IntelliNote – Artificial Intelligence enhanced Online Lecture Note-Taking including Flowchart and Diagram Detection

Dr. Devidas Thosar
Department of Artificial Intelligence and Machine Learning
G.H.Raisoni College of Engineering and Management,
Pune, India
Email: devidas.thosar@raisoni.net

Priyanshi Patil
Department of Artificial Intelligence and Machine Learning
G.H.Raisoni College of Engineering
and Management, Pune, India
Email: priyanshi.g.patil@gmail.com

Rashi Bajpai
Department of Artificial Intelligence and Machine Learning
G.H.Raisoni College of Engineering
and Management, Pune, India
Email: rashibajpai2004@gmail.com

Renuka Mune
Department of Artificial Intelligence and Machine Learning
G.H.Raisoni College of Engineering
and Management, Pune, India
Email: renukamune45@gmail.com

***Abstract*— Now-a-days with the advancement of technology, education has become more accessible to everyone. Learners can find many types of videos related to any topic they want to learn or know about. Videos can be lengthy sometimes when we are time constrained. IntelliNote is designed to generate transcription of any provided valid YouTube link and then generates the summary out of the Transcript. Along with transcription of video, IntelliNote also detects objects such as Flowcharts and Diagrams in the provided video link and saves them locally. This research paper contains the purpose, its proposed solution, methodology, working and results of the generated model.**

Keywords – Yolo Model, Fast API, Assembly AI, Object Detection, Transcription, Summarization, Automatic Speech Recognition, WebSockets

I. INTRODUCTION

E-learning systems are gaining increased popularity due to their massive scalability and their immense potential of providing non-disrupted and affordable learning 24/7. Modern learning and teaching embrace the usage of Information and Communication Technology (ICT) and other tools in education. The existence of the technological aspects helps in improving the interest of learners, developing creativity among learner, and encourages the critical thinking of the learners.

In recent years, artificial intelligence (AI) has changed many areas, including education. AI-powered tools are making a big difference in how students learn and interact with educational content. One such tool is the AI-powered note maker, which helps make the process of taking notes during online lectures much easier. This tool will help learners in making/taking notes of the online lectures that they are attending more effectively. This helps learners to take notes in a time efficient manner. It uses smart technologies like Natural Language Processing (NLP) and Machine Learning (ML) to turn the spoken words from lectures into well-organized, easy-to-read notes. It also comprises the important diagrams and flowcharts in the online lectures. Thus, learners will be able to overview the diagrams and flowcharts in a systematic manner.

The learners can learn in-depth theoretical, practical, and research-specific concepts of their subjects with the advent of this technology. An excellent example of this is Revisely, where learners can save hours and enhance their productivity in note taking and document editing. It is like ChatGPT except Revisely generates content from images as well. Although Revisely is a great AI tool it cannot make or generate structured notes from online classes.

Let us understand how IntelliNote can be helpful for learners. Suppose you have an exam today and as time is ticking, you do not have the time to go through the topics. And you also do not want to watch videos. You have already tried many other Transcription extensions or websites that provide you the Transcript of the video link that you provide. But what if the video had some key diagrams or flowcharts that must be remembered? The available Transcription tools do not provide diagrams/ flowcharts or object detection in general from the video link.

IntelliNote is the solution to this problem. With IntelliNote, you can not only get your YouTube Video Transcribed and Summarized in your preferred Language, but can also go through the important images from the videos. The Transcript and Images get saved automatically after they are transcribed and they can also be edited according to the user's choice. Also, the users can name the Transcription themselves so that the Transcriptions can be distinct and differentiable. The Transcription is performed with the help of Assembly AI API and Yolov8n model is used for image detection. We are also using WebSocket Connection to let the user know the live status of his request being processed, send messages and completion notifications to the user.

II. LITERATURE REVIEW

Note taking can be time consuming and includes use of paper and ink. Sometimes we might lose the notes that we created or we might not be able to find them in the nick of time. Traditional learning methodologies in education have transformed through the integration of Artificial Intelligence. This literature review synthesizes current research on AI's impact on traditional ways of making notes, shortcomings, challenges, and outcomes.

Table 2.1. shows the summary of the research papers that we reviewed.

TABLE 2.1: LITRATURE REVIEW

Sr No.	Paper Title	Author	Publisher and Published Year	Limitations
1.	The Developing of the System for Automatic Audio to Text Conversion	Oleh Basystiuk, Natalya Shakhovska	Published Year: March 5–6, 2021, Publisher: IT&AS'2021: Symposium on Information Technologies & Applied Sciences, Bratislava, Slovak Republic	The algorithm uses prefix and hash functions, which can be tricky to implement and fine-tune for different situations
2.	Artificial intelligence inspired multilanguage framework for note-taking and qualitative content-based analysis of lectures	Munish Saini, Vaibhav Arora	Published Online: 18 July 2022 Publisher: Springer Science+Business Media, LLC, part of Springer Nature 2022	Factors like not being able to hear the professor clearly, distractions in the classroom, and problems with the learning environment can make it hard to take effective notes.
3.	Artificial intelligence, text generation tools and ChatGPT	Thomas Lancaster	Published Year: 2023 Publisher: International Journal for Educational Integrity	Using AI to generate text might not always produce accurate or high-quality results, which could result in submitting information that is misleading or incorrect.

A. *Drawbacks of Existing System*

After reviewing these research papers, the drawbacks of these existing systems are – They do not provide notes that contain flowcharts or diagrams that are included in the online lecture videos. The notes also contain auto generated words aside from actual words from the lecture. This can sometimes be misleading and the sentence formulation may change if incorrect words are auto-generated.

