

Report for homework 1



1772991, Richmond odotei, manasseh

1522685, YAO WATEBA, APPETI

1. 1772991, RICHMOND ODOTEI, MANASSEH

2. 1522685, YAO WATEBA, APPETI

**Part\_1\_1**

There are three (3) search engines, 222 queries with 222 ground truth values representing relevant documents for each query having query ids from 1 to 225 with query ids 31,119,215 missing. The ground truth for query id 157 has the highest number of relevant documents with 36 and 13 query ids with only one relevant document - [14, 22, 49, 81, 93, 95, 98, 105, 112, 135, 136, 142, 216]. All 222 queries were run on all 3 Search Engines and each returned 200 ranked documents per query, thus, 44,400 documents by each Search Engine.

**The P@K Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Search Engine | Mean(P@1) | Mean(P@3) | Mean(P@5) | Mean(P@10) |
| SE\_1 | 0.041 | 0.053 | 0.113 | 0.19 |
| SE\_2 | 0.137 | 0.142 | 0.142 | 0.149 |
| SE\_3 | 0.128 | 0.152 | 0.159 | 0.158 |

**R-Precision Table**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Search Engine** | **Mean(R-Precision\_Distribution)** | **min(R-Precision\_Distribution)** | **1°\_quartile (R-Precision\_Distribution)** | **MEDIAN(R-Precision\_Distribution)** | **1°\_quartile (R-Precision\_Distribution)** | **MAX(R-Precision\_Distribution)** |
| SE\_1 | 0.023 | 0 | 0 | 0 | 0 | 0.667 |
| SE\_2 | 0.255 | 0 | 0.429 | 0.25 | 0.429 | 1 |
| SE\_3 | 0.179 | 0 | 0.333 | 0.143 | 0.333 | 1 |

**MRR Table**

|  |  |
| --- | --- |
| Search Engine | MRR |
| SE\_1 | 0.081 |
| SE\_2 | 0.486 |
| SE\_3 | 0.395 |

**nDCG Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Search Engine | Mean(nDCG@1) | Mean(nDCG@3) | Mean(nDCG@5) | Mean(nDCG@10) |
| SE\_1 | 0.032 | 0.025 | 0.024 | 0.023 |
| SE\_2 | 0.302 | 0.241 | 0.227 | 0.187 |
| SE\_3 | 0.239 | 0.172 | 0.164 | 0.142 |

**Part\_1\_2**

There are three (3) search engines, 222 queries with 198 ground truth values representing relevant documents for each query having query ids from 1 to 200 with query ids 31,119 missing. There were no ground truth values for query ids from 201 to 225 and thus we excluded them while running Part\_1\_2 codes. The ground truth for query id 157 has the highest number of relevant documents with 36 and 12 query ids with only one relevant document - [14, 22, 49, 81, 93, 95, 98, 105, 112, 135, 136, 142]. The total number of documents returned was 44,400 documents by each Search Engine, but we used 39600 for our solution.

To assess the quality of the three different Search-Engines using the ground-truth and query-results, we use P@k, R-Precision, MRR and nDCG metrics. Since the “TetraSocialApp” can only display four outputs of the search result we restrict the k argument to k, thus we present P@4, nDCG@4, R-Precision for the first 4 results. The results after execution are;

* Precision@4:
  + SE\_1: 0.112
  + SE\_2: 0.1
  + SE\_3: 0.12
* R-Precision for the first 4 results
  + SE\_1: 0.284
  + SE\_2: 0.206
  + SE\_3: 0.289
* Mean Reciprocal Rate
  + SE\_1: 0.498
  + SE\_2: 0.417
  + SE\_3: 0.503
* Normalized Discounted Cumulative Gain (nDCG@4)
  + SE\_1: 0.24
  + SE\_2: 0.178
  + SE\_3: 0.245

From the above results, it can be observed that the scores of Search Engine 2 was consistently the lowest throughout all the metrics. On the other hand, Search Engine 3 had the highest scores throughout all the metrics followed by Search engine 1 slightly performing lower than the Search engine 3. Based on the outcome of the results Search Engine 3 module would be best for the ‘TetraSocialApp’.