

Programmierpraktikum

Sommersemester 2007 - 13.07.2007

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

Heute

- Vorstellung des Zwischenstandes der Applikation
- Roundup
 - Was haben wir dieses Semester gemacht?
 - Was waren die Ziele der Veranstaltung
 - Bewertung

Roundup

- Worum ging es im Programmierpraktikum?
 - praktische Programmierfertigkeiten
 - Algorithmen
 - mathematische Zusammenhänge
 - Methoden der Softwareentwicklung

Roundup

- Softwareentwicklung per IDE
 - Eclipse
 - EPIC
 - SVN, Subclipse

Roundup

- ‚guter‘ Code in Perl
 - Guidelines wider ‚write only‘ Code
 - verschiedene Pragmata (use strict etc.)
 - Namenskonventionen
 - Perl-Module
 - Dokumentation

Roundup

- weitere Perl Tools
 - Hilfsmodul, z.B. List::Utils
 - CPAN

Roundup

- Datenstrukturen in Perl
 - Skalare
 - Arrays
 - Hashes
 - verschachtelte Datenstrukturen
 - Referenzen

Roundup

- Tokenisierung
 - symbolisch vs. statistisch
 - Problemfälle: Abkürzungen, Klitika, Formatierung von Zahlen

Roundup

- Textnormalisierung
- Abbildung von Token auf Äquivalenzklassen
- Porter Stemmer
- Soundex

Roundup

- Information Retrieval
 - Recall
 - Precision
 - $tf*idf$

Roundup

- WordNet
- Semantic Web
 - OWL
 - RDF

Roundup

- Graphen
 - Graphentheorie
 - kürzeste Pfade (Dijkstra Algorithmus)
 - Breitensuche
 - Tiefensuche
 - Adjazenzlisten, Adjazenzmatrizen

Roundup

- Laufzeit von Algorithmen
 - O-Notation

Roundup

- XML
 - Datenstrukturen
 - Markupssprachen im Allgemeinen
 - XML für linguistische Daten
- RSS
- (X)HTML

Roundup

- XML
 - DTD XSD, XSLT, XPath...
 - Grammatiken, Chomsky-Hierarchie
 - XML-Verarbeitung mit Perl

Roundup

- Heaps
 - Priority Queue
 - Min Heap, Max Heap
 - Binomial Heap, Fibonacci Heap
 - Merge-Find Sets (Mischwälder)
 - Goldener Schnitt

Roundup

- Vektorraummodell im Information Retrieval
 - Dokumentvektoren
 - Inverted Index
 - Kosinus-Ähnlichkeit

Roundup

- effiziente Datenübertragung / Klassifikation von Daten
 - Fourier Transformation
 - LPC, Quantisierung
 - Clustering, Centroiden
 - k-means, LBG

Roundup

- Datenkompression
 - Statistische Kodierung
 - Arithmetische Kodierung
 - Kodierung über Wörterbücher
 - statisch vs. adaptiv

Roundup

- Datenkompression
 - Entropie
 - Huffman-Coding
 - Tries

Roundup

- Datenstrukturen zur Speicherung von Strings
 - Automaten-Theorie, Hashes
 - Tries
 - Suffix Trees, Suffix Arrays
 - Patricia Trees

Roundup

- Klassifikation
 - hierarchisches Clustering: bottom-up, top-down
 - k-means
 - Naive Bayes Classifier
 - Verfahren nach Cavnar & Trenkle (City Block Metrics)

Roundup

- Methoden der Softwareentwicklung
 - Prozessmodelle
 - Phasen der Softwareentwicklung
 - Artefakte

Roundup

- Methoden der Softwareentwicklung
 - Modellierung
 - Lastenheft, Pflichtenheft
 - UML-Diagramme (Klassendiagramm, Aktivitätsdiagramm, Use Case Diagramm, Sequenzdiagramm usw.)

Ziele

- Worum ging es im Programmierpraktikum?
 - praktische Programmierfertigkeiten
 - Algorithmen
 - mathematische Zusammenhänge
 - Methoden der Softwareentwicklung

Ziele

- Worum ging es im Programmierpraktikum?
- üblicherweise nur theoretisch bekannte Algorithmen praktisch zu implementieren
- Kennenlernen realer Anwendungen und Verfahren für die Bereiche Information Retrieval, Data Mining und Signalverarbeitung

Ziele

- Worum ging es im Programmierpraktikum?
- Umsetzung mathematischer Formalismen in Programmcode
- Methoden der Softwareentwicklung praktisch zu üben

Noch Fragen?

• ...

Abschließende Bewertung

- Evaluation
- Meinungen
- Verbesserungsvorschläge
- eigene Ideen

**Vielen Dank für Ihre
Aufmerksamkeit!**