III. METHODOLOGY

IntelliNote uses FastAPI to create a web-based service that processes YouTube videos by first downloading them, extracting the audio and converting it in the form which can be then transcribed, summarized and by using YOLO model, object detection can be done and the results can be communicated with the client in real-time using WebSockets.

Along with standard libraries like os, re, subprocess, pathlib.Path, asyncio and logging, the other major libraries like assemblyai is used for transcription services, OpenCV is used for video and video processing, youtube_dl is used to download videos from YouTube and other platforms, FastAPI for building APIs with Python, YOLO model for real-time object detection and typing.Dict for providing type hinting for dictionaries are also used.

First, we create an instance of FastAPI application, which will handle incoming requests and route them to appropriate endpoints. The FastAPI runs on localhosthost:8000. When the user first visits the website, he can see 4 options i.e. Login, Sign Up, Transcribe and Features.

A. *Client Side:*

The new user is asked to Sign In, in which the user will enter details like Username, Email ID and Password. These details are then stored in the Database for future Login purposes. If the user is a returning user, then for Logging In, he will have to enter his Email ID and Password. Once the user enters these details, the Email Id and Password are verified with the Database, if the Email and Password match, then user has logged in successfully and if they do not match, error message is generated saying that the Email and Password do not match.

After the Sign-In and Login process, the user can then move on to transcribing the video.

For Video Transcription, the user enters the YouTube video link in the "Enter YouTube Link" box, then the user enters the name of the transcript that he wants to save as followed by entering the Summary Length and choosing the preferred language.

B. *Server Side:*

The end-point expects form data containing the YouTube video URL, desired transcript name, summary length, language, and a unique request_id.

'app' is the FastAPI instance. Logging is configured to display information-level logs and above. The 'logger' will be used throughout the application to log messages. The server takes the YouTube link and downloads the video in '.webm' format using yt-dlp which downloads the videos from YouTube or any other sources. It constructs the download options to fetch the best available video and audio streams. Each step sends a status update via WebSocket.

After the download is completed, ffmpeg extracts the best available audio stream from the video, converting it first into MP3 file with a specified quality. AssemblyAI requires audio in WAV format, so the MP3 file is converted into WAV format. Ffmpeg converts the MP3 file to WAV format asynchronously. The WAV audio is transcribed into text in user preferred language using AssemblyAI's transcription service. The AssemblyAI API key is crucial for the transcription service. The API key is fetched from an environment variable for security. If it is not found, a default key is used for development purposes. And we have used the default key for testing and development process. The 'transcriber' object handlesaudio transcription tasks. The application needs specific directories to store transcripts and detected images. So we specify the paths for storing them. The transcribed text is summarized based on the user-selected length.

Yolov8n Model is used for Image Detection where every 200th frame of the video is processed and detected frames are saved as images in the deignated output folder. It returns a list of paths to the saved images. We have used yt_dlp to extract the title of the video without downloading it and invalid characters are removed and this title is used as a name for the detected images folder.

To handle real-time communication between the server and the clients, we defined a 'ConnectionManager' class. Websocket endpoint allows clients to establish a connection by providing a unique request_id. The ConnectionManager class maintains active WebSocket connections, allowing the server to send messages and completion notifications to specific clients based on a unique request_id. This setup ensures that multiple users can interact with the server simultaneously without interference.

The transcript, summary and detected images are displayed on the website once all the steps are completed and is sent to the client via WebSocket. The transcript and detected images are saved automatically after the transcription is completed for easy access to the users as well for future editing purposes.

If any step fails, appropriate error messages are sent back to the client, and the error is raised to be handled by FastAPI. Temporary fileslike the downloaded video and audio files are deleted to conserve storage. To run the FastAPI application, we have used uvicorn, to run the script directly and start the FastAPI server on all available network interfaces(0.0.0.0) at port 8000.

Flowchart 3.1. shows the working of the system in a systematic manner.

C. Mathematical Functions --

1. Non-Maximum Suppression (NMS) Algorithm:

Type of Algorithm: Post-processing in Object Detection

Purpose: Reduces multiple overlapping bounding boxes to a single detection.

Working Principle:

YOLOv8 uses NMS to eliminate redundant bounding boxes by keeping only the one with the highest confidence score and discarding the rest that overlap by a significant threshold (IoU - Intersection over Union).

Mathematical Function: IoU is a mathematical function that measures the overlap between two bounding boxes. It's computed as:

