LBSC 878 Oard/**Soergel**

Doug Oard oard@glue.umd.edu
Dagobert Soergel ds52@umail.umd.edu

April 20, 1999

A selection of comprehensive exam questions in information retrieval

1 Discuss the design of a system for the following task: Process an incoming stream of news and select news items of high importance for a general audience in the US (news items that would be candidates for the front page of the Washington Post or the New York Times).

Background: Users of bibliographic retrieval systems are usually interested in substantive data that help them solve a problem. Often data from two or more documents are helpful **in combination** even though no single document in the group, taken be itself, would be helpful or considered relevant.

2 Question 1. In light of this information, discuss the definition of document relevance and the measures and procedures used for evaluating bibliographic retrieval systems

3 Question 2. Discuss the design of a system that would

(a) Include such pairs of documents in the retrieved set (for simplicity, this question is restricted to pairs) and

(b) Assist the user with identifying such pairs in the retrieved set.

4 This question addresses the relationship of the classification of visual data and the classification of textual data

(a) Discuss the ways in which the classification of visual data and the classification of textual data are inherently similar or dissimilar

(b) Take some approach that has been used for the classification of visual data and examine its applicability to the classification of textual data. Similarly, take some approach that has been used for the classification of textual data and examine its applicability to the classification to the classification of visual data.

5 Design an information system for learner-directed exploration of themes in history

6 Consider the case of an information retrieval system that is designed to retrieve digital representations of aerial photographs for use by cartographers when creating maps and charts. Develop a design for the indexing and retrieval techniques in which the following issues are addressed and the corresponding design decisions are justified, both in terms of technical feasibility and in terms of utility for the intended purpose:

(a) What indexing features will be used? You are free to choose manually assigned features, automatically derived features, or any combination of the two approaches.

(b) How will the indexing features be obtained for each aerial photograph?

7 Design an information system for finding organizational units (down to small work groups) in a large organization such as the federal government. How would you evaluate the system.

8 Design a system for the automatic processing of documents that discuss scientific issues that have policy implications. The system should produce descriptors to characterize the following aspects of the document: The scientific subject matter, the viewpoint of the author / which side of an issue the author is on how strongly, explicitly stated policy implications / recommendations, policy alternatives supported by the scientific evidence presented (whether or not these alternatives are explicitly stated in the document). In addition to just producing descriptors, the system might also produce a summary covering these aspects.

9 Discuss ways of grouping retrieved images for display to the user. Relate the suitability of each display to the different purposes and cognitive abilities users may have.

10 Discuss ways in which the structure of a set of objects (documents, hypertext units, images) can be made explicit to the user.

11 (a) Design an interface that allows the user to interact with a conceptual structure, such as a classification, during the retrieval process.
(b) How could one measure the effect of using such an interface on the user's own cognitive structures.

12 A user retrieving images is not restricted to accepting the images as they are but can use

powerful image editing software to modify images. It may even be that the user is looking for inspiration, a model for creating his own image, rather than an image to be used directly. How does this affect the definition of image relevance?

13 Relevance studies have developed lists of document characteristics that people use to make relevance judgments. Based on your own knowledge, develop a hypothetical list of image characteristics that you think might be used that way.

14 Assume a writer wants to put together a volume of poems and images. She has a collection of poems and wants to find for each poem an image to match it. Discuss the design of a retrieval system for images that would support this task.

15 Several alternative versions of what is basically the same question:

Compare and contrast the need for browsing and the nature of browsing in bibliographic and text retrieval on the one hand and image retrieval on the other.

OR (more complex but more appropriate)

Discuss the need for browsing and the nature of browsing as a function of the [type of] information need and of the medium (esp. text versus images).

OR

Discuss principles for the arrangements of image representations (such as thumbnails) in response to a user's request.

16 An image retrieval system has a function that lets the user ask for "like" images. Discuss what it means for two images to be "like" each other.

17 Assuming that a powerful image recognition system is available, discuss the structure of a knowledge base that would assist with image recognition and subsequent iconographic analysis.

18 Design an expert system that builds a user model with a minimum of direct questions to the user relying instead as much as possible on the analysis of the regular user-system interaction during the course of a search and on appropriate "world knowledge".

19 An information company is developing an expert intermediary system to serve as a front-end to a system like DIALOG. A test version of the system is completed. The company calls you in to design a formative evaluation which would give them feedback for improving the system.

20 Design a study for this purpose: Formulate study objectives and sketch a design for data collection and analysis. What human subjects would you use? What would be your criteria for selecting them? How would you select search topics. What observations would you make, what data would you collect for each search session? How would you analyze the data to come to recommendations for improvements.

21 Picture an information system to support practicing physicians with information useful for dealing with individual patients. The system's database consists of the full text of medical textbooks and journal article. Its output consists of useful passages extracted from that text base. Sketch your approach to designing such a system.

22 Develop a set of indexing instruction that would make sure that aspects relevant for practicing physicians are captured by the indexers

23 Design an expert system for automatic request-oriented indexing.

24 Discuss the role of classification in a hypertext system.

25 Design a study to compare the retrieval performance of a system such as ERIC or MEDLINE with controlled vocabulary searching and with free-text searching. The purpose of the study is to gain insights which would help to improve both modes of searching.

26 How could models of human information processing, especially memory, be applied to an analysis of how people make relevance judgements?

(a) How could such an analysis account for any discrepancies that might exist between the subjective relevance judgements of a user and the "objective" relevance which is ascertained in terms of the potential contribution to a solution of the user's problems?

(b) How could such an analysis be used to build a computer system that would evaluate a user's relevance judgement?

27 Consider the hypothesis that an index language that mirrors the mental patterns of its users has a positive effect on ISAR system performance.

(a) Based on your knowledge of the functioning of ISAR systems, give arguments for and against of this hypothesis.

(b) Design a study to test this hypothesis.

28 Design a user system interface that would support problem-oriented retrieval in a free-text retrieval system.

.