AutoTagLearn: An NLP-Based System for Intelligent Keyword Extraction and Conceptual Flowcharting in Academic Content

By:
Niharika Ravi Kumar
Roll no: 31

Guide:
Mrs. Leena Chakraborty

PROBLEM STATEMENT

1. Students struggle with unorganized, untagged educational content.
2. Manual tagging is time-consuming and often inconsistent.
3. Lack of intelligent resource discovery through semantic search.
4. Need: An automated, AI-driven platform to manage and recommend academic resources based on context.

LITERATURE REVIEW

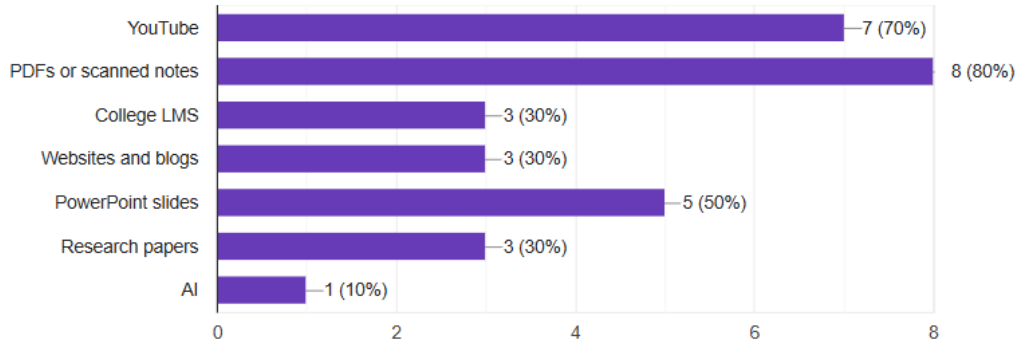
1. Topic Modeling for Educational Documents using BERTopic
Publisher: Elsevier, 2021
Key Idea: Applied unsupervised BERTopic to identify broad academic themes.
Limitation: Topics lacked granularity and learner-specific context.
What We Improve: AutoTagLearn grounds tagging in **actual course/subject modules**, offering **fine-grained, context-driven tags**.

2. Enhancing Tagging of Learning Content using Semantic Similarity Measures
Publisher: IEEE, 2021
Key Idea: Used TF-IDF and cosine similarity for static tag matching.
Limitation: Struggled with contextual understanding.
What We Improve: AutoTagLearn leverages **contextualized embeddings (BERT)** to offer **richer, more accurate tag suggestions**.

SURVEY REVIEW

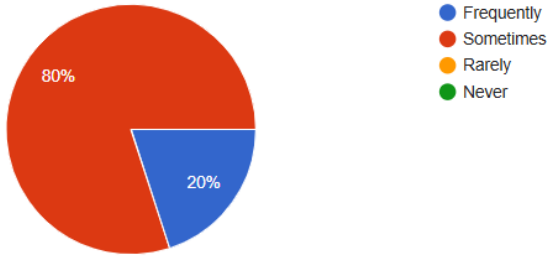
What type of resources do you use most frequently for studying?

10 responses



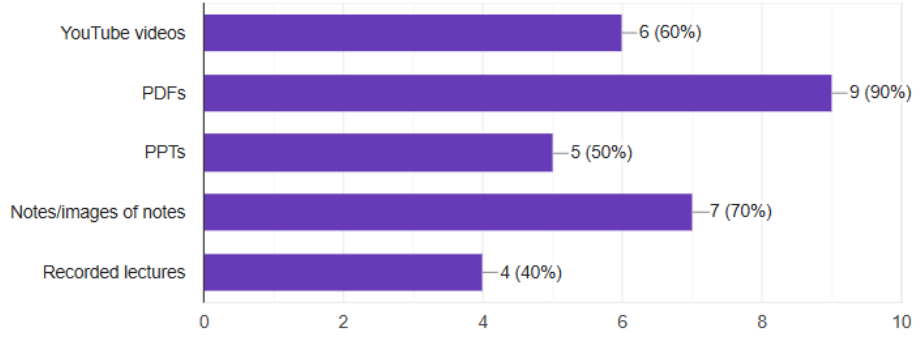
Do you face difficulty in finding topic-specific content within long videos or documents?