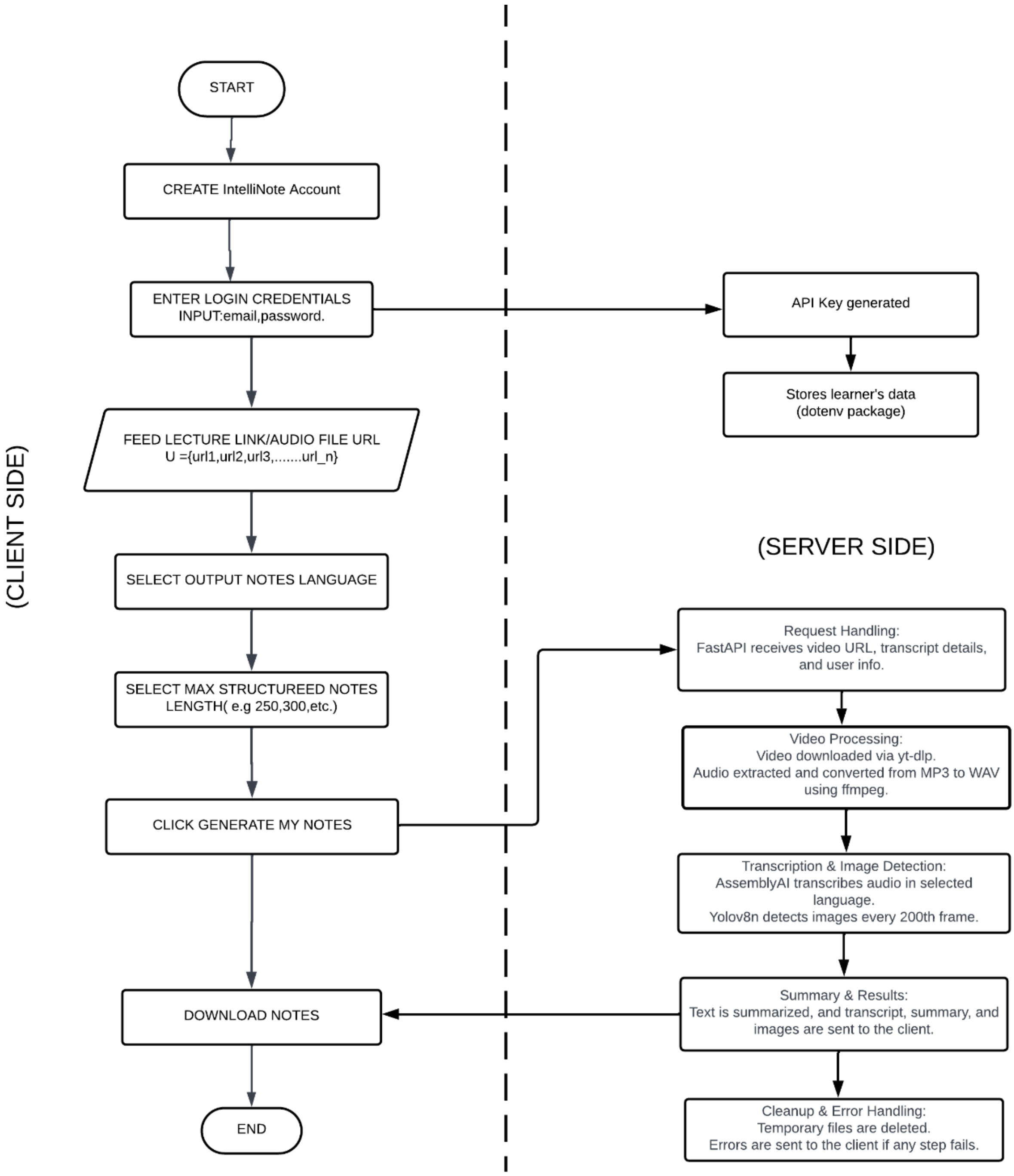
$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

2. Convolutional Neural Networks (CNNs):

YOLOv8 uses CNNs to extract features from input frames. The convolution operations help in detecting spatial hierarchies in images (edges, shapes, textures, etc.).

Mathematical functions used:

1. Convolution: A sliding window function that applies a filter (kernel) to compute feature maps from the input image.
2. Activation Functions: Non-linear functions like ReLU (Rectified Linear Unit), which replaces negative values with zero to introduce non-linearity, and Sigmoid for probability outputs.


START
CREATE IntelliNote Account
ENTER LOGIN CREDENTIALS INPUT:email,password.
FEED LECTURE LINK/AUDIO FILE URL U ={url1,url2,url3,.......url_n}
SELECT OUTPUT NOTES LANGUAGE
SELECT MAX STRUCTUREED NOTES LENGTH(e.g 250,300,etc.)
CLICK GENERATE MY NOTES
DOWNLOAD NOTES
END
(CLIENT SIDE)
API Key generated
Stores learner's data (dotenv package)
(SERVER SIDE)
Request Handling: FastAPI receives video URL, transcript details, and user info.
Video Processing: Video downloaded via yt-dlp. Audio extracted and converted from MP3 to WAV using ffmpeg.
Transcription & Image Detection: AssemblyAI transcribes audio in selected language. Yolov8n detects images every 200th frame.
Summary & Results: Text is summarized, and transcript, summary, and images are sent to the client.
Cleanup & Error Handling: Temporary files are deleted. Errors are sent to the client if any step fails.

FLOWCHART 3.1: SYSTEM ARCHITECTURE

3. Pooling: Used to downsample feature maps, reducing their dimensions while preserving important features.

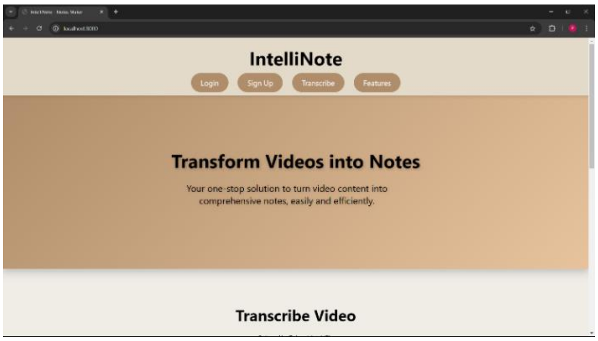
3. Object Detection Confidence Score Calculation:

Sigmoid function: This transforms the raw model outputs into probabilities, which represent how likely a particular object is present in a bounding box.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

IV. RESULTS

IntelliNote helps user to generate transcript, summary and detect images from the provided YouTube link. The below figures show the website interface, the working of the system and the results produced as where as how they get stored locally automatically.

