



Aktualisierungen und Erweiterungen der Gesprochenen Wikipedia durch Sprachsynthese

Abschlussarbeit (Master oder Bachelor)

Motivation: Von Menschen gesprochene oder vorgelesene Inhalte (PodCasts, Audiobücher, ...) sind gegenüber Sprachsynthese wesentlich besser verständlich und angenehmer zu nutzen, weil der menschliche Vorleser wesentlich intelligenter ist und damit besser vorlesen kann als ein Computer. Audioaufnahmen sind allerdings inflexibel falls sich die gelesenen Inhalte nach der Aufnahme ändern (oder anders gelesen werden sollen).



Kurzfassung: Die gesprochene Wikipedia¹ umfasst etwa 900 vorgelesene Artikel (insgesamt 270 Stunden Audiomaterial für Deutsch); gegenüber maschinell vorgelesenen Artikeln² sind sie deutlich verständlicher und angenehmer zu konsumieren, veralten aber schnell durch die konstante Überarbeitung der Wikipedia. Einige Vorleser sprechen so viel Material, dass sich daraus Sprachsynthesestimmparameter generieren lassen. Diese Arbeit soll die Anpassung und Erweiterung gelesener Inhalte mittels (passender) Sprachsynthese untersuchen um die Vorteile von Mensch und Maschine zu verbinden.

Ziel der Abschlussarbeit ist es, vorgelesene aber veraltende Wikipedia-Artikel mittels kontextangemessener Sprachsynthese aktuell zu halten. Hierfür soll einerseits der Sprechstil und die Prosodie des Vorlesers möglichst gut auf die zu synthetisierenden Abschnitte übertragen werden. Weiterhin soll auch die Vorlesestimme möglichst gut getroffen werden, indem aus dem Material eine Sprachsynthesestimme generiert wird, die für die Synthese genutzt wird. Dies würde es auch ermöglichen, beliebige andere, nicht originär vorgelesene Artikel in hoher Qualität zu synthetisieren.

Stichworte: Sprachsynthese, Mensch-Maschine-Interaktion, Sprechschnittstelle, Wikipedia

¹http://de.wikipedia.org/wiki/Portal:Gesprochene_Wikipedia

²<http://www.pediaphon.org>

Der Arbeitsprozess könnte sich in folgende Schritte gliedern:

- Aneignung von Wissen über gesprochene Sprache, Sprachsynthese (z.B. [1, 2])
- Einarbeitung in die Wikipedia-API und Analyse der Artikel-Edits in Hinblick auf die zu synthetisierenden Abschnitte
- Prototypische Perfektionierung des Alignments (sowie ggfs. Erstellung weiterer notwendiger Annotationen) sowie detaillierte Analyse und Klassifizierung der Edits für einige wenige Testartikel
- Analyse der vorliegenden Text-Audio-Alignments [3] und Rückbezug auf das originäre Wikipedia-HTML
- ggfs. Einarbeitung in das Sprachsynthesetraining zum Beispiel mit dem Mary Voice Builder, Festvox, oder HTS zur Erstellung von HMM- oder Unit-Selection-Stimmen
- Entwurf einer Methode um für fehlende Abschnitte eine kontextangemessene Prosodie zu generieren und Resynthese fehlender Abschnitte im Artikel
- Evaluation der Resynthese in quantitativen Korpus- und qualitativen Hörexperimenten

Die erfolgreiche Bearbeitung des Themas erfordert die Fähigkeit sich in teils umfangreiche vorhandene Softwareprojekte, Datensammlungen und Algorithmen einzuarbeiten, gute Programmierkenntnisse in JAVA sowie Kenntnisse auf der Kommandozeile (oder die Bereitschaft sich diese zu erarbeiten). Vorwissen aus den Lehrveranstaltungen Sprachverarbeitung, Computerlinguistik und/oder User Interface Design ist nicht notwendig, aber von Vorteil.

Die Abschlussarbeit kann auf Deutsch oder Englisch erarbeitet und verfasst werden; relevante Literatur ist überwiegend auf Englisch verfügbar; Daten sind für Deutsch und Englisch verfügbar. Bei jeder Abschlussarbeit streben wir die Veröffentlichung der Forschungsergebnisse auf einer internationalen oder nationalen Konferenz an.

Literatur

- [1] Daniel Jurafsky und James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2. Aufl. Pearson International, 2009.
- [2] Paul Taylor. *Text-to-Speech Synthesis*. Cambridge Univ Press, 2009.
- [3] Arne Köhn, Florian Stegen und Timo Baumann. „Mining the Spoken Wikipedia for Speech Data and Beyond“. In: *Proceedings of LREC 2016*. 2016.

Kontakt

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

URL dieses Dokuments:

