

MOSIG – MSIAM – 2021-2022 Information Access and Retrieval – GBX9MO23

Georges Quénot – Philippe Mulhem – Jean-Pierre Chevallet

1st February 2020 – 9h00-11h00 (9:00am-11:00am) – 2 hours

Course materials, the two papers related to the examination, personal notes, and calculators (without network capabilities) are allowed.

The examination consists in questions related to two scientific papers and/or to the contents of the course:

[1] B. Mansouri, R. Zanibbi, D. W. Oard, Learning to Rank for Mathematical Formula Retrieval, ACM SIGIR '21.

[2] Jiaxin Wu and Chong-Wah Ngo, Interpretable Embedding for Ad-Hoc Video Search, ACM Multimedia 2020.

In the following questions, we expect real explanations with details, and not only excerpts from the papers. You should spend about 5 minutes per question and we expect concise answers.

Please use separate sheets for answers to questions (one for each paper).

Questions related to paper 1.

Q1.1: Explain (by giving examples, if needed) why the proposal of the paper is related to the principles of Information Retrieval.

Q1.2: Explain why classical IR representations on terms are unable to represent accurately mathematical formulas.

Q1.3: Present briefly the representations used for the formulas, and their complementarity.

Q1.4: List the similarity features proposed in the paper, and explain shortly the advantages/drawbacks of these similarities.

Q1.5: The part 3.2 presents the *unification* on formulas. Explain why the authors this unification is useful for formula retrieval. Is it related this to a kind of synonymy on words?

Q1.6: Explain why using a Learning to Rank (LtR) approach is a good idea in this paper.

Q1.7: Explain shortly the objective of the different evaluation measures $P@5$, MAP and $nDCG@5$. Why $nDCG$ measures are well adapted to the test collection ARQMath (see part 2.2)?

Q1.8: Comment briefly the results of Table 6 compared to the state-of-the-art results of table 3.

Q1.9: What would you suggest as interesting future works (*different from the ones proposed in the conclusion*) for this paper?

Questions related to paper 2.

Q2.1: What is the TRECVID Ad hoc Video Search task?

Q2.2: What is the main contribution in the work described in this paper?

Q2.3: What are the two tasks referred to in the “dual task” network?

Q2.4: On which prior work is based the Visual-Textual Embedding part?

Q2.5: How many sub-networks are there in the Visual-Textual Embedding part and which are they, (a) for the text stream and (b) for the video stream?

Q2.6: How is the mappings in the common space implemented?

Q2.7: (a) What is the difference between the loss used for classification and the classical BCE loss (equations 8 and 9)? (b) Why did the authors make this modification?

Q2.8: (a) How are the matching scores computed for the “embedding” part and the “concept” part? (b) What is the main difference?

Q2.9: (a) Which data sets are used for the training of the system? (b) For the evaluation of the system?

Q2.10: (a) Which “backbones” are used for the visual part? (b) For the textual part? (c) What is a “backbone” in this context?

Q2.11: What is the difference between the “classical” concept-based approach and the one proposed in this paper?

Q2.12: How does the concept-based approach proposed by the author compare with the classical one?