Search Engine Based On Permuterm Index

Course Project Report, Information Retrieval

DEADLINE HITTERS

J S Sunil - S20160010033

N.Prudhvi Krishna - S20160010057

K.Sai Prasanna Kumar - S20160010047

K.Upendra Sainath Reddy - S20160010038

10.09.2018
Information Retrieval

ABSTRACT

The main area in spotlight is to implement a search engine based on permuterm index to handle wildcard query. Google's search with wildcard queries doesn't give us proper results.This will help us get familiar with its importance in practical scenarios.

CONCEPTS THAT ARE FOCUSED IN OUR PROJECT

To implement a search engine based on permuterm index , we use the following concepts of IR :

PERMUTERM INDEXES

We use a special symbol **$** to mark the end of a term – Term “hello” is represented as **hello$**

A permuterm index contains various rotations of each term augmented with $ all linked to the original vocabulary term – The permuterm vocabulary: the set of rotated terms in the permuterm index

IMPACT OF THIS PROJECT AND WHO OTHERS WILL BENEFIT FROM OUR WORK

1.A user is uncertain about the spelling of a query term

EX: S*dney uncertain about Sydney or Sidney

2. A user is aware of multiple variants of spelling a term and (consciously) seeks documents containing any of the variants

EX: Color versus colour

3. A user searches documents containing variants of a term that would be caught by stemming, but is unsure whether the search engine conducts stemming.

EX: “judicia*” judicial versus judiciary

4. A user is uncertain about the correct rendition of a foreign word or phrase – “Universit* Stuttgart”