



Contact during the exam:
Dept. of Computer and Information Science
Heri Ramampiaro, 73593440

IT2801 INFORMATION RETRIEVAL EXAM

Wednesday, 25th of May 2005.
Duration: 09.00 – 13.00 (4 hours)

ENGLISH

Allowed aid: D – No printed or written materials allowed. Only approved calculator is allowed.

Result deadline: 15th of June 2005.

Give **short and concise** answers to all questions. Short sentences are preferred rather than long explanation.

Problem I (25%)

1. Data vs. Information retrieval
 - a. What 3 characteristics that describe the differences between data retrieval and information retrieval (2%)
 - b. Explain briefly what are the aims of information retrieval. (1.5%)
 - c. Give two examples each on typical data, and information retrieval applications (1.5%)
2. What are the roles of index terms within information retrieval? Justify your answer. (4%)
3. What 5 important text operations should be used to facilitate the retrieval information? Explain what are the goals of these operations. (5%)
4. Explain shortly what the principles behind signature files. Use an example to support your explanation. (5%)
5. Explain what is the Rocchio equation used for. Use an example to support your explanation. (6%)

Problem II (25%)

1. Draw a block diagram (with rectangles and arrows) of a general architecture for multimedia information systems (MIRS). Explain briefly what are the roles of each block. (6%)
2. In the textbook there is a figure of the multimedia data model. What three layers does this model consist of? Provide at least one example on something that belongs to each layer. (6%)
3. Feature is a central notion in multimedia retrieval. What are the goals of features? What three requirements do feature extraction have to meet? (4%)
4. We can classify audio into speech and music.

- a. Draw a table that explains the differences between these two audio types. You have to include at least 4 features. (2%)
 - b. Show how you can use these differences to classify audio by using step-by-step classification. Hint: Draw a flow chart that shows the steps. (2%)
 - c. Is it possible to find the differences between speech and music based on the frequency spectrum? Justify your answer. (2%)
5. Provide a list of relevant features that you can use in retrieval of audio. (3%)

Problem III (30%)

1. Compare the Boolean and the vector-based models. What model would you prefer if you were building a text retrieval system, where it is important to present a ranked list based on relevance to the user. (5%)
2. Which two standard measures that we can use to evaluate an information retrieval system? Explain. (4%)
3. Extension of queries.
 - a. What are the aims of query extension? Explain. (2%)
 - b. Explain briefly the principle behind User Relevant Feedback. You may use a figure to support your explanation. (3%)
 - c. What are the differences between automatic local analysis and automatic global analysis? (4%)
4. Explain the principle behind inverted index. You may use example to support your answer. (6%)
5. Suffix tree and suffix trie are used to build indexes. Build a suffix trie and a suffix tree for the following text line: (6%)
 "A tornado can cause severe destructions. Twister is a movie about a tornado"

Problem IV (20%)

Answer with correct/wrong on the following statements. Each **correct** and **justified** answer will be given 2 points. **Each wrong answer** gets **-1.5** point. While **unexplained** and **no answer** is **0** point.

1. Colour pixels are suitable to compare two pictures.
(Correct/Wrong)
2. It is not practical to compare the distance between two audio files by summing the differences between samples from the two files.
(Correct/Wrong)
3. R-frame is representation of an average picture from a picture collection.
(Correct/Wrong)
4. Thesaurus is quite suitable for query improvements.
(Correct/Wrong)
5. Crawlers are used in both centralised and distributed web search engines.
(Correct/Wrong)
6. Video information can be retrieved by methods for other media types.
(Correct/Wrong)
7. An SQL-database is quite suitable for multimedia information retrieval.
(Correct/Wrong)
8. Audio retrieval systems can use techniques known from text retrieval systems.
(Correct/Wrong)
9. Movement information is not useful as feature for retrieval of video.
(Correct/Wrong)
10. Micon is an important feature for images, and is used in image retrieval. It is equivalent with index terms for text retrieval.
(Correct/Wrong)