Conversion and Analysis From Multiple Data Sources

Aaron Hammond 1[1] (aaron.hammond1@students.mq.edu.au)
[1] Macquarie University, Sydney, NSW

▶ **Background** This solution is designed to create a quick and simple method of converting and analysing various forms of data.

▶ **Intention:** The goal of my proof of concept is to utilise a combination of tools and scripts to effectively convert multiple data types into a single file format for subsequent text analysis in three steps.

▶ **Applicability:** This solution will be of use to any research that combines data from audio, literature and personally gathered information such as field notes and finds value in analysing these sources together.

▶ **Significance:** Allows for comparative textual analysis of multiple sources of data such as audio, notes and academic literature. It combines these three sources together with minimal steps, effort and can be performed on multiple files without issue.

▶ **Execution:** The process requires the use of Ubuntu's Windows distribution to link software: Joplin, xPDF, Mozilla Deepspeech. Voyant Server is then utilised to analyse the results.

▶ *More information available in the Github repository HammondA_PoC_Design*

Aaron Hammond 1[1] (aaron.hammond1@students.mq.edu.au)
[1] Macquarie University, Sydney, NSW

1. Place audio files into the interviews folder and PDF files into the articles folder

2. Open Ubuntu and execute the script that transcribes audio with Mozilla Deepspeech, converts PDFs using xPDF, exports fieldnotes from Joplin, and places all files within the Ready folder

3. Open Voyant Server and select all files within Ready for analysis

Mozilla's Project Deepspeech

is an open source Speech-To-Text engine, using a model trained by machine learning techniques based on Baidu's Deep Speech research paper. Project DeepSpeech uses Google's TensorFlow to make the implementation easier.

It is only supported on Linux distributions.
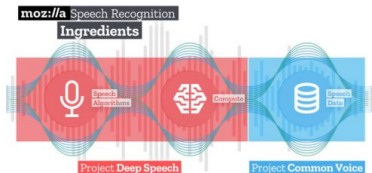
More information can be found at https://github.com/mozilla/DeepSpeech

Aaron Hammond 1[1] (aaron.hammond1@students.mq.edu.au)
[1] Macquarie University, Sydney, NSW

Joplin is an open source note taking and to-do application with synchronisation capabilities.

⇒ https://joplinapp.org/

Aaron Hammond [1] (aaron.hammond1@students.mq.edu.au)
[1] Macquarie University, Sydney, NSW
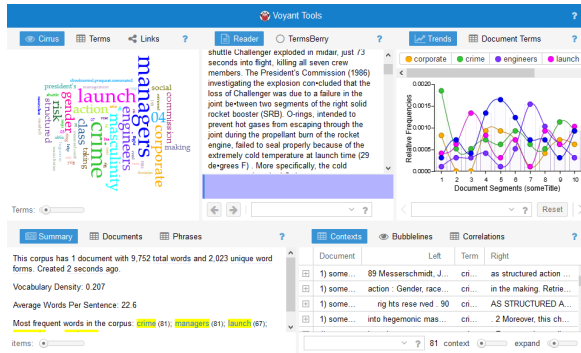
Voyant is an open-source application for performing text analysis. It supports scholarly reading and interpretation of texts or corpus.
⇒ http://docs.voyant-tools.org

Aaron Hammond 1[1] (aaron.hammond1@students.mq.edu.au)
[1]Macquarie University, Sydney, NSW

This proof of concept utilises the Linux distribution for Windows in order to create a seamless mode of converting various types of data. This is achieved by simply placing the required files in the corresponding folder and running a script using Ubuntu. The files will all be processed alongside a sync and export of Joplin. The export and converted files will then be placed in a single location ready for liaison with Voyant in order to perform the analysis.