Internship openings at Glose

Glose's mission is to improve reading and learning for students. With Glose Education, students have access to thousands of digital books on their computers, tablets or smartphones. Students and teachers can interact around the text by exchanging comments or voice notes, highlighting passages or reacting with emojis.

How to apply:
By email at jobs@glose.com, with an up-to-date resume attached.

Glose in a few words it's :

- A digital bookstore of **1.5 million books distributed in 200 countries**
- **Thousands of students and teachers** who read and discuss texts on our platform
- The official platform for the distribution of **CNED courses** (National Center for Remote Education)

Glose takes interest in all research topics that **improve the reading experience** of its users. Glose's R&D therefore covers a wide variety of cutting-edge technologies:

- Recommendation of texts according to their complexity
- Recommendation of texts according to their content
- Syntax analysis (POS tagging)
- Named Entity Recognition (NER)
- End Of Sentence segmentation (EOS)
- Automatic analysis of read alouds
- Automatic generation of comprehension tests

The R&D team is looking to grow to bring more innovations into the hands of our students. For all the subjects below, we expect the same work from the intern:

Expectations:

- Make a state of the art review for the task at hand
- Identify potential solutions
- Implement a prototype
- Test the prototype and evaluate it against our data

Topic 1: Automatic feedback for read alouds

This is a research subject linked with Glose's "À Haute Voix" project. This very ambitious project will allow students to improve their reading and comprehension skills as well as their oral fluency.

Prosody is an essential skill to master reading. Prosody is how we read a text giving it life and meaning. This requires mastering the pauses, accents, pace, intonations and intensity of the voice. This therefore requires a good understanding of the syntax, semantics and punctuation of the text. Research has also established a link between prosody and text comprehension.

Glose has developed an audio recording feature within its platform. This allows students to record read alouds on any text and share them with their class.

The goal of the project is to develop an algorithm that automatically analyzes the read alouds and gives students feedback in a detailed report:

- Correspondence of the support text and the spoken text
- Pronunciation
- Intonation
- Pace
- Expressiveness

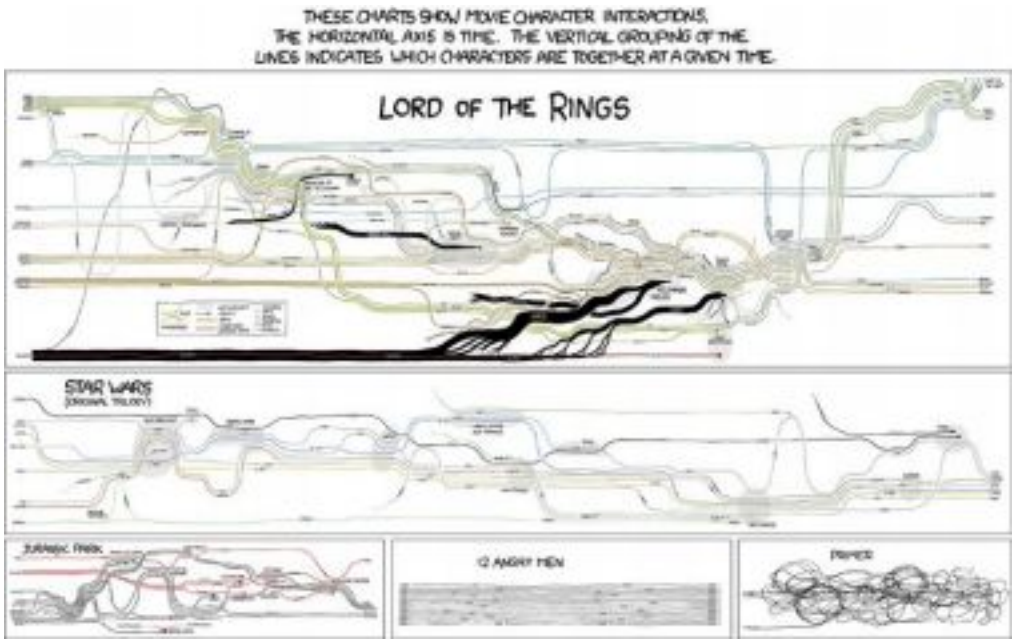
Where it is relevant, the algorithm should be able to indicate where the error was made (non-respect of punctuation, wrong pronunciation, bad intonation, etc.). The subject covers a variety of subjects that you will need to familiarize yourself with: **Automatic Speech Recognition, Natural Language Processing and Signal Processing.**

References:

When Speech Input is Not an Afterthought: A Reading Tutor that Listens: https://www.cs.cmu.edu/~listen/pdfs/Mostow_Aist_pui97_final.pdf
Predicting Multidimensional Subjective Ratings of Children' Readings from the Speech Signals for the Automatic Assessment of Fluency: https://www.aclweb.org/anthology/2020.lrec-1.39/
Performance Analysis of Several Pitch Detection Algorithms on Simulated and Real Noisy Speech Data: https://www.eurasip.org/Proceedings/Eusipco/Eusipco2017/papers/1570342455.pdf
The Kaldi Speech Recognition Toolkit: https://infoscience.epfl.ch/record/192584

Topic 2: Topic extraction from a text

Our catalog is currently analyzed to extract named entities. We want to improve this and extract the important "topics" of each text in order to build - above these entities and topics - graphs representing in an innovative way the content of the books. (see for example the illustration below).



Another line of work will be the construction of a recommendation graph between the different books to suggest to our users their next readings. The volume of data to be processed will have to be taken into account in the choice of methods and their implementation (> 1M books).

References:

Comparison of topic extraction approaches and their results: https://link.springer.com/article/10.1007/s11192-017-2306-1
A Survey on Recent Advances in Named Entity Recognition from Deep Learning models: https://arxiv.org/abs/1910.11470
Scientific Article Recommendation by using Distributed Representations of Text and Graph: https://dl.acm.org/doi/abs/10.1145/3041021.3053062

Topic 3: Automatic generation of comprehension questionnaires

Comprehension tests are widely used by teachers. They allow students to test their understanding and remember important elements of the passage they have just

read.

Our R&D team wishes to develop comprehension tests for our reading platform. So as soon as a student finishes a chapter, he can answer a few questions to assess his understanding of the chapter.

The goal will be to automatically generate these questions. It is therefore necessary to select the important content, create a questionnaire and select distractors (the wrong answers in the multiple choice questions).

The automatic generation of comprehension tests is a vast subject which is increasingly active in the scientific community. Depending on the method used, the algorithms make use of various concepts:

- Text simplification
- Content selection
- Syntax analysis
- Semantic analysis
- Classification of sentences
- etc.

We will have to identify the most promising methods, adapt them to our use case and test them on our data.

References: A Systematic Review of Automatic Question Generation for Educational Purposes: https://link.springer.com/article/10.1007/s40593-019-00186-y

Topic 4: Detection of salient passages / simplification of texts:

Study the algorithms for detecting the salient points of a text to increase our reading interface with markers. We can highlight important passages in a book - to help our students focus on the difficult parts for example - then use these algorithms to automatically generate summaries at any scale on the book (book / chapter) which will allow our readers to easily dive back into the context of their reading without having to reread the entire previous chapter.