

# Ekstraksi Key Moments Otomatis pada Video Perkuliahan di Fasilkom UI Menggunakan Optical Character Recognition dan Topic Modelling = Automatic Key Moments Extraction on Lecture Videos in Fasilkom UI with Optical Character Recognition and Topic Modelling

Parluhutan, Matthew Tumbur, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920534703&lokasi=lokal>

---

## Abstrak

Pandemi COVID-19 mengubah pola kehidupan manusia, termasuk sistem perkuliahan yang berubah ke metode daring. Video perkuliahan dengan salindia menjadi salah satu pilihan sarana penyampaian materi kuliah secara daring. Penelitian ini bermaksud menguji keabsahan rancangan sistem yang mampu melakukan segmentasi temporal sesuai topik secara otomatis pada video perkuliahan. Sistem yang diajukan dibagi menjadi tiga sub-sistem yang memanfaatkan teknologi keyframe extraction, optical character recognition (OCR), dan topic modelling. Pertama, video perkuliahan akan diubah menjadi kumpulan keyframe dengan memanfaatkan metode Slide Detector yang dimodifikasi. Selanjutnya, akan dilakukan ekstraksi teks dari frame-frame tersebut menggunakan Tesseract OCR dengan preprocessing tambahan. Akhirnya, BERTopic dengan beragam algoritma clustering dan LDA diuji kemampuannya dalam topic modelling yang berguna untuk mengambil topik yang koheren dari teks tersebut. Penelitian pada tahap keyframe extraction menunjukkan bahwa terdapat peningkatan recall sebesar 0,235-0,25 dari 0 dan precision sebesar 0,619-0,75 dari 0 pada beberapa video pada Slide Detector termodifikasi. Sebaliknya, penelitian pada tahap OCR menunjukkan bahwa tambahan preprocessing belum bisa membantu meningkatkan performa Tesseract OCR. Pada tahap terakhir, ditemukan bahwa BERTopic lebih unggul daripada LDA dalam menarik topik yang koheren untuk use case penelitian ini. Agglomerative dan KMeans clustering ditemukan lebih optimal untuk kasus video perkuliahan jika dibandingkan dengan metode density-based. Augmentasi data dengan takaran yang sesuai diperlukan untuk mendapatkan hasil sedemikian rupa pada tahap ini. Secara umum, sistem dengan tiga bagian yang diusulkan pada penelitian ini sudah mampu melakukan segmentasi video perkuliahan sesuai tujuan, namun, video perkuliahan bersalindia merupakan dataset yang sangat heterogen dan merancang sebuah sistem yang mampu memanfaatkan dataset tersebut adalah tantangan tersendiri.

.....The COVID-19 pandemic changed the lifestyle of many people, including university lectures that moved to online delivery. Lecture videos with slides became an option to deliver lecture materials online. This work attempts to show a proof of concept for a system design that is able to automatically segment a lecture video temporally based on the topic. The proposed system is divided into three subsystems that make use of keyframe extraction, optical character recognition (OCR), and topic modelling techniques. First, a lecture video will be converted to a collection of keyframes using a modified Slide Detector technique. Next, those frames will be processed using Tesseract OCR with some additional preprocessing steps to extract text. Lastly, BERTopic with various clustering techniques and LDA will be used for topic modelling to obtain a coherent topic from the text extracted earlier. The research in the keyframe extraction step shows that there is an increase of 0.235-0,5 points from 0 for recall and 0,619-0,75 points from 0 for precision for certain videos using the modified Slide Detector. On the other hand, the research in the OCR step shows that the additional preprocessing is not yet able to help increase the performance of Tesseract OCR. At the last step,

BERTopic proves to be better than LDA to obtain the coherent topic for this system's use case.

Agglomerative and KMeans clustering is better for lecture videos compared to density-based methods.

Appropriate amounts of data augmentation is needed to obtain the best results at this step. Overall, the three-part system in this research is able to segment lecture videos as intended, however, lecture videos with slides is a dataset that is very heterogeneous and designing a system to handle all types of videos is a large challenge.