Bronco ID:

Last Name:

First Name:

1. IR and text.

a. Querying a database is easier than querying text documents because (1) most of the information in text documents is unstructured whereas the information in databases is highly structured and organized; and (2) database queries have well-defined semantics that can be mapped to specific results whereas the meaning and information contained in text documents can be more opaque and more difficult to define than that contained in databases.

b. IR technologies have previously relied on text descriptions of multimedia content rather than the content itself. However, advances have been made that utilize direct comparison of multimedia content as part of an IR system. Some examples include YouTube’s protected content recognition system and various image recognition technologies (Google Vision, Amazon Rekognition, etc.).

2. Scope.

a. Web search engine – World Wide Web for relevant pages.

b. Vertical search engine – specialized web search where the domain is restricted to a specific topic or field. E.g. Amazon.

c. Vertical search engine – search of content within a corporate intranet to include relevant web pages, emails, documents, etc.

d. Desktop search engine – search of files contained on an individual computer.

e. Peer-to-peer (P2P) search engines – peer-to-peer search differs from the other types in that it involves finding information stored in networks of nodes without any centralized control. The network of nodes and their related content constitute the P2P search engine. Napster was an example of a P2P search engine.

3.

a. Task : classification. Classification uses a defined set of labels and automatically assigns those labels to documents.

b. Task : ad hoc search. Ad hoc search is a search based on a user query. This could be a search for research papers in Google Scholar.

c. Task : question answering. Question answering is similar to search but is aimed at finding an answer to a question rather than a search for content relevant to a query.

d. Task : filtering. Filtering involves detecting content of interest based on users’ profiles and history and subsequently providing an alert to a user indicating new or related documents.

4.

a.

b.

c.

d.

5. Precision and Recall

recall = {hits/(hits + misses)} \* 100%

precision = {hits / (hits + noise)} \* 100%

a. r = {2/(2+1)} \* 100% = 66%

p = {2/(2+1)} \* 100% = 66%

b. r = {3/(3+0)} \* 100% = 100%

p = {3/(3+2)} \* 100% = 60%

c. r = {2/(2+1)} \* 100% = 66%

p = {2/(0+2)} \* 100% = 100%

d. r = {0/(0+3)} \* 100% = 0%

p = {0/(0+2)} \* 100% = 0%

6. PageRank

A screenshot of a computer

Description automatically generated

7.