# CS4065 - Multimedia Search and Recommendation

## Exam

# Core & Systems

4 July, 2018, 13.30 -16.30 h

- This exam has 5 questions, for which a total of 40 points can be obtained.
- The allotted time for this exam is 3 hours.
- Use of a calculator is permitted.
- For each of the 5 questions, start your answer on a new page.

*This page intentionally left blank*

# 1. Multimedia (4 sub-questions, 9 points in total)

Last year, the lectures of CS4065 were recorded on Collegerama. Example metadata and video screenshots are given in Figure 1 and Figure 2.

CS4065_01 Introduction

Info | Watch

A. Hanjalic
C.C.S. Liem

Tuesday, April 25, 2017
3:45 PM WEDT
1 Hour 23 Minutes 29 Seconds
231 views

**Figure 1.** Sample metadata for the recorded CS4065 2016-2017 lectures.

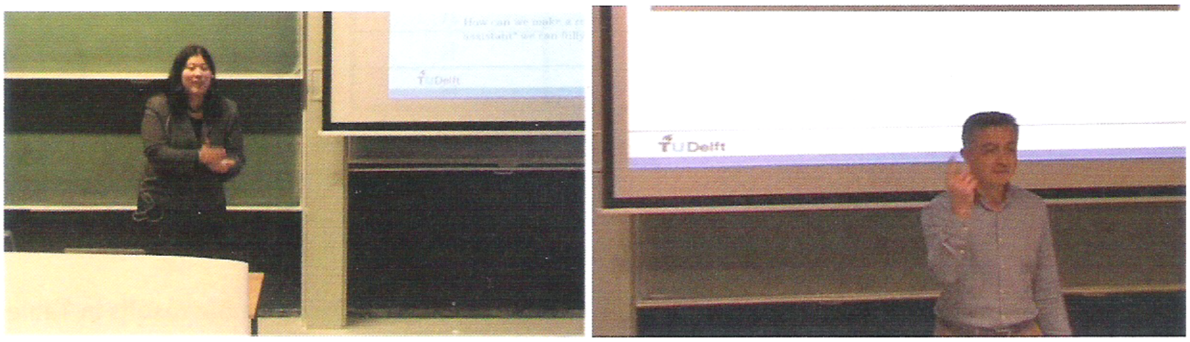


**Figure 2.** Sample video screenshots for the recorded CS4065 2016-2017 lectures.

As you know, two lecturers are involved with the course. Unfortunately, from the metadata in Figure 1, one cannot clearly tell which of the two lecturers is actually represented in a given lecture video.

Imagine you will build a content-based video search engine for the CS4065 2016-2017 lecture videos, in which it should be possible to retrieve (entire) lecture videos based on a query for a dedicated lecturer (aka, '`Cynthia Liem`' or '`Alan Hanjalic`').

**(a) (3 pt)** Make a clear sketch of a system diagram for the content-based lecture video search engine as described above. Describe what components are in it, and how these components interact.

**(b) (2 pt)**: The videos encompass multimodal data. If you had to pick just one modality based on which you would distinguish between the two lecturers of this course, which one would it be? Explain your answer.

**(c) (2 pt)**: Considering your answer to Question 1 (b), describe what feature representation would be appropriate for the given modality, in order to distinguish between the two lecturers.

**(d) (2 pt)**: Considering your answer to Question 1 (c), describe how you would determine the (dis)similarity between two given lecture videos. That is, give a rough outline of how the feature vector(s) for the videos will be computed, indicate what (dis)similarity measure you would apply, and what part(s) of the feature vector(s) will be used for the (dis)similarity computation.

## 2. Search (3 sub-questions, 6 points in total)

A query $q_1$ is submitted to two search engines $S_1$ and $S_2$, which both operate on the same dataset that contains 6 relevant items. Table 1 lists the top-10 results for each of the three systems of interest. A relevant item is indicated as R, while a non-relevant item is indicated as N.

| rank | $S_1$ | $S_2$ |
|---|---|---|
| 1 | R | N |
| 2 | R | N |
| 3 | N | N |
| 4 | N | R |
| 5 | N | R |
| 6 | N | R |
| 7 | N | N |
| 8 | R | R |
| 9 | N | R |
| 10 | R | N |

**Table 1.** Top-10 results for search engines $S_1$ and $S_2$ in response to a query $q_1$.

**(a) (2 pt)** Compute the Precision (P) and Recall (R) for both results lists.

**(b) (2 pt)** We now compare the performance of the two search engines based on the results in Table 1. We use the Average Precision (AP) for this purpose, defined as:

$$AP(q) = \frac{1}{|I|}\sum_{r=1}^{N}\{Precision(r) \cdot Relevance(r)\}$$

with $|I|$ being the total amount of relevant items in the dataset for query $q$ and with $N$=10.