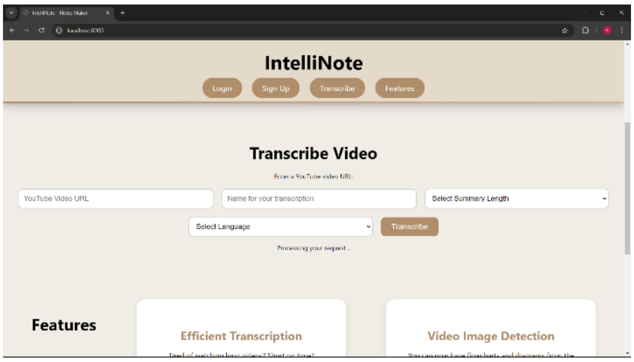


Transform Videos into Notes

Your one-stop solution to turn video content into comprehensive notes, easily and efficiently.

Transcribe Video

(FIGURE 4.1 WEBSITE INTERFACE)



Transcribe Video

Enter a YouTube video URL:

YouTube Video URL | Name for your transcription | Select Summary Length

Select Language

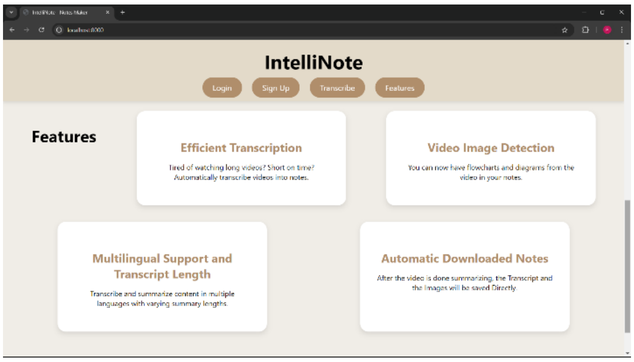
Processing your request...

Features

Efficient Transcription

Video Image Detection

(FIGURE 4.2: TRANSCRIPTION AREA INTERFACE)



Features

Efficient Transcription

Tired of watching long videos? Short on time? Automatically transcribe videos into notes.

Video Image Detection

You can now have flowcharts and diagrams from the video in your notes.

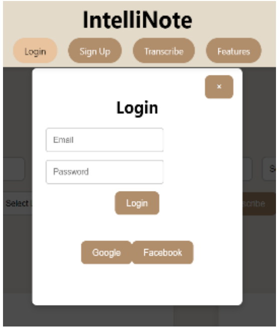
Multilingual Support and Transcript Length

Transcribe and summarize content in multiple languages with varying summary lengths.

Automatic Downloaded Notes

After the video is done summarizing, the Transcript and the Images will be saved Directly.

(I FIGURE 4.3: FEATURES OF INTELLINOTE)

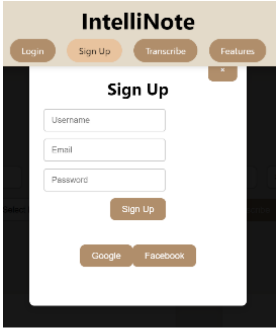


Login

Email

Password

(FIGURE 4.4: LOGIN PAGE)



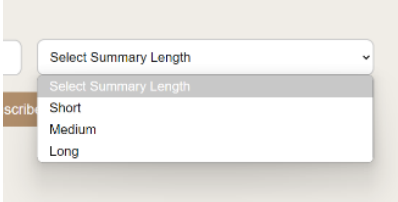
Sign Up

Username

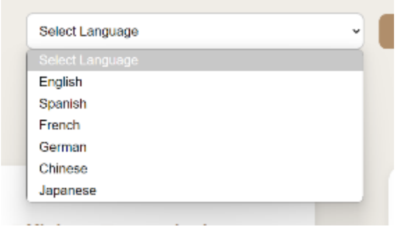
Email

Password

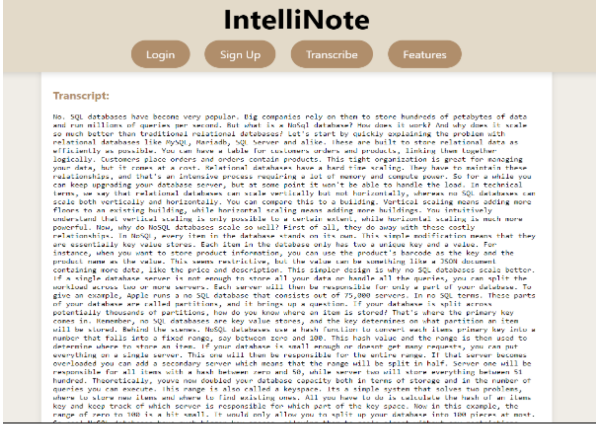
(FIGURE 4.5: SIGN UP PAGE)



(FIGURE 4.6: SUMMARY LENGTH)



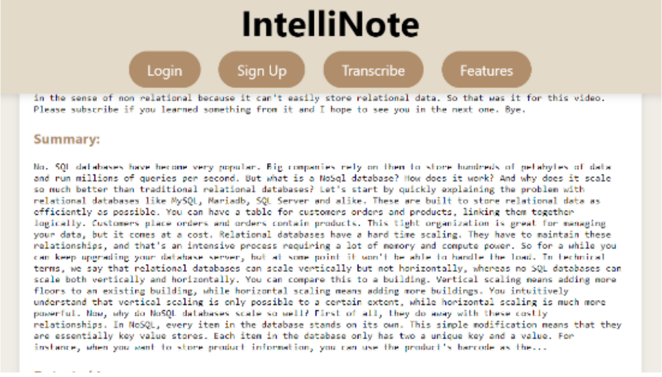
(FIGURE 4.7: LANGUAGE CHOICE)