10 responses



Which of the following formats would you like to see auto-tagged for easier navigation? (Select all that apply)

10 responses

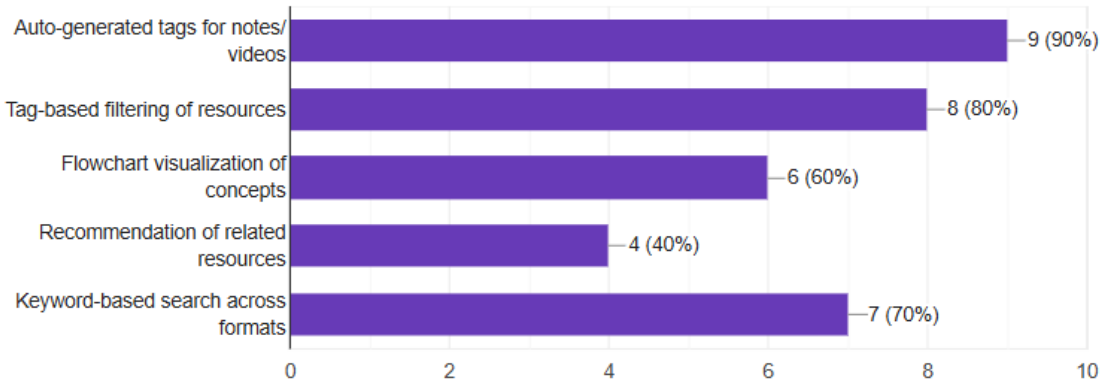


SURVEY REVIEW

Copy chart

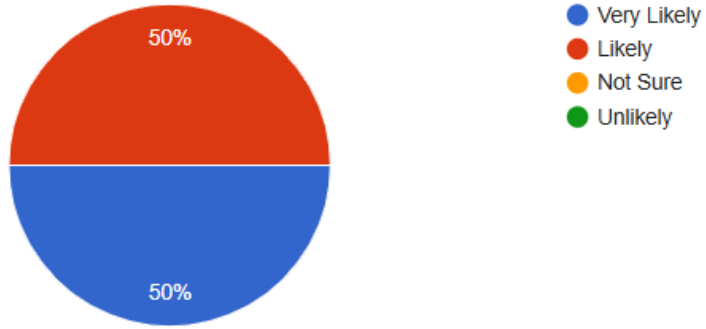
What features would you expect from a tool like AutoTagLearn? (Select all that apply)

10 responses



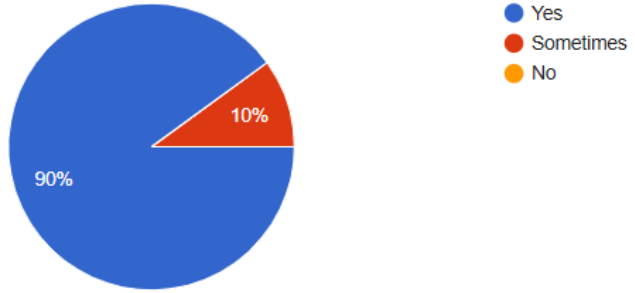
How likely are you to use such a tool if it's freely available?

10 responses



Do you prefer visual tools like flowcharts for understanding inter-topic connections?