Compute $AP(q_1)$ for search engines $S_1$ and $S_2$. Rank the search engines based on your outcomes.

**(c) (2 pt)** The AP values for the two search engines are not the same. Explain what information caused the difference between them.

## 3. Recommendation (4 sub-questions, 9 points in total)

We consider a service in which food items are rated on a scale from 1 (lowest) to 5 (highest). In Table 2, we see a user-item matrix indicating the ratings by 5 users (rows) of 5 items (columns).

| | Apple | Broccoli | Cauliflower | Date | Eggplant |
|---|---|---|---|---|---|
| Alice | 3 | 2 | 3 | ??? | 1 |
| Bob | 5 | 1 | 1 | 5 | 1 |
| Chris | 1 | 2 | 1 | 1 | 1 |
| Debby | 5 | 2 | 5 | 4 | 2 |
| Erin | 3 | 1 | 3 | 1 | 3 |

**Table 2.** User-item matrix for five users and 5 items

It is not known yet to what extent Alice likes Date. For this, neighborhood-based prediction will be applied. In assessing the distance between users, the Pearson correlation will be employed. The Pearson correlation $r$ between user $a$ and $b$ is given as

$$r(a,b) = \frac{\sum_{p \in P}(s_{a,p} - \bar{s_a})(s_{b,p} - \bar{s_b})}{\sqrt{\sum_{p \in P}(s_{a,p} - \bar{s_a})^2}\sqrt{\sum_{p \in P}(s_{b,p} - \bar{s_b})^2}}$$

with $s_{x,p}$ being the rating score given by user $x$ to item $p$.

Based on the closest neighbors $b$ within a relevant user neighborhood $N$, a prediction of a user $a$'s rating of item $p$ is obtained as follows:

$$pred(a,p) = \bar{r_a} + \frac{\sum_{b \in N} r(a,b) \cdot (r_{b,p} - \bar{r_b})}{\sum_{b \in N} r(a,b)}$$

**(a)** **(3 pt)** Calculate the predicted rating $pred(Alice, Date)$, when considering neighborhood $N$ to contain the **two** closest neighbors.

In recommender systems, we generally distinguish **content-based** and **collaborative** recommendation.

**(b)** **(2 pt)** Explain the difference between content-based and collaborative recommendation, and which technique would work best in which situation.

*Please turn over for further questions.*

Imagine that a large grocery chain (e.g. Albert Heijn or Jumbo) wants to build a recommender system for its customers, that will give in-store recommendations of what products to buy.

**(c)** **(2 pt)** For the given grocery chain scenario, give an example of a situation in which *relevance* and *utility* would be different. Explain this difference.

Student Sam builds a recommender system based on collaborative filtering. As initial setup data, Sam considers interaction data from users in the grocery chain's online web store: if an online user clicked on an item in that web store, Sam's system considers this to be a product transaction for the user. Then, as users will go to a physical store, the items in their shopping basket will be counted as their consumption profile, and collaborative filtering suggestions will be made.

**(d)** **(2 pt)** Critically discuss Sam's approach to the given recommender problem. If you would be asked to build the recommender system, to what extent would you take the same approach? Mention **at least one strong and one weak point** about Sam's proposed approach, including a discussion on **why** you find these points strong/weak.

# 4. Compression and compact data representation (4 sub-questions, 8 points in total)

In perceptual audio coding techniques and their related compression schemes (e.g. MP3, AAC), a phenomenon known as **masking** plays an important role.

**(a) (2 pt)** Explain what masking is, and why it is relevant for audio compression.

A general outline of the JPEG compression scheme is illustrated in Figure 3.