Transcript:

No. SQL databases have become very popular. Big companies rely on them to store hundreds of petabytes of data and run millions of queries per second. But what is a NoSql database? How does it work? And why does it scale so much better than traditional relational databases? Let's start by quickly explaining the problem with relational databases like MySQL, Mariadb, SQL Server and alike. These are built to store relational data as efficiently as possible. You can have a table for customers orders and products, linking them together logically. Customers place orders and orders contain products. This tight organization is great for managing your data, but it comes at a cost. Relational databases have a hard time scaling. They have to maintain these relationships, and that's an intensive process requiring a lot of memory and compute power. So for a while you can keep upgrading your database server, but at some point it won't be able to handle the load. In technical terms, we say that relational databases can scale vertically but not horizontally, whereas no SQL databases can scale both vertically and horizontally. You can compare this to a building. Vertical scaling means adding more floors to an existing building, while horizontal scaling means adding more buildings. You intuitively understand that vertical scaling is only possible to a certain extent, while horizontal scaling is much more powerful. Now, why do NoSQL databases scale so well? First of all, they do away with these costly relationships. In NoSQL, every item in the database stands on its own. This simple modification means that they are essentially key value stores. Each item in the database only has two a unique key and a value. For instance, when you want to store product information, you can use the product's barcode as the key and the product name as the value. This seems restrictive, but the value can be something like a JSON document containing more data, like the price and description. This simpler design is why no SQL databases scale better. If a single database server is not enough to store all your data or handle all the queries, you can split the workload across two or more servers. Each server will then be responsible for only a part of your database. To give an example, Apple runs a no SQL database that consists out of 75,000 servers. In no SQL terms. These parts of your database are called partitions, and it brings up a question. If your database is split across potentially thousands of partitions, how do you know where an item is stored? That's where the primary key comes in. Remember, no SQL databases are key value stores, and the key determines on what partition an item will be stored. Behind the scenes. NoSQL databases use a hash function to convert each items primary key into a number that falls into a fixed range, say between zero and 100. This hash value and the range is then used to determine where to store an item. If your database is small enough or doesnt get many requests, you can put everything on a single server. This one will then be responsible for the entire range. If that server becomes overloaded you can add a secondary server which means that the range will be split in half. Server one will be responsible for all items with a hash between zero and 50, while server two will store everything between 51 hundred. Theoretically, youve now doubled your database capacity both in terms of storage and in the number of queries you can execute. This range is also called a keyspace. Its a simple system that solves two problems, where to store new items and where to find existing ones. All you have to do is calculate the hash of an items key and keep track of which server is responsible for which part of the key space. Now in this example, the range of zero to 100 is a bit small. It would only allow you to split up your database into 100 pieces at most.

(FIGURE 4.8: TRANSCRIPTION RESULT)

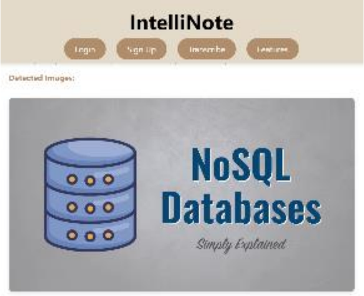


in the sense of non relational because it can't easily store relational data. So that was it for this video. Please subscribe if you learned something from it and I hope to see you in the next one. Bye.

Summary:

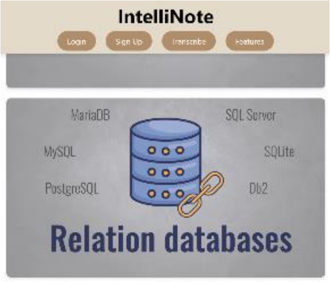
No. SQL databases have become very popular. Big companies rely on them to store hundreds of petabytes of data and run millions of queries per second. But what is a NoSql database? How does it work? And why does it scale so much better than traditional relational databases? Let's start by quickly explaining the problem with relational databases like MySQL, Mariadb, SQL Server and alike. These are built to store relational data as efficiently as possible. You can have a table for customers orders and products, linking them together logically. Customers place orders and orders contain products. This tight organization is great for managing your data, but it comes at a cost. Relational databases have a hard time scaling. They have to maintain these relationships, and that's an intensive process requiring a lot of memory and compute power. So for a while you can keep upgrading your database server, but at some point it won't be able to handle the load. In technical terms, we say that relational databases can scale vertically but not horizontally, whereas no SQL databases can scale both vertically and horizontally. You can compare this to a building. Vertical scaling means adding more floors to an existing building, while horizontal scaling means adding more buildings. You intuitively understand that vertical scaling is only possible to a certain extent, while horizontal scaling is much more powerful. Now, why do NoSQL databases scale so well? First of all, they do away with these costly relationships. In NoSQL, every item in the database stands on its own. This simple modification means that they are essentially key value stores. Each item in the database only has two a unique key and a value. For instance, when you want to store product information, you can use the product's barcode as the...

