Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Pronunciation Discovery in Large Spoken Databases
Thesis Proposal (Master, possibly Bachelor)

Motivation: Large Spoken Databases are used to investigate the correspondence between spoken words and written text. One example is the Spoken Wikipedia[1] which contains hundreds of hours of speech for hundreds of articles, spoken by many different speakers, thus giving a broad impression of read speech in general. One of the peculiarities worth investigating are the many ways in which words can be pronounced. This is particularly relevant for atypical material, such as abbreviations (e.g., 'AIDS' is enunciated but 'HIV' is spelled), and special characters ('*' can be pronounced "asterisk", "star", or even "born on" if followed by a four-digit number).

Abstract: The goal of the thesis is to investigate the pronunciation variety for textual material with a focus on atypical word tokens. Experiments will be performed on the Spoken Wikipedia Corpus [1] using English and/or German language data. The purpose is to extend the corpus alignment coverage and to improve pronunciation models for textual material. The process may be structured as follows:

- acquire an understanding of reading styles in the Spoken Wikipedia
- analysis of the problem and development of a plan of attack considering the available data[2] and software (e. g. [2, 3, 4])
- manually annotate some development and test data
- acquire knowledge in the relevant software and implement additional solutions (e. g. implement restricted speech recognition search in previously un-aligned portions of the corpus using context dependent pronunciation candidates)
- test and evaluate the solutions and generalize conclusions from the data
- improve the methods for pronunciation modelling, e. g. taking context into account
- evaluate the improved pronunciation model in a speech synthesis context

Successful candidates have developed or will develop a solid understanding of the (bash) commandline, diligent data management, and a talent for staing organized. Previous contact with speech processing, computational linguistics and machine learning is profitable but not essential.

---

[1] http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Spoken_Wikipedia
[2] http://nats-www.informatik.uni-hamburg.de/SWC/

The thesis work can be developed and written in English or German; relevant literature is almost exclusively available in English. We aim and actively encourage our students to publish their relevant research results at a international or national conference.

Keywords: Speech processing, Pronunciation modelling, Big data, Wikipedia.

Literatur

[1]   Arne Köhn, Florian Stegen und Timo Baumann. „Mining the Spoken Wikipedia for Speech Data and Beyond". In: *Proceedings of LREC 2016*. 2016.

[2]   Florian Schiel. „MAUS goes iterative". In: *Proceedings of the LREC*. 2004.

[3]   Maximilian Bisani und Hermann Ney. „Joint-sequence models for grapheme-to-phoneme conversion". In: *Speech Communication* 50.5 (2008), S. 434–451. ISSN: 0167-6393. DOI: 10.1016/j.specom.2008.01.002.

[4]   Willie Walker u. a. *Sphinx-4: A Flexible Open Source Framework for Speech Recognition*. Techn. Ber. SMLI TR2004-0811. Sun Microsystems Inc., 2004.

Kontakt

Timo Baumann (*baumann@inf…*), Prof. Wolfgang Menzel

URL dieses Dokuments: