**CS6200-Information Retrieval**

**(Fall 2016)**

Members

**Ashok Koduru**

**Sravanthi Kethireddy**

**Frenia Pinto**

Under the guidance of

**Professor Nada Naji**

**Introduction**

The project demonstrates designing and building information retrieval systems, evaluating and comparing their performance levels in terms of retrieval effectiveness. The algorithms that are implemented for retrieval are:

1. Tf-idf
2. Cosine similarity
3. Lucene
4. BM-25 algorithm

The project also includes implementation of the query expansion technique using Pseudo Relevance feedback. Using stopping and stemming on corpus, two other runs are produced.

The runs produced by the retrieval models are evaluated using:

1. MAP
2. MRR
3. P@K, K = 5 and 20
4. Precision and Recall

**Contribution of the team-members:**

Ashok:

Sravanthi:

Frenia:

**Literature and Resources**

For implementation of the query expansion, ‘Pseudo Relevance Feedback’ approach is being used. The expansion terms generated by pseudo-relevance feedback will depend on the whole query, since they are extracted from documents ranked highly for that query, but the quality of the expansion will be determined by how many of the top-ranked documents in the initial ranking are in fact relevant. The derivational/inflectional variants, thesauri, ontologies are used to generate language-specific terms.

We have used this approach because derivational/inflectional variants are used to add the variants (parts of speech) terms to the query which may change the entire meaning of the query terms entered by the user. The thesauri and ontology adds synonyms to the query terms. This approach may not allow the user to find the exact document he is looking for. Also, pseudo relevance feedback is most effective and widely used.

**Implementation and Discussion**

**Results**

**Conclusions and Outlook**

**Bibliography**