(FIGURE 4.9: SUMMARY RESULT)

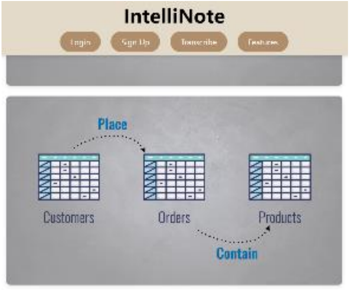


Detected Images:

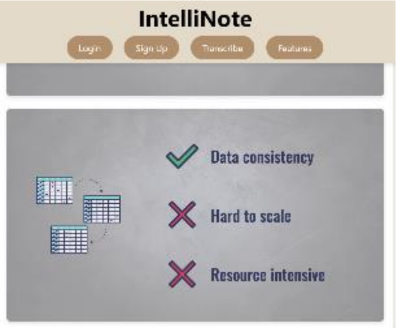
FIGURE 4.10: DETECTED IMAGE 1)



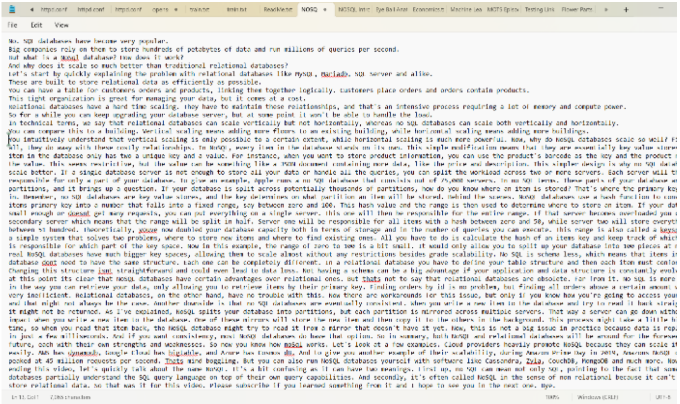
(FIGURE 4.11: DETECTED IMAGE 2)



(FIGURE 4.12: DETECTED IMAGE 3)



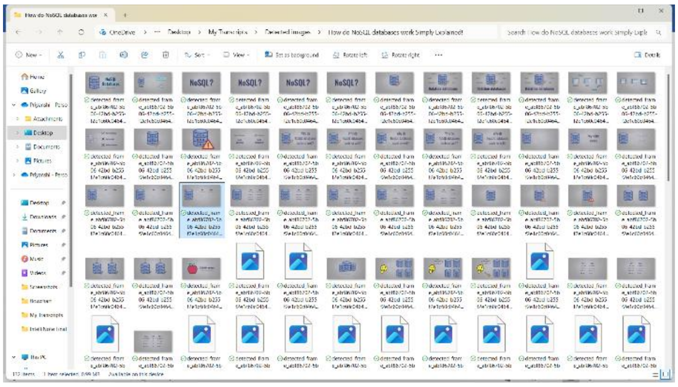
(FIGURE 4.13: DETECTED IMAGE 4)



```
No. SQL databases have become very popular.
Big companies rely on them to store hundreds of petabytes of data and run millions of queries per second.
But what is a NoSQL database? How does it work?
And why does it scale so much better than traditional relational databases?
Let's start by quickly explaining the problem with relational databases like MySQL , MariaDB, SQL Server and alike.
These are built to store relational data as efficiently as possible.
You can have a table for customers orders and products, linking them together logically. Customers place orders and orders contain products.
This tight organization is great for managing your data, but it comes at a cost.
Relational databases have a hard time scaling. They have to maintain these relationships, and that's an intensive process requiring a lot of memory and compute power.
So for a while you can keep upgrading your database server, but at some point it won't be able to handle the load.
In technical terms, we say that relational databases can scale vertically but not horizontally, whereas no SQL databases can scale both vertically and horizontally.
You can compare this to a building. Vertical scaling means adding more floors to an existing building, while horizontal scaling means adding more buildings.
You intuitively understand that vertical scaling is only possible to a certain extent, while horizontal scaling is much more powerful. Now, why do NoSQL databases scale so well? Fi
All, they do away with these costly relationships. In NoSQL, every item in the database stands on its own. This simple modification means that they are essentially key value store
item in the database only has two a unique key and a value. For instance, when you want to store product information, you can use the product's barcode as the key and the product n
the value. This seems restrictive, but the value can be something like a JSON document containing more data, like the price and description. This simpler design is why no SQL datab
scale better. If a single database server is not enough to store all your data or handle all the queries, you can split the workload across two or more servers. Each server will th
responsible for only a part of your database. To give an example, Apple runs a no SQL database that consists out of 75,000 servers. In no SQL terms. These parts of your database ar
partitions, and it brings up a question. If your database is split across potentially thousands of partitions, how do you know where an item is stored? That's where the primary key
in. Remember, no SQL databases are key value stores, and the key determines on what partition an item will be stored. Behind the scenes. NoSQL databases use a hash function to conv
items primary key into a number that falls into a fixed range, say between zero and 100. This hash value and the range is then used to determine where to store an item. If your dat
small enough or doesnt get many requests, you can put everything on a single server. This one will then be responsible for the entire range. If that server becomes overloaded you c
secondary server which means that the range will be split in half. Server one will be responsible for all items with a hash between zero and 50, while server two will store everyth
between 51 hundred. Theoretically, youve now doubled your database capacity both in terms of storage and in the number of queries you can execute. This range is also called a keys
a simple system that solves two problems, where to store new items and where to find existing ones. All you have to do is calculate the hash of an items key and keep track of which
is responsible for which part of the key space. Now in this example, the range of zero to 100 is a bit small. It would only allow you to split up your database into 100 pieces at m
real NoSQL databases have much bigger key spaces, allowing them to scale almost without any restrictions besides grade scalability. No SQL is schema less, which means that items in
database dont need to have the same structure. Each one can be completely different. In a relational database you have to define your table structure and then each item must confor
Changing this structure isnt straightforward and could even lead to data loss. Not having a schema can be a big advantage if your application and data structure is constantly evolv
at this point its clear that NoSQL databases have certain advantages over relational ones. But thats not to say that relational databases are obsolete. Far from it. No SQL is more
in the way you can retrieve your data, only allowing you to retrieve items by their primary key. Finding orders by id is no problem, but finding all orders above a certain amount w
very inefficient. Relational databases, on the other hand, have no trouble with this. Now there are workarounds for this issue, but only if you know how you're going to access your
and that might not always be the case. Another downside is that no SQL databases are eventually consistent. When you write a new item to the database and try to read it back straig
it might not be returned. As I've explained, NoSQL splits your database into partitions, but each partition is mirrored across multiple servers. That way a server can go down witho
impact when you write a new item to the database. One of these mirrors will store the new item and then copy it to the others in the background. This process might take a little bi
time, so when you read that item back, the NoSQL database might try to read it from a mirror that doesn't have it yet. Now, this is not a big issue in practice because data is repl
in just a few milliseconds. And if you want consistency, most NoSQL databases do have that option. So in summary, both NoSQL and relational databases will be around for the foresee
future, each with their own strengths and weaknesses. So now you know how noSQL works. Let's look at a few examples. Cloud providers heavily promote NoSQL because they can scale it
easily. AWS has dynamodb, Google Cloud has bigtable, and Azure has Cosmos dB. And to give you another example of their scalability, during Amazon Prime Day in 2019, Amazons NoSQL d
peaked at 45 million requests per second. Thats mind boggling. But you can also run NoSQL databases yourself with software like Cassandra, Zyla, CouchDB, MongoDB and much more. Now
ending this video, let's quickly talk about the name NoSQL. It's a bit confusing as it can have two meanings. First up, no SQL can mean not only SQL, pointing to the fact that some
databases partially understand the SQL query language on top of their own query capabilities. And secondly, it's often called NoSQL in the sense of non relational because it can't
store relational data. So that was it for this video. Please subscribe if you learned something from it and I hope to see you in the next one. Bye.
```

