1. Introduction
   1. Focus of the Search Engine – Matthew
   2. Architecture of the Search Engine – Subahasis
   3. Responsibilities
      1. Crawler – Stephen
      2. Index/Relevance Models - Ram
      3. User Interface – Subahasis
      4. Query Expansion – Matthew
      5. Clustering – Wyatt

|  |  |
| --- | --- |
| Responsibility | Solution |
| What you learned? | {Everyone fills this out} |
| What was your experience? |  |
| What were difficulties you faced? |  |
| How did you resolve them? |  |

1. Crawling – Stephen
   1. How many webpages gathered
   2. How the webpages were gathered
   3. how he passed the collection to index creation
   4. describe clearly how many webpages were crawled in the search engine
   5. details of the webpages that were crawled
   6. how were duplicates handled
   7. how was hyper link information provided to the students that generated the index and the relevance model
2. Indexing and Relevance
   1. How you assembled the index
   2. include a picture of how you assembled the index
   3. describe the web graph and how it was constructed
      1. Give statistics (see project template)
   4. show how information from the web graph was connected to the graph
   5. describe in detail two relevance models that you created and provide the weighting schemes that you have used
   6. give an example topic based page ranks computed
   7. discuss the hits score and show which webpages have obtained the largest score
   8. how interaction with user interface in generating queries to test the relevance models and to display the results of your search engine
      1. State clearly how many queries you have used,
      2. how you have generated them
      3. how you have judged the results of your relevance models
   9. Collaboration with clustering to improve relevance models
3. User Interface
   1. Design of user interface
   2. how you have worked with the student that has generated the index – how you have accessed the relevance models to provide the results in you user interface
   3. number of queries you have used for testing the search engine.
      1. How many were used in collaboration with the student that built the relevance models and how many did you generate on your own.
   4. collaborate with the student that produced clusters
      1. how use the clustering information for relevance and presentation on the interface
   5. How do you think you search engine compares to Google and Bing.
      1. Explain your judgments.
   6. how did you use the results of clustering in presenting the results of your search engine in the user interface
   7. how you have decided to select the queries for the demonstration of your search engine
   8. Provide three examples of the queries and the results produced by your search engine, as well as the results of Google and Bing.
4. Clustering
   1. how you have designed the flat clustering
      1. how many predefined clusters did you select
   2. What did you do with the results of clustering
   3. did you incorporate them in the relevance models
   4. did you provided to the user interface results that were obtained when clustering is used
   5. how did you use the results of agglomerative clustering.
   6. How many clusters did you obtain
   7. How were they presented on the user interface
   8. How many queries did you experiment with
   9. State clearly how many queries you have used to test the impact of the results of each clustering method, how you have generated them and how you have judged the results of your relevance models
   10. Discuss how you have decided to select the queries for the demonstration of your search engine
   11. Provide three examples of the queries and the results produced by your search engine and the clusters that you have created
5. Query Expansion
   1. Describe how you have selected 20 queries to test the Rocchio algorithm of your search engine.
      1. List them in your reports.
   2. Give examples of the web pages that you found relevant and those that you found irrelevant – and explain your judgements.
   3. Show also the modified queries that resulted by applying Rocchio to your original queries
   4. Discuss the 50 queries that you have used for pseudo-relevance feedback
   5. For each of the three methods, i.e. associative clustering, metric clustering and scalar clustering show
      1. (1) examples of 3 queries, the local document set and
      2. (2) the local vocabulary and set of local stems as well as their vocabularies;
      3. (3) show the values of the correlations you computed for the queries, and discuss you selection of the clusters and
      4. (4) show the resulting expanded queries.
   6. show the results of the search engine on your expanded queries and discuss them
   7. Elaborate on how you have collaborated with the student responsible for the user interface to expose the results of your expanded queries as well
   8. Discuss which queries and their expansion you selected for the demonstration of the project.
6. Discussion – all team {Matthew}
   1. Assumptions
   2. Algorithms/ Data Structures
   3. Collaboration - How we collaborated
7. Conclusion {Matthew}
   1. In conclusion…