Plan Overview

A Data Management Plan created using DMPTool

DMP ID: https://doi.org/10.48321/D1BF4BE9E7

Title: Topic Tagging in Educational Videos Using Text-Based Search Techniques

Creator: Soongho Han - **ORCID:** 0009-0007-5443-9173

Affiliation: Iie Varsity College

Funder: The Independent Institute of Education (iie.ac.za)

Template: Digital Curation Centre

Project abstract:

This research project delves into the technical intricacies of developing robust algorithms for automatic speech-to-text transcription in educational videos. The primary objective is to employ advanced signal processing techniques and deep learning models to accurately convert spoken language in videos into textual format. These transcriptions will serve as the foundation for implementing sophisticated topic tagging and indexing mechanisms within the educational video content. By leveraging cutting-edge advancements in speech recognition and natural language processing, the project aims to revolutionize the accessibility and navigability of educational video resources through precise text-based search functionalities.

Start date: 04-02-2024

End date: 11-04-2024

Last modified: 04-12-2024

Copyright information:

Topic Tagging in Educational Videos Using Text-Based Search Techniques

I want to ask people these two questions.

1.   How can natural language processing algorithms be effectively applied to extract key themes and concepts from educational video transcripts for topic tagging purposes?

2.   What is the most suitable text-based search techniques for accurately categorizing topics within educational videos, considering factors such as efficiency, accuracy, and scalability?

1.   Primary Data Collection

-        Literature Review: Primary data will be collected through a comprehensive review of relevant academic literature, including seminal sources and recent scholarly publications. The literature review will provide theoretical frameworks, empirical findings, and best practices related to topic tagging in educational videos, text-based search techniques, and related topics.

-        Metadata Analysis: In addition to scholarly literature, primary data may involve the analysis of metadata from existing educational videos available on online platforms (such as Youtube). Metadata such as titles, descriptions, tags, and user engagement metrics can provide insights into current practices and trends in topic tagging.

2.   Secondary Data Collection

Sampling: Convenience sampling will be used to recruit participants for the online surveys. Recruitment efforts will target educational communities, online learning platforms, social media groups, and professional networks to reach a diverse pool of participants.

- Detailed Dataset Descriptions: Describe the structure and contents of your datasets, including the types of data collected (quantitative or qualitative), the nature of the variables, and any specific data formats used. This should also include any file naming conventions or identifiers used for the dataset components.
- Methodology Documentation: Document the methodology used in data collection, analysis, and processing. This includes the specific techniques and tools used for data scraping, natural language processing, and any statistical or thematic analysis methods. It's crucial to detail the algorithms or models developed, including versioning information of any software or tools used.
- Ethical Considerations: Document any ethical considerations and approvals, including informed consent forms and how participant confidentiality is maintained. This information is crucial for ethical reproducibility.
- Citations and References: Include complete references for any sources cited in your dataset documentation. This allows future researchers to trace back to your foundational sources.

I need to ensure that all participants provide informed consent before participating in the research. This includes explaining the purpose of the study, the procedures involved, potential risks, benefits, and the rights to withdraw at

any time without penalty.

I am going reference and cite for any material and data not created by me. So others can see where the data come from. And if I need to use private data, I can ask people who created it for the permission.

It can be stored on computer and external hard drive and also cloud storage.

so If my computer and hard drive are damaged, I can still get data from the cloud storage.

I can put the password on the project document, so then no one can open it without the password.

Most important information will be the solution about the research topic. So I want to keep for 1 year.

I am going to use my computer, external HDD and google drive.

Google drive provides allocated free disk storage, so there is no payment and I can preserve my document as long as it is saved on Google drive.

I can send my document through email and whatsapp if someone wants and I can also post it on such as blog.

And I can also share the link from the Google drive, so then people can download. And if my close people want my project, I can give it with USB.

There is no restriction for my project. But when someone use my project, it must be rephrased.

I am the only one responsible.

I need just computer.

Planned Research Outputs

Text - "Topic Tagging in Educational Videos Using Text-Based Search Techniques"

This research project delves into the technical intricacies of developing robust algorithms for automatic speech-to-text transcription in educational videos. The primary objective is to employ advanced signal processing techniques and deep learning models to accurately convert spoken language in videos into textual format. These transcriptions will serve as the foundation for implementing sophisticated topic tagging and indexing mechanisms within the educational video content. By leveraging cutting-edge advancements in speech recognition and natural language processing, the project aims to revolutionize the accessibility and navigability of educational video resources through precise text-based search functionalities.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Topic Tagging in Educational Videos Using Text-Bas ...	Text	2024-12-03	Open	None specified		Creative Commons Attribution 4.0 International	None specified	No	No