Ln 13, Col 1   7,065 characters   100%   Windows (CRLF)   UTF-8

(FIGURE 4.14: LOCALLY SAVED TRANSCRIPT)


How do NoSQL databases wor
OneDrive
Desktop
My Transcripts
Detected images
How do NoSQL databases work Simply Explained!
New
Sort
View
Details
NoSQL?

(FIGURE 4.15: LOCALLY SAVED IMAGES)

V. DISCUSSION ANALYSIS OF RESULT

Figures 4.1, 4.2 and 4.3 show the website interface.

Figures 4.4 and 4.5 show the Login and Sign Up dialogue. For now the only option availabe for Logging In or Signing Up is by entering the Email Id. The Facebook and Google options do not work right now , but it will be worked upon in the future. If the user forgets the Password while logging in, it does not have a feature like Forgot Password.

Figure 4.6 allows the user to choose the length of the summary. This will help them get important points from the video to go through quickly.

Figure 4.7 allows the user to select the language in which he wants his transcript to be. Right now, there are only a few languages but more languages will be added in the future implementations.

Figure 4.8 showcases the transcript that is generated from the user 's entered Youtube Link. The transcription right now is in paragraph form and does not have a structured manner to it. But it gives accurate transcription results.

Figure 4.9 is the summary generated from the transcript accoding to the user specified length. It still needs to be worked upon.

Figures 4.10, 4.11, 4.12 and 4.13 are some of the detected images from the YouTube video. The images get saved automatically in a folder with the name of the youtube video. The transcription is limitated to video transcription only for now and the data of the processed links is not being saved int the database.

Figure 4.14 shows the transcript that is saved locally on the computer. This is saved in the '.txt' format locally and can then be editied further. The user can later make them into structured notes and also add the detected images to make full proof notes of the topic. The transcipt gets saved by the name that the user inputs.

Figure 4.15 shows a folder with the video title in which detected images of the video are saved. This makes it easier to access the images of a particular topic avoiding the clutter of images.

This transcription will not only help the students to prepare for exams, but will also help professors to generate notes from their own recorded online videoas as study material for the students. This application can be useful for Online course creators for generating notes from their past lectures. This application is also useful for competitive exam learners to utilize YouTube news videos enhancing their study routine without watching the video.

Without having to watch the video lecture, learners can easily review key points and concepts from the lectures. This application will help students with hearing impairment to generate textual concepts of the videos. Learners will be able to see the flowcharts or diagrams from the video lecture in the generated notes.

VI. FEATURES OF OUR SYSTEM

IntelliNote provides the following features to the users:

1. Efficient Transcription:
As AssemblAI's API is used, the transcription generated is efficient and has corect words with an actual meaning to them. This feature will help the learners to understand the words and concept as it is without being confused.

2. Automaticaly Downloadable:
As soon the transcription is done, it gts saved locally. This contains the Transcript in the '.txt' format and the detected images in either '.jpg' or '.png' format. This will help them to learn when they do not have access to the internet.

3. Multilingual and Easy to Use Interface:
We provide multilingual feature to the users, so that they can learn and understand the notes in their native language or in language they are comfortable in.

4. Object Detection:
We provide not only transcription and summary to the user but also provide object detection from the provided YouTube video link. This will help capture important images like flowcharts, diagrams, memory-maps etc. and then they can be saved locally. This will eliminate the issue of having to watch the entire video to see if the user has missed any important point.

