DCU

DUBLIN CITY UNIVERSITY

SEMESTER 1 EXAMINATIONS 2013/2014

MODULE: CA437 – Multimedia Information Retrieval

PROGRAMME(S):

      CASE   BSc in Computer Applications (Sft.Eng.)
      DME   B.Eng. in Digital Media Engineering
      ECSAO   Study Abroad (Engineering & Computing)

YEAR OF STUDY: 4,O

EXAMINERS:

            Dr. Liadh Kelly   (Ext:6725)
            Dr. James Power
            Dr. Ian Pitt

TIME ALLOWED: 2 Hours

INSTRUCTIONS: Answer 4 questions. All questions carry equal marks.

QUESTION 1   *[TOTAL MARKS: 25]*

Q 1(a)   **[3 Marks]**
What is the objective of an information retrieval system?

Q 1(b)   **[6 Marks]**
Users' information needs from information retrieval systems vary from verificative to explorative, precise to vague, and from shifting to static. Briefly explain this statement.

Q 1(c)   **[6 Marks]**
Before indexing documents several pre-processing steps can be performed. What is tokenization in this process? What is conflation? Why are conflation procedures used? List the two main classes of conflation.

Q 1(d)   **[10 Marks]**
Explain using a simple example the use of inverted files for query-document matching efficiency in information retrieval systems. Your example should include the uses of hashing algorithms for search term lookup.

[End of Question 1]

QUESTION 2   *[TOTAL MARKS: 25]*

Q 2(a)   **[4 Marks]**
What is precision and what is recall in information retrieval system evaluation?

Q 2(b)   **[7 Marks]**
Precision and recall can be calculated at different cut offs.

What does it mean to calculate precision and recall at different cut offs? Illustrate your answer with an example calculation.

Why are precision and recall calculated at defined cut offs for different systems?

Q 2(c)   **[10 Marks]**
Describe the concept of term weighting in classical best-match information retrieval. Include details of the components included in term weighting methods.

Q 2(d) [4 Marks]

What is the Robertson offer weight ($ow(i)$, where i is a term) and what is it used for in the context of the Okapi retrieval model?

[End of Question 2]

QUESTION 3 *[TOTAL MARKS: 25]*

Q 3(a) [4 Marks]

Give a concise definition of a document summary. In your answer contrast the possibilities for depth versus coverage in the summary when compared with the document being summarised.

Q 3(b) [4 Marks]

What is a query biased summary? What is the hypothesis underlying generation of query biased summaries?

Q 3(c) [6 Marks]

What are the advantages and disadvantages of the document translation and the query translation approaches to cross-language information retrieval?

Q 3(d) [11 Marks]

What is a question answering system?

Describe the design of a data-based question answering system. In your answer explain how quantity of information compensates for the quality of the data in answering questions from data sources of variable quality such as the WWW. Include examples to illustrate your answer.

[End of Question 3]

QUESTION 4 *[TOTAL MARKS: 25]*

Q 4(a) [5 Marks]

How are spoken document retrieval systems affected by errors in the transcription of words compared to retrieval systems working with a perfect accurate transcript of the same document?

Q 4(b) [6 Marks]
How can GPS and 'date-time' be used to facilitate user search in personal photograph search applications?

Q 4(c) [10 Marks]
What is the purpose of shot boundary detection methods?

Why is shot boundary detection useful in information retrieval for video data?

What are the basic features used to detect shot boundaries?

What problems are typically encountered in shot boundary detection for edited real-world video data such as films?

Q 4(d) [4 Marks]
How might information extraction methods be applied for video question answering? Your answer should consider separately the characteristics of the different media associated with video data streams.

[End of Question 4]

QUESTION 5 *[TOTAL MARKS: 25]*

Q 5(a) [4 Marks]
Speech recognition systems typically comprise two fundamental components: Acoustic models and a language model. Explain what acoustic models and a language model are in this context.

Q 5(b) [9 Marks]
Using examples explain the 3 levels of indexing used for image retrieval.

Q 5(c) [5 Marks]
What is enterprise search?

List two sources of information that are typically indexed by an enterprise search application.

Q 5(d) [3 Marks]

Typically enterprise search does not make 'all' of an organisation's content available for search. Provide three constraints on the content which can be made available.

Q 5(e) [4 Marks]

Would you regard precision or recall as more important in enterprise search applications? Explain your answer.

[End of Question 5]

QUESTION 6 [TOTAL MARKS: 25]

Q 6(a) [6 Marks]

Explain the difference between HTML and XML using examples to illustrate your answer.

Q 6(b) [10 Marks]

Why are link-based methods such as PageRank used in web search engines?

Outline the operation of the PageRank algorithm using an example.

Q 6(c) [5 Marks]

Analysis of the HTML markup of a web page can be used to modify search term weights such as those in the vector-space model or the BM25 model, in a web information retrieval engine. Using examples, explain how this is possible.

Q 6(d) [4 Marks]

It can be difficult to make good document snippet summaries for use in the ranked output of a search engine.

Suggest methods that might be used to overcome the limitations of snippet summaries to enable effective presentation of multimedia content in a search engines ranked list.

[End of Question 6]

[END OF EXAM]