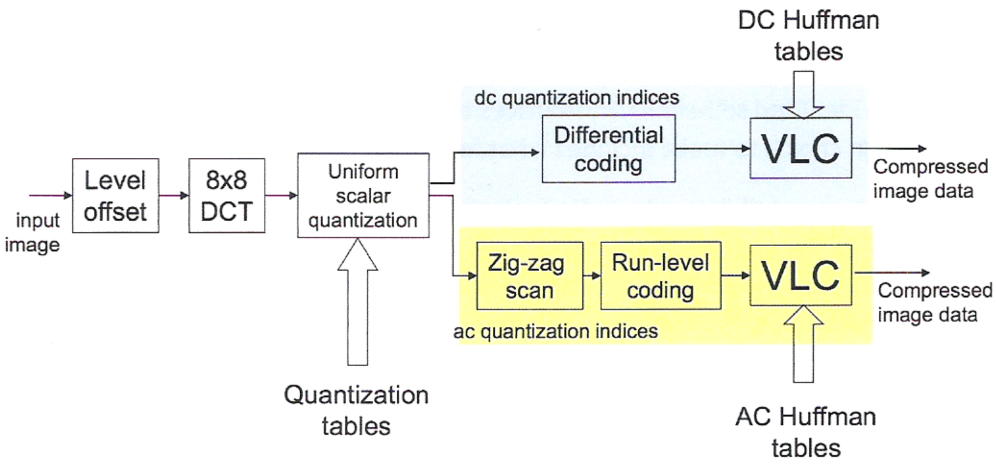


**Figure 3.** JPEG compression scheme

This scheme distinguishes between 'DC' and 'AC' quantization indices. The color-marked part of the scheme indicates how DC quantization indices will further be compressed.

**(b) (2 pt)** Explain **to what information in an image** the DC quantization indices refer, and why it is a good choice to apply **Differential Coding** to this data.

We consider four sentences $S_A$, $S_B$, $S_C$ and $S_D$:

$S_A$: "It is easy to recognize speech"

$S_B$: "It is easy to wreck a nice beach"

$S_C$: "It is nice to visit a beach"

$S_D$: "It is nice if this question is easy"

**(c) (2 pt)** Express these four sentences as sets of 2-shingles, taking words as tokens.

**(d) (2 pt)** Encode your answer to question (c) in the form of a **characteristic matrix**, that can e.g. be used for Min-Hashing purposes.

## 5. Similarity and matching (4 sub-questions, 8 points in total)

We want to perform musical structure analysis by considering the information in a self-similarity matrix.

**(a) (2 pt)** Sketch an idealized self-similarity matrix for a song with the following structure:

- Section A1: 20 seconds;
- Section B1: 20 seconds;
- Section A2: 15 seconds, same content as Section A1, but gradually speeding up;
- Section C: 5 seconds;
- Section A3: 15 seconds, exact repeat of Section A2;
- Section D: 10 seconds.

When moving from idealized self-similarity matrices to the actual computation of real-world similarity matrices, one design choice to make includes what features to use.

**(b) (2 pt)** Mention **two possible music audio features** that could be used for structure-related self-similarity analysis. For each of them, indicate **what information** they reflect, and thus, **what characteristics** a song should have, for the feature to reflect useful information in a self-similarity analysis.

In audio fingerprinting according to the Shazam scheme, to determine what song in a database may have been queried, an analysis of time consistent matches is performed. In Figure 4, scatterplots of matching hash location of a query against two database sound files are illustrated.

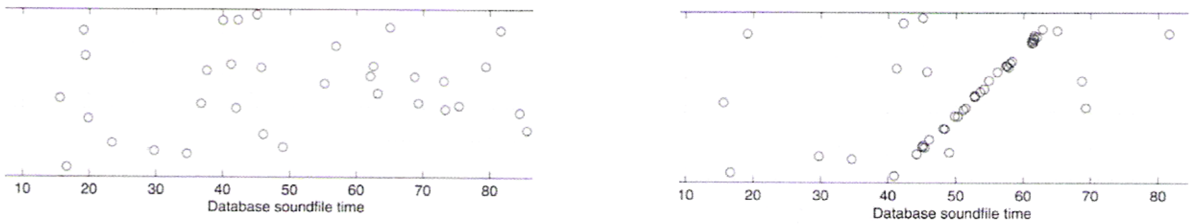


**Figure 4.** Scatterplots of matching hash locations of a query against two database sound files.

**(c) (2 pt)** Explain how the matching information as shown in Figure 4 is used, in order to perform a fast identification of which song was queried.

Finally, another 'matching' scenario that was discussed in the lectures considered the automated ranking of suitable job candidates, based on videos in which these candidates present themselves.

**(d) (2 pt)** Discuss **at least two considerations** that an engineer should take into account when building an automated job candidate ranking system as described above, in order for the resulting system to be valid and acceptable.

*This is the end of the exam.*

*Before handing in your answers, please verify that your name and student number are indicated on each answer sheet.*