Issouf kindo

Information Retrieval

1. Evaluation
   1. Problem 1.1
      1. **F-score** combines both recall and precision, so systems that favor are penalized for whichever is lower. So, for a specific user query if both precision and recall are weighted evenly, F1-score represents the fraction of relevant documents retrieved to the total number of relevant documents in the corpus and the number non-relevant documents retrieved.
      2. **Precision at K**: K is an arbitrary number chosen according to the goal of the project. Precision at K corresponds to the number of relevant results among the top K retrieved documents
      3. **Average precision**: Average precision computes the average value of precision at K scores.
      4. **R-precision** is the precision at R where R is the number of relevant documents to the given query.
   2. Problem 1.2

The motivation behind NDCG (Normalized Discounted Cumulative Gain

) is to help order the result of an IR model from the most important document to less important document

* 1. Problem 1.3

Get more information from the user to better specify user’s query will help the IR model better recommend webpages to the user. Additional query extension possibilities should be shown to help rephrase and direct the search

1. Link Analysis
   1. Problem 2.1

Web page is more important if it has more in-links

Web page is more important if it has in-links from important pages

* 1. Problem 2.2: Stochastic matrix:

Assumptions:

The violet bubbles are jumps (not considered)

A, B, C, D, E, F are webpages

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F |
| A | 0 | 0 | 0 | ½ | 0 | 0 |
| B | 0 | 0 | 1 | ½ | 1/3 | ½ |
| C | 0 | 1 | 0 | 0 | 0 | 0 |
| D | 0 | 0 | 0 | 0 | 1/3 | 0 |
| E | 0 | 0 | 0 | 0 | 0 | 1/2 |
| F | 0 | 0 | 0 | 0 | 1/3 | 0 |