

Student Research Symposium 2004

Jonathan Brown

Retrieval of Authentic Documents for Reader-Specific Lexical Practice

When a teacher gives a reading assignment in today's language learning classrooms, all of the students are almost always reading the same text. Although students have different reading levels, it is impractical for a single teacher to seek out unique texts matched to each student's abilities. In this presentation, I describe REAP, a system designed to assign each student individualized readings by combining new techniques in reading difficulty estimation[1] and detailed student and curriculum modeling[2] with the large amount of authentic materials on the Web. REAP is designed to be used as an additional resource in teacher-led classes, as well as to be used by reading comprehension researchers for testing hypotheses on how to improve reading skills for L1 as well as L2 learners. I describe how researchers can use this tool to get fine-grained control over selection of reading materials, so that they can more easily test these new learning hypotheses.

Vocabulary acquisition is the primary factor we use in matching texts to a student's abilities. These abilities are modeled as a histogram of words. We also model each desired curriculum level as a histogram of words, learned from a corpus of texts that the students would normally read. Differences between the student model and that of the next desired skill level indicate where the student needs to focus. The system can also prioritize different criteria during the search. For instance, the system can retrieve documents based solely on the vocabulary terms needed to progress toward the next level, thereby focusing on curriculum. REAP can also take into account other goals, such as student interests, special topics, or an upcoming test, all represented as word histograms. This allows teachers and researchers to decide what they want the students to focus on for each session.

[1] K. Collins-Thompson and J. Callan. (2004.) "A language modeling approach to predicting reading difficulty." In Proceedings of the HLT/NAACL 2004 Conference. Boston.

[2] J. Brown and M. Eskenazi. (2004.) "Retrieval of Authentic Documents for Reader-Specific Lexical Practice." In Proceedings of InSTIL/ICALL Symposium 2004. Venice, Italy.