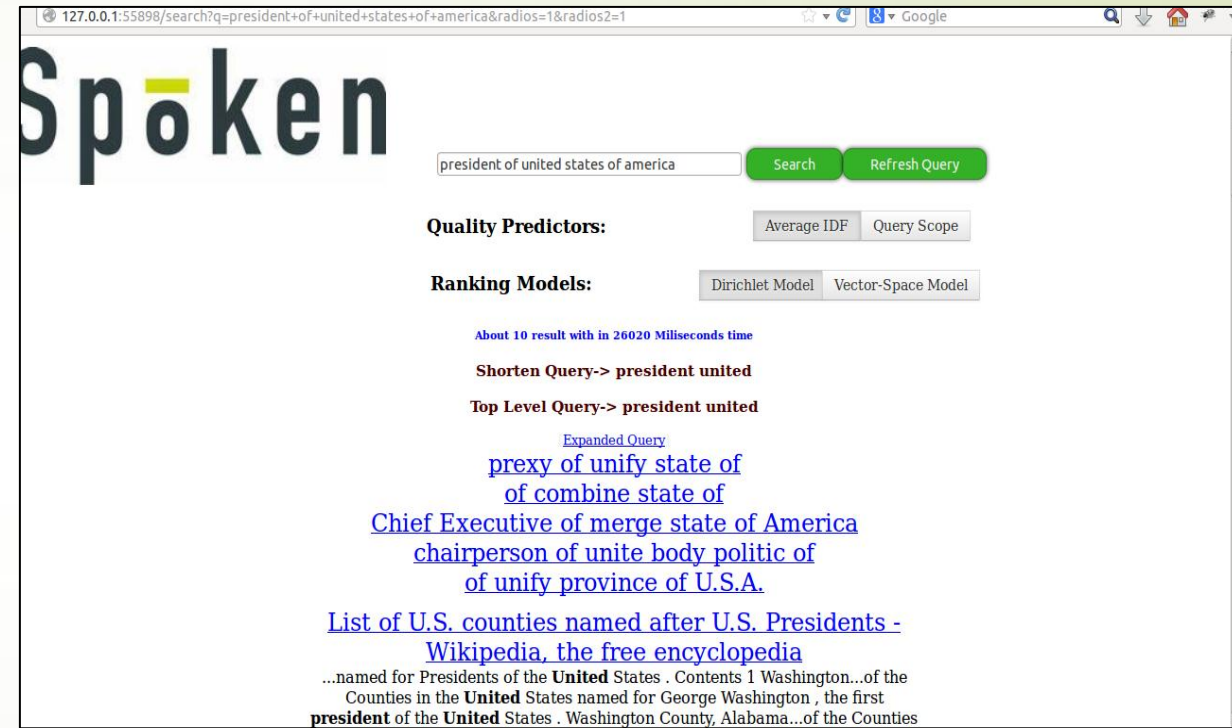


Results with some expanded queries

# Evaluation

**Query**: Indira Gandhi

| Rank | Doc-Dirichlet | R vs NR | Doc-Vector Space | R vs NR |
|------|---------------|---------|------------------|---------|
| 1 | Nehru Gandhi Family | 3 | Nehru Gandhi Family | 3 |
| 2 | Jatropha Carcus | 0 | Pupul Jayakar | 2 |
| 3 | Pupul Jayakar | 2 | Jatropha Carcus | 0 |
| 4 | Deptt of Law | 0 | Allahabad Airport | 1 |
| 5 | Cross Maidan | 1 | National FilmFare Award | 0 |
| 6 | Anti-Zionism | 1 | Tobacco | 1 |
| 7 | Allahabad Airport | 1 | Shankunugham | 1 |
| 8 | Tobocco | 1 | Kaithal | 1 |
| 9 | Tributo –a-soda | 0 | Tributo-a-soda | 0 |
| 10 | Shankunugham | 1 | Puteri | 0 |