VII. DRAWBACKS

The proposed system has the following drawbacks:

1. UnStructured Notes:
Having notes in the paragrapgh form can complicate things sometimes and might distract the user while reading them.

2. Time Complexity:
The longer the videos, the more time it takes for generating transcription, summary and takes even longer time to detect images for every frame and then save them. This increases the time complexity and may sometimes, time out the write function and generate an error.

3. Audio dependent Translation:
As of now, only the videos that have their audio in English are getting transcribed, and hence the videos that are in other languages are not getting transcribed currently.

4. Output Language:
The solution provides only a few languages for transcription. Regional Indian Languages are not included.

5. Storing User Data:
The transcripts that are generated by the user are not saved in the database as for now and so the user cannot access the old transcriptions again but will have to transcribe them again.

6. Transcription:
The transcription is getting saved in '.txt' format and not in word form. The transcription and images atre getting saved seperately and not together.

VIII. OUR SYSTEM VS EXISTING SYSTEM

1. Current note-taking systems typically depend on manual entry, which means users must type or write notes themeselves. This can be a slow process and often leads to important details being overlooked. IntelliNote changes the game by using AI to transform lecture audio into organized notes, create relevant diagrams, and **support real-time collaboration**, significantly improving the speed and accuracy of note-taking.

2. Existing systems often face challenges in providing effective **multilingual support,** usually needing extra tools or manual translation. In contrast, IntelliNote can deliver smooth multilingual support by utilizing AI for real-time speech-to-text conversion in various languages, guaranteeing precise and organized notes no matter the language used during the lecture.

3. In current systems, **detecting images and diagrams** often necessitates manual input or the use of external tools, which can hinder effective note-taking. IntelliNote stands out by automatically identifying and creating pertinent images and diagrams from lecture materials through machine learning, thereby simplifying the note generation process with little need for user involvement.

4. Existing systems typically create unstructured notes, placing a significant burden on users to organize the content. In contrast, IntelliNote automatically produces **well-structured notes** by categorizing lecture material, organizing key points, and connecting them with relevant

diagrams. This ensures a clear, concise, and coherent flow of information without the need for manual formatting.

5. Existing systems usually necessitate manual processes for downloading notes, often missing automated features. IntelliNote, on the other hand, offers an **automatic notes download option**, allowing users to quickly download well-structured and formatted notes, along with relevant diagrams, right after a lecture, without any manual effort.

6. Current systems typically do not offer built-in video lecture summarization, which means users have to go through the content themselves to condense it. In contrast, IntelliNote utilizes **AI-powered summarization** to automatically create brief summaries of video lectures, highlighting important points and key insights, which greatly cuts down the time required for review.

IX. FUTURE IMPLEMENTATION

1. Structure the Notes:
We will structure the notes point-wise so that it beomes easy to understand understand and eye pleasing while studying.

2. Time Complexity:
Reduce the time complexity for the entire process no matter the length and duration of the input video. This will make it easy for users to get transcription of long videos and save time.

3. Summary:
Use Artificial Intelligence and Machine learning to improve the summary so that it gives information which is important only.

4. Convertible Transcript:
We will try to give more options to the user so that he can download the transcription in either '.txt', '.doc' or 'pdf' format according to the user preferences.

5. Multilingual Support:
We will implement multilingual support for Indian languages so that the users can read or understand the notes in their comfortable language.

6. Remove Audio Dependency:
We will improve the proposed solution in such a manner that every and any video that has audio other than English can be transcribed.

7. User Data Storage:
We will provide users the option to Sign In or Log In using their google or facebook acount. We will also add forget password and security questions for the ease of user's login. We will also keep a record of all the transcriptions that the user does when he is Signed In or Logged In so that if in future the user wants to access the transcription of any topic if the files are deleted locally, then he will not face any problem.

8. Provide Transcription for other source:
In future, we will implement Text to Sturctured Notes as well, so that the users can not lose any study material when they are learning.

9. Generate Flowcharts and memory maps based on the transcription to easily memorize the concepts of the topics.

X. CONCLUSION

The proposed solution aims to provide Transcript and also save the detected images for learning purposes. It not only provides Transcript and detected images but also provides summary of the transcript. Users can choose the summary length and also select the language in which they want their notes to be in.

The interface of the website and the working are simple and minimal for the ease of assistance of the user. The Transcription time and object detection depends on the length of the video. As of now, only the videoas that have English audio are getting transcribed, and only the videoas that are available on the YouTube.

IntelliNote will help users save time and can accessed anytime and anywhere. This solution will help not only students but also teachers or other enthusiasts that wish to learn something new.

XI. REFERENCES

[1] Oleh Basystiuk, Natalya Shakhovska, Violetta Bilynska, Oleksij Syvokon, Oleksii Shamuratov, Volodymyr Kuchkovskiy, The Developing of the System for Automatic Audio to Text Conversion, CEUR-WS(CEUR Workshop Proceedings), Lviv Polytechnic National University,12 Bandera str., Lviv, 79013, Ukraine, Vol-2824, pages 8, 2022

[2] Munish Saini, Vaibhav Arora, Madanjit Singh, Jaswinder Singh, Sulaimon Oyeniyi Adebayo, Artificial Intelligence inspired multilanguage framework for note-taking and qualitative content-based analysis of lectures, Springer Link, Vol-28, pages 1141-1163, 2023.

[3] Thomas Lancaster, Artificial Intelligence, text generation tools and ChatGPT – does digital watermarking offer a solution? International Journal for Educational Integrity, BMC Part of Springer Nature, Article No. 10, pages 14, 2023.