

Leveraging AI for Semi-Automatic Scoring of Running Records



Abstraction and Objective

Running records is a literary assessment method employed by teachers in order to gauge a student's reading level. Leveraging AI for semi-automatic scoring of running records (LASARR) is the working title for the project that utilizes Watson for its speech recognition. (LASARR) seeks to properly record all spoken speech regardless of accent and dialect, and to be able to grade the spoken text. The purpose is not to completely automate the process, but to expedite it, and allow teachers to have a reference for growth comparisons.





Methods

One of the main challenges with this project is understanding where in the text the subject is. We have devised a general algorithm to perform this task:

- Determine list of all possible mappings between individual words
- Determine all permutations of potential word vector maps
- Determine optimal permutation by finding most "continuous" word vector map

We have packed the algorithm into the text-tracker Python package, available for free on pypi.org by running the command pip install text-tracker from the command line.

Results

The algorithm runs extremely fast on small sample sizes, running several low reading level pages in under a tenth of a second. The package will be further optimized in the future using multithreading and graphics acceleration techniques.

This example ran a short sentence in under a millisecond.

Citations

Running Records: Ross, John A. "Effects of running records assessment on early literacy achievement." The Journal of Educational Research 97.4 (2004): 186-195. Watson Speech to Text: Santiago, Felipe, Pallavi Singh, and Lak Sri. Building Cognitive Applications with IBM Watson Services: Volume 6 Speech to Text and Text to Speech. IBM Redbooks, 2017.

[(venv) (base) Coopers-MacBook-Pro:Text Tracker Example coopersanders\$ python example.py Example written sentence:
one day a bird met a hippo
Example spoken sentence:
one day a bird a word bird met a hippo
Word map vector:
[0, 1, 2, 3, 2, nan, 3, 4, 5, 6]
Time elapsed: 0.97 ms