# Evaluation(contd.)

**Query:** fa cup
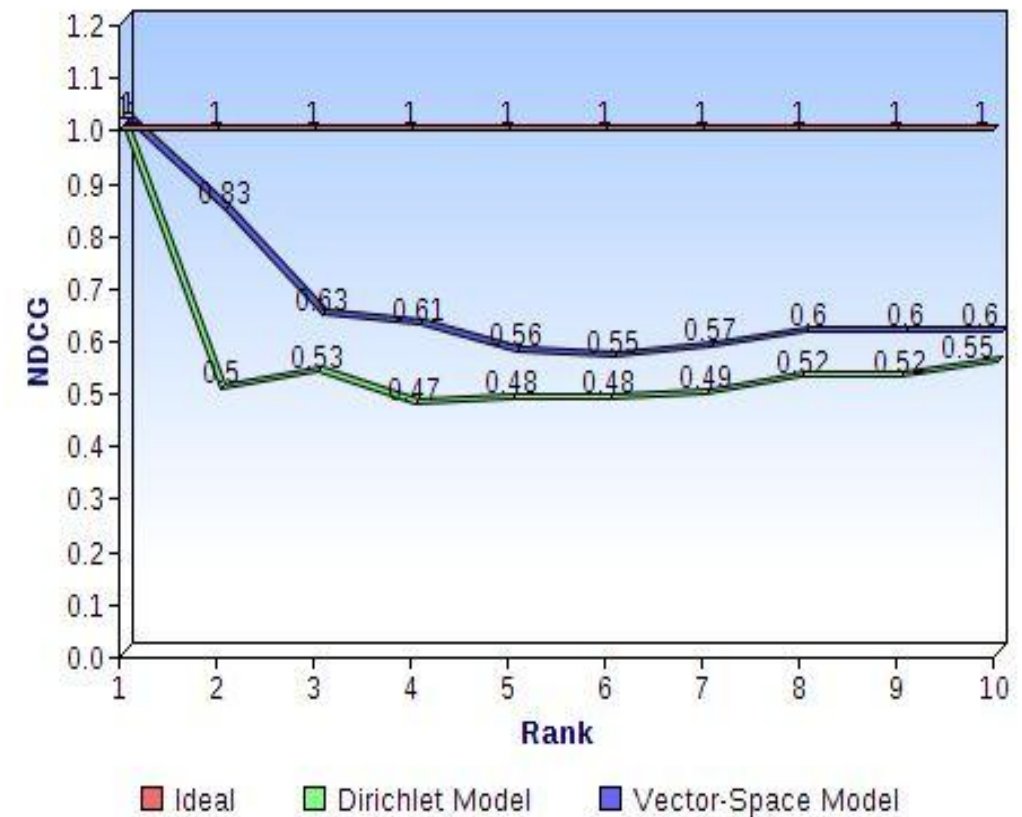
| Rank | Doc-Dirichlet | R vs NR | Doc-Vector Space | R vs NR |
|------|---------------|---------|------------------|---------|
| 1 | 1972-73 | 3 | Tottenham Hotspur | 3 |
| 2 | Frederik Lujenberg | 2 | 1972-73 | 3 |
| 3 | Tottenham Hotspur | 3 | Frederik Lujenberg | 2 |
| 4 | Isthmian | 2 | George Kay | 2 |
| 5 | Little Com.. | 1 | Timeline of Magic | 0 |
| 6 | FA Cup | 3 | Epsom & En… | 2 |
| 7 | George Kay | 2 | Toscana | 0 |
| 8 | J. Hyundai Motors | 0 | FA cup | 3 |
| 9 | Archie Godall | 1 | Stu Marshall Cup | 0 |
| 10 | Timeline of Magic | 0 | Stanely Cup | 0 |
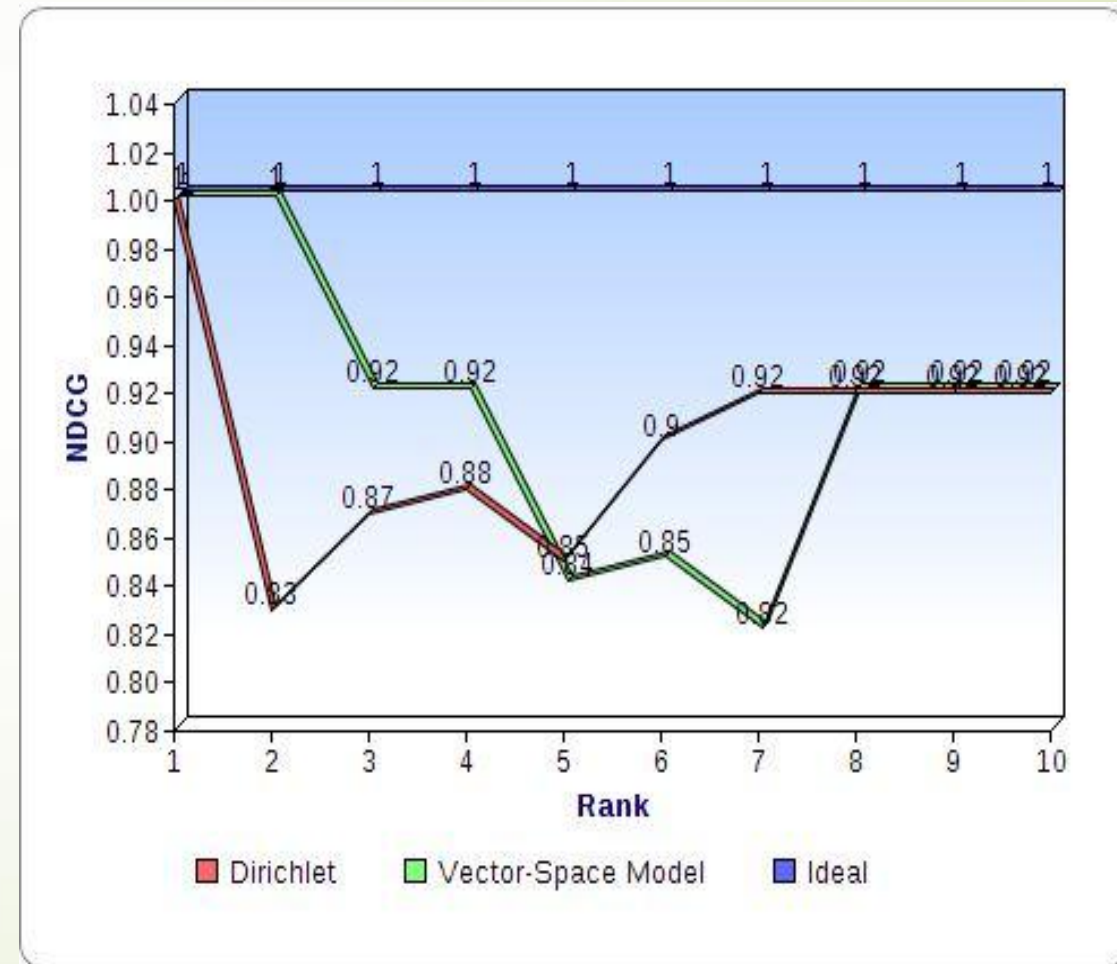
# Evaluation (contd.)

