SlideReader: Automated Slide Analysis from Lecture Footage (Master Thesis)

Slides are crucial documents for lecturers. Well-designed slides efficiently summarize complex information, making them valuable study materials. Additionally, when integrated with lecture footage, static elements offer deeper insights, such as summarizing data tables or interpreting charts. Well-produced lecture footage offers an excellent resource to generate a dynamic more interactive slide set (e.x. Assigning auditory explanation for each element in the slide). When slides are combined with the reading order and audio, it aids both sighted and visually impaired students in better understanding static slide materials. Numerous other applications can be derived, including but not limited to slide-QA, summarization, pitch preparation, and slide authoring.

In this research, we aim to develop the SlideReader, a system that can analyze lecture recordings to retrieve the element-wise slide reading order and audio explanation employing state-of-the-art models.

1. Dataset Preparation:
    a. Collect high-quality lecture videos along with their corresponding digital slides.
    b. Utilize heuristics for data curation and annotation
2. Model Training and Experimentation:
    a. Train diverse neural networks, with a focus on transformers and LLMs.
3. Performance Evaluation:
    a. Conduct a quantitative & qualitative evaluation of the trained models.

Throughout your research, you'll benefit from experts' guidance and support. You will have access to a powerful computing cluster to facilitate your experiments. **Significant findings will be submitted as a research paper at a prestigious conference.**

Requirements:

1. Demonstrated interest in the topic. **Related work [1-5].**
2. Experience with deep learning models, such as transformers and LLMs.
3. Familiarity with the Linux operating system and comfort with terminal commands.

If interested, please send your application, which should include your CV and transcript of records (optional: motivation letter) to:

Omar Moured, omar.moured@kit.edu

References:

1. https://ieeexplore.ieee.org/document/7333955
2. https://dl.acm.org/doi/pdf/10.1145/3544548.3580921
3. https://github.com/BradyFU/Awesome-Multimodal-Large-Language-Models
4. https://cvhci.anthropomatik.kit.edu/~mhaurile/papers/2019_ICDAR_WiSe.pdf
5. https://cvhci.anthropomatik.kit.edu/~mhaurile/papers/2019_WACV.pdf