STT Explorer: A Look Beyond Wer

Katsiaryna Mlynchyk and **Gabriel Hunziker** and **Lars Mosimann**
Ralph Scheu and **Mark Cieliebak**
mlyn@zhaw.ch

Abstract

STT Explorer introduces a new dimension to evaluating and inspecting STT (Speech-to-Text) results. While traditional metrics like Word Error Rate (WER) give a general idea of performance, they often miss the subtleties of STT output. STT Explorer is meant to fill this gap.

Through visual representation, the STT Explorer allows users to easily compare the performance of different STT engines by examining the alignment between reference and predicted texts. Users can upload STT results in JSON or CSV format, or input them manually. The tool then automatically aligns the texts, enabling a detailed examination of each sentence. Mistakes are color-coded for clarity, and alignments are displayed both succinctly (as dots) and in full (as words).

Current version of the tool utilizes a phonetic-based alignment algorithm, supports m-to-m alignment, meaning multiple words can align with multiple words if phonetically sensible. Additionally, the tool accommodates the alignment of multiple hypotheses from various STT engines. Detailed corpus statistics are also provided for further analysis.