- Also for a set of displayed results, the precision of the clicked document and therefore the average precision can also be calculated out.
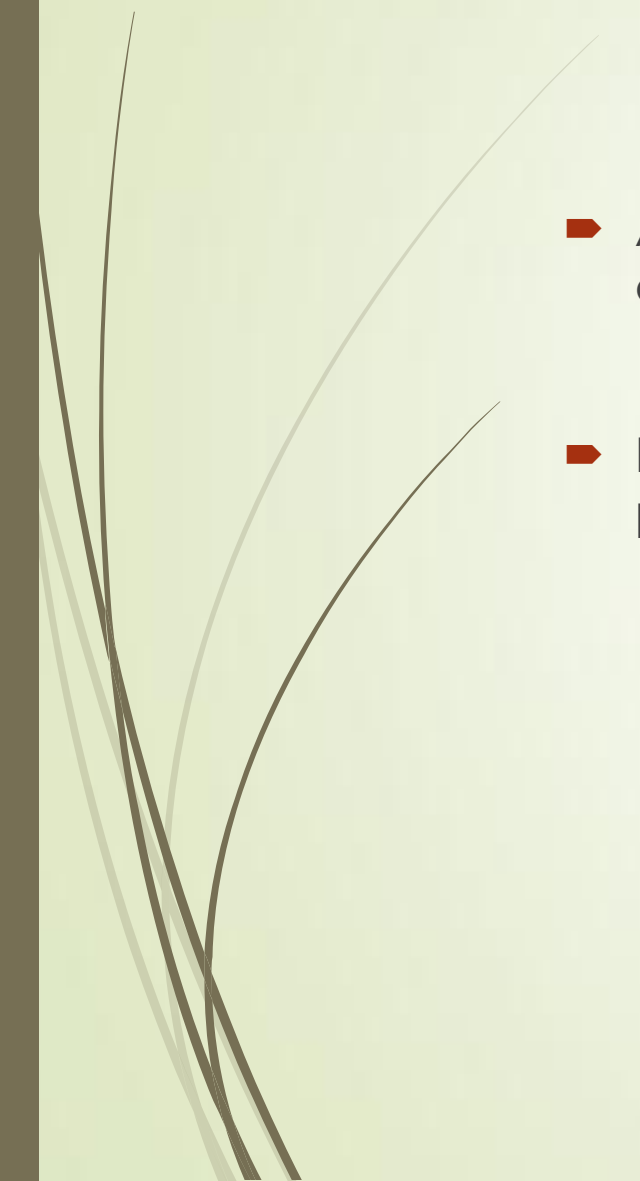
- Oops- only visible in console.

@ Javadoc   Declaration   Console ✕

App [Java Application] /usr/lib/jvm/java-7-openjdk-amd64/bin/java (05-May-2014 12:17:01 AM)
```
2014-05-05 00:17:03.091::INFO:  Logging to STDERR via org.mortbay.log.StdErrLog
2014-05-05 00:17:03.097::INFO:  jetty-6.1.5
2014-05-05 00:17:03.124::INFO:  Started SocketConnector@0.0.0.0:37453
Server: http://127.0.0.1:37453
No Expansion or Similar Word exist that contain the word form 'how'
No Expansion or Similar Word exist that contain the word form 'you'

Reduce Verbose Query
String passes to search box
No Expansion or Similar Word exist that contain the word form 'how'
No Expansion or Similar Word exist that contain the word form 'you'

Reduce Verbose Querypresidential middle
String Passes to Search Box  presidential middle
Precison of document ranked 2 have value  0.5
Average Precison of overall clicked is 0.5
Precison of document ranked 6 have value  0.3333333333333333
Average Precison of overall clicked is 0.41666666666666663
Precison of document ranked 10 have value  0.3
Average Precison of overall clicked is 0.37777777777777777
```

# Results

- Area between the ideal curve and the nDCG curve represents the quality of the IR technique.

- Plots from evaluation shows that vector-space model gives almost similar performance to Dirichlet Scoring method.