10 responses



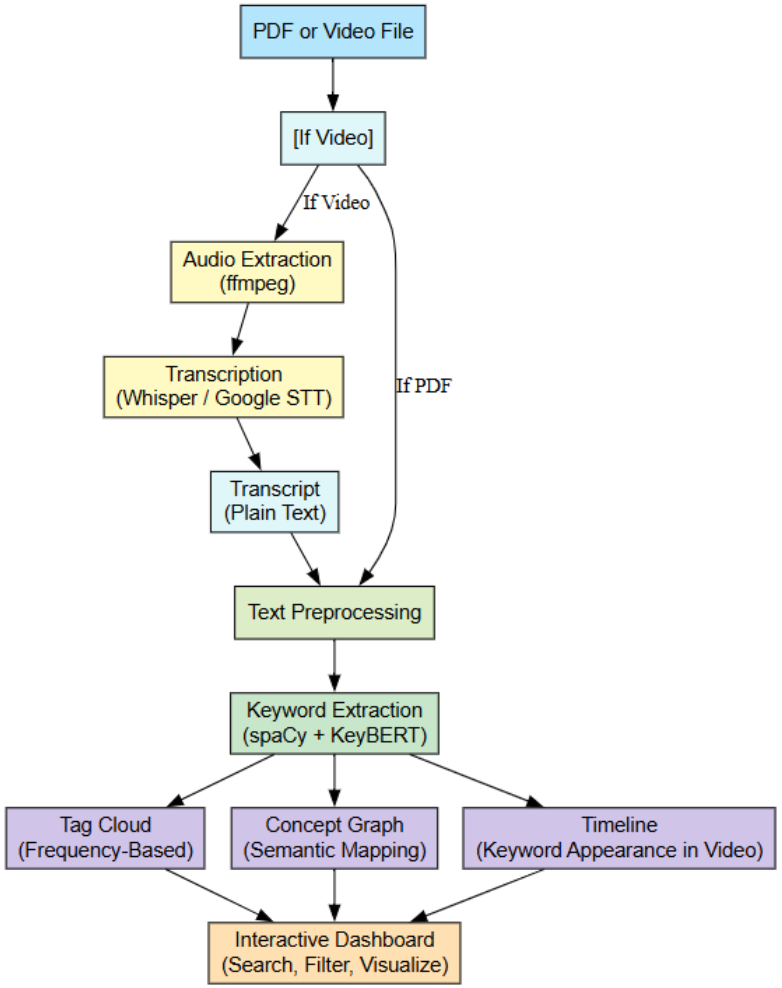
Objective of the Project

1. To build a web-based application that:
   - Automatically tags uploaded educational content using NLP.
   - Allows users to upload, manage, and search study materials.
   - Recommends relevant content based on user behavior and tags.
2. Improve accessibility and discoverability of learning materials.

