inesc id lisboa

L²F – Spoken Language Systems Laboratory

Digital Talking Books

Digital Talking Books is a spreading commodity amongst users. They provide fast access to the audio or written texts of multimedia documents.

As an output of this activity, L2F has a platform to time align text and speech files in one passage, faster than real time. Audio files can be reasonably long (over 2 hours), avoiding partitioning the narration into smaller files.

Goal. To allow Digital Talking Books (DTBs) to be used by a wider audience of users, by providing a browsing interface for navigation in multimedia documents. DTBs are thus an important tool for e-learning and e-inclusion.

Summary. This activity is the result of two nationally funded research projects, IPSOM (form Nov 2000 to Nov 2004) and RICOBA (still undergoing) that started with a consortium with L2F (coordinator), LASIGE of the Faculty of Sciences of the University of Lisbon and the National Library. The undergoing idea is to have the audio and text files synchronized at the word/topic/sentence level, thus allowing the two instances of a book (spoken and written) to be accessed simultaneously. This concept was further extended to encompass different varieties of Portuguese (project LIFAPOR with UFRGS, Brazil) and also parallel texts in other languages.

Description. Audio Books (AB) started early as vinyl or cassette recordings. However, looking for any information in these recordings was, at least, a trial-and-error experience. Computers, with their powerful storage, indexing and retrieval capabilities, were the obvious choice for AB, thus transforming them into Digital Talking Books, by the inclusion of metadata. Forced-alignment speech recognition technologies and Weighted Finite State Transducers, allowed us to synchronize (time align) simultaneously the audio and the text versions of the DTBs. The audio version of a DTB, previously recorded and manually edited to remove reading errors, is time aligned against its written version in less than real-time. A multimodal navigation and information retrieval interface, plus the audio and text files comprise the DTBs that can be further complemented with short videos or other multimedia data.

More information is available by email to info@l2f.inesc-id.pt or directly from the website http://www.l2f.inesc-id.pt/.