Testing Whisper AI as a Tool for Digital Library Workflows

Kenneth Cotich

The Claremont Colleges Library

Follow this and additional works at: https://scholarship.claremont.edu/library_staff

Recommended Citation
Cotich, Kenneth, "Testing Whisper AI as a Tool for Digital Library Workflows" (2024). Library Staff Publications and Research. 78.
https://scholarship.claremont.edu/library_staff/78
Testing Whisper AI as a Tool for Digital Library Workflows

Kenneth Cotich, PhD
Open Publishing and Digital Production Associate
What is Whisper AI?

Whisper is an automatic speech recognition (ASR) system trained on 680,000 hours of multilingual and multitask supervised data collected from the web.

Approaches human level robustness and accuracy on multilingual speech recognition.

Coded in Python and tested using Pytorch, an Open Source deep learning framework.

Hosted in an Open Source repository that’s constantly developed.
Workflow Experimentation

Tested on legacy and homeborn Audio visual (AV) content in the digital library

Transcriptions provide accessibility/improve search indexers

OpenAI Whisper - GitHub Quick Start
Install - CPU with <3.5GHz, ≥32GB RAM
Integrate multilingual model

Text files, VTT and subrip files for closed captioning

Update records in the CCDL
Example: [Stitching the 24-hour City: Life, Labor & the Problem of Speed in Seoul](#)

Example: [Drucker Archives](#)
So this overpass is the one that connected the central east part of Seoul to the airport or the highways that go to overseas or that go to other parts of South Korea.
[Pause]
>> The Dean of the Graduate School of Business
and in that capacity it is my pleasure to welcome you here.
You're a very roly crowd.
We hold faculty meetings in this room and our faculty is
about this size.
We have a very large faculty and I can assure you it's a lot more
difficult to get them to come to order than it is to get you.
Dealing by the way has been referred to from time
to time as trying to herd cats, which gives you some sense
of what it's like to deal with tenured faculty.
One of our former tenured faculty is a man
by the name of Peter Drucker.
He is indeed a gentleman who is a legend in his own time
and needs no introduction so I'm not going to give him one.
I'm simply going to turn the microphone over to him.
He will speak for 30 or 20, 30 minutes and we'll then go
to Dr. Harrington Parker and I will introduce Dr. Parker
after Peter holds forth.
For those of you who don't know the drill,
we will have our two main panelists talk
and perhaps interact a bit with each other.
We'll break at 10:45 for coffee, come back at 11 o'clock,
1105 depending on how busy you are at that point and go
to comments and questions and discussion
from the floor followed by lunch and so forth.
Peter, the floor is yours.
>> Thank you very much, Dick West.
Let me first say that I am delighted to be back once more
at what for 20 odd years was my home
and where I still feel now at home than any other place.
And I'm delighted to be able to welcome as a panelist Dr. Harrington
Parker who is doing pioneering work and very distinguished work
in one of our leading major high tech companies in thinking
through working out fundamental new approaches
to the organization of work and of people.
And you know if you look back at the way we have been organizing
Findings

Errors: names, noise, audio quality, processing speed

More languages available through Whisper than in Adobe Premiere

The more RAM and GPU increases processing time

Improves the uploading process

Background procedure

Word error rate (WER) Less than 5%

60 min mp3=18 min text file (6495 words)

87 min mp4=38 min text file (12357 words)
New Directions

TCCS IT and Cloud Machine
Add external GPUs to Digital Library workstations
Classroom implementation
Oral Histories
Live streaming speech to text transcription
Thank you!

“whispers
I’ll be back... with your text file