

Programmierpraktikum

Sommersemester 2007 - 13.04.2007

Björn Wilmsmann, B.A.
Sprachwissenschaftliches Institut
Ruhr-Universität Bochum
bjoern@wilmsmann.de

Allgemeines

- E-Mail: wilmsmann@linguistics.rub.de
- Telefon: 0234-32-25116
- Raum: GB 3/154
- Sprechzeiten: Nach Vereinbarung
- 14 Termine (letzter Termin: 13.07.2007)

Syllabus

13.04.2007	Einführung, Formalia
20.04.2007	Einführung in Eclipse
27.04.2007	Tokenisierung
04.05.2007	Stop Word Identification
11.05.2007	Named Entity Recognition
18.05.2007	Soundex Algorithmus
25.05.2007	Porter Stemming Algorithmus
01.06.2007	WordNet

Syllabus

08.06.2007	Huffman Coding
15.06.2007	Erwerb von Ontologien
22.06.2007	Clustering, Textkategorisierung
29.06.2007	Clustering, Textkategorisierung
06.07.2007	Applikation: Akronym-Generator
13.07.2007	Applikation: Akronym-Generator

Literatur

- Baeza-Yates, Ricardo / Ribeiro-Neto, Berthier (1999): *Modern Information Retrieval*. Boston, MA: Addison-Wesley.
- Cavnar, William B. / Trenkle, John M. (1994) *N-Gram-Based Text Categorization*. In: *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*. 161-175. Las Vegas, NV.
- Chakrabarti, Soumen (2003): *Mining the Web – Discovering Knowledge from Hypertext Data*. San Francisco, CA: Morgan Kaufmann.
- Christiansen, Tom / Torkington, Nathan (2003): *Perl Cookbook*. Cambridge, MA: O'Reilly.
- Halama, André (2004a), *Flache Satzverarbeitung*. In Klabunde, Ralf. et al. (Hrsg.), *Computerlinguistik und Sprachtechnologie. Eine Einführung*. Heidelberg: Elsevier.
- Halama, André (2004b): *Das Perl Tutorium für Computerlinguisten*. Bochum, Ruhr-Universität: <http://www.linguistics.ruhr-uni-bochum.de/~kiss/seminare/materialien/compling/perl2006.pdf>.
- Jurafsky, Daniel / Martin, James H. (2000): *Speech and Language Processing*. Upper Saddle River, NJ: Prentice Hall.

Literatur

- Knuth, Donald E. (1998): *The Art of Computer Programming*. Boston, MA: Addison-Wesley.
- Lutz, Mark (2006): *Programming Python*. Cambridge, MA: O'Reilly.
- Manning, Christopher D. / Schütze, Hinrich (1999): *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Martelli, Alex / Ascher, David / Ravenscroft, Anna (2005): *Python Cookbook*. Cambridge, MA: O'Reilly.
- Python.org: *Python Documentation*: <http://www.python.org/doc/>.
- Soundex: <http://en.wikipedia.org/wiki/Soundex>.
- Wall, Larry / Christiansen, Tom / Orwant, Jon (2000): *Programming Perl*. Cambridge, MA: O'Reilly.

Ziele

- Vorstellung verschiedener Problemstellungen und Algorithmen
- Einüben von Programmierfertigkeiten
- Kennenlernen von Prinzipien der Softwareentwicklung

Ziele

- Kennenlernen und Nutzung der Eclipse IDE
- Praktische Realisierung der vorgestellten Verfahren
- Realisierung einer Anwendung (geplant: Akronym-Generator)
- Evtl. Umsetzung als Web-Applikation

Implementierung

- Im Kurs wird hauptsächlich Perl verwendet werden
- Aufgaben können sowohl in Perl als auch in Python gelöst werden
- Implementierung als Modul - zwecks Wiederverwendbarkeit in späterer Anwendung

Studienleistung

- Aufgaben
 - Lösung in kleinen Gruppen von 2 bis 3 Personen
 - 10 Programmieraufgaben (je 7 %)
 - Implementierung einer vollständigen Applikation am Ende des Semesters (30 %)

Studienleistung

- **Bewertungskriterien**
 - **mathematische Korrektheit des Programmcodes**
 - **eleganter Programmierstil**
 - **Dokumentation**
 - **bei vorheriger Modellierung: Nähe zum Modell**

Prüfungsleistung

- vollständiger Entwurf, Implementierung und Dokumentation einer Anwendung
- das Thema ist im Rahmen der maschinellen Sprachverarbeitung frei wählbar, aber vorher mit dem Dozenten zu besprechen

Für die nächste Sitzung

- Eclipse IDE besorgen (im Studienbüro, GB 3/157)
- Installation der Eclipse IDE
- Wiederholung: Was ist Tokenisierung?
 - Siehe dazu auch: Halama (2004a), Kap. 3.3.1 (pp. 218-223) (ebenfalls im Studienbüro, GB 3/157)

Für die nächste Sitzung

- Für Windows-Nutzer: Perl (und Python) installieren
- <http://www.activestate.com/Products/ActivePerl/>
- <http://www.activestate.com/Products/ActivePython/>

Heute

- Eintragen in die Liste

**Vielen Dank für Ihre
Aufmerksamkeit!**