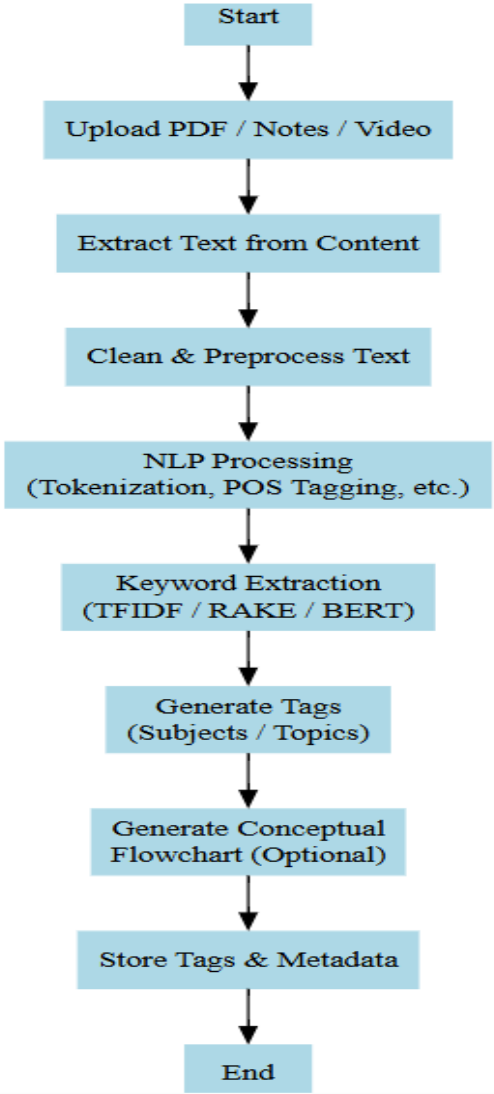
Tools and Technologies Used

Component	Tool/Technology	Justification / Purpose
Frontend	HTML, CSS, JS (Bootstrap)	Lightweight, fast development, responsive UI
Backend	Flask	Lightweight Python framework, integrates well with ML modules
NLP/ML	spaCy / Transformers	For tokenization, entity recognition, auto-tag generation
Database	Firebase / SQLite	Realtime DB for quick deployment; easy to manage user data
Storage	Firebase Storage / Cloudinary	For storing uploaded files efficiently
Visualisation	Graphviz / Mermaid.js	For UML & system architecture diagrams
Hosting	GitHub Pages + Render	Free hosting for frontend and backend

System Architecture Diagram



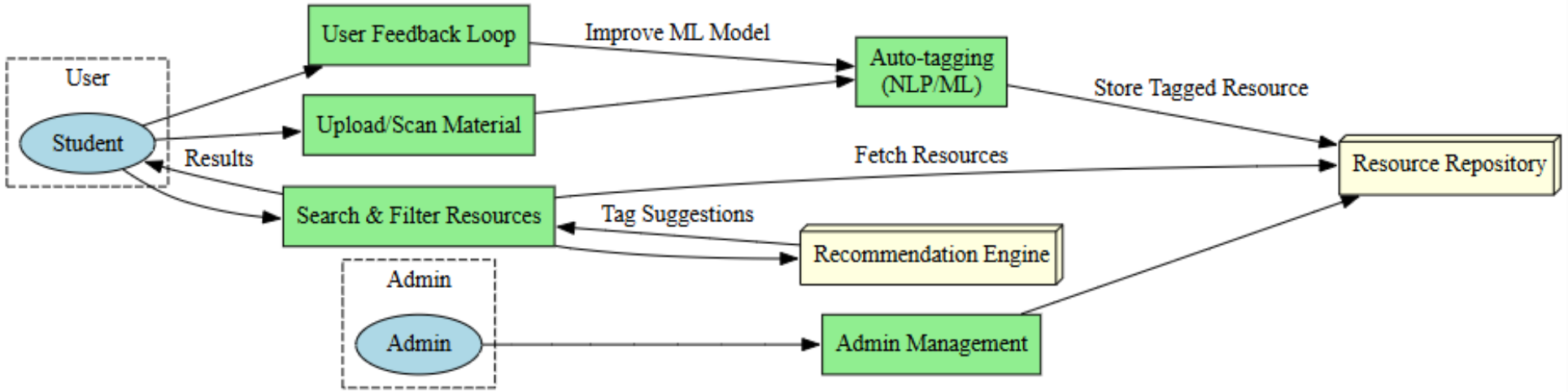
NLP & AutoTagging Workflow



Module Division (Phase-wise Breakdown)

Phase	Modules Covered	Description
Phase 1	User Management	Login, signup, user session, roles
Phase 2	Resource Upload & Management	Upload PDFs, notes, videos, detect type
Phase 3	Auto-Tagging (ML/NLP Integration)	Process content, assign subject/topic tags
Phase 4	Search & Filter	Tag-based and relevance-based content search
Phase 5	Recommendation Engine	Suggest content based on behaviour/tags
Phase 6	Feedback & Evaluation	Ratings, comments, and model fine-tuning

Data Flow Diagram (DFD Level 1)



Conclusion

- AutoTagLearn aims to streamline educational content discovery using AI.
- Addresses accessibility and organization challenges for learners.

- Future Scope:

Multi-language support
Question-answer generation from notes
Voice/audio tagging

THANK YOU