

Web Mining Übung

- www.ke.tu-darmstadt.de/lehre/ss11/web-mining/uebungen
 - zusätzliche Informationen, Registrierung, Upload, Übungsblätter
- Aufgaben
 - aus dem Bereich Data-, Text- und Web-Mining
 - Crawling, Textanalyse, Textklassifizierung, Clustering, Information Extraction, etc.
 - großer Spielraum bei Lösungsfindung
 - Programmierung ist notwendig
 - aber die Programme sind nur Mittel zum Zweck
- Umfang
 - Aufgaben teilweise zeitintensiv, je nach Programmiersprache, Vorkenntnisse, Programmierfähigkeit etc.
 - dafür praktischer Einsatz der Techniken aus der Vorlesung

Web Mining Übung

- Zeitplan
 - 6 Aufgaben
 - ca. alle 2 Wochen
 - Abgabe sonntags
 - Übungsstunde dienstags
 - neue Übungsblätter in der Regel Anfang der Woche
 - erste Übung Abgabe 1. Mai, Besprechung 3. Mai
- Beurteilung:
 - 10+2 Bonuspunkte je Übung, max. 60 Punkte
 - Verbesserungen bis zu einem Notengrad sind möglich
 - plus „Wissensbonus“ natürlich
 - nur bei bestandener Klausur!

Web Mining Übung

- Gruppenarbeit möglich
 - Gruppengröße max. 3
- Registrierung
 - über Upload-Seite
 - wird bald freigeschaltet!
 - RBG-Login notwendig
 - ein Teammitglied meldet die Gruppe für alle an
 - theoretisch bis 1. Mai
 - Gruppenwechsel nur nachträglich möglich

Web Mining Übung

- Ablauf Übungsstunde
 - Durchbesprechen der abgegebenen Lösungen
 - Gruppen stellen abwechselnd ihre Lösungen vor
 - mindestens ein Teammitglied muss anwesend sein
 - und vorführen können!
- Abgabe
 - über Upload-Seite
 - ein Teammitglied lädt für alle hoch
 - letzter Upload wird gewertet
 - Upload einer Zip-Datei

Web Mining Übung

- Abgabe enthält
 - Lösungsdokument
 - PDF-Datei
 - präsentierfähig
 - Ergebnisdateien
 - Tabellen, Grafiken etc.
 - Quelldateien und Programm
 - kompilierbar und ausführbar
 - Beispieldateien
 - benutzter Korpus, Webseiten, Wörterbücher, etc.

Web Mining Übung

- Form des PDFs
 - Lösung ohne mündliche Erklärung nachvollziehbar!
 - Benotung soll allein anhand PDF-Datei möglich sein
 - zusätzliches Material nur als Nachweis oder für Detailfragen
 - Referenzen auf benutzte Beispieldateien angeben

Web Mining Übung

- Tabellen, Diagramme, Graphen:
 - vollständige Beschriftung der Achsen
 - kurze Beschreibung

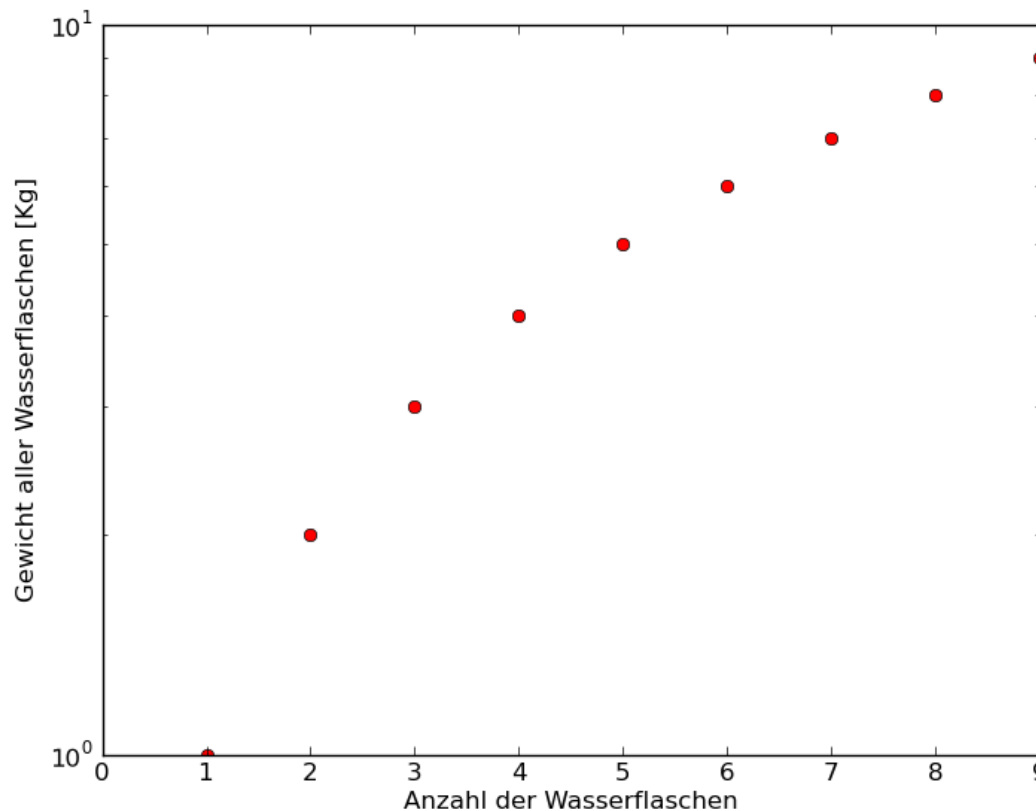


Abb. 2: Gewicht in Abhängigkeit in Abh. der Flaschen in sinnlosem Datensatz A

Web Mining Übung

- Codelistings, Pseudocodes:
 - Aufzeigen der Funktions / Funktionsweise
 - nicht einfach den Source Code hineinkopieren!
 - unbedingt Erklärung!

```
perl -n -l -a -e 'BEGIN{$x=0} {$x+=$F[$#F]} END{print $x}' preisliste.txt
```

Abb. 1: Aufsummierung der Preise (letztes Wort einer Zeile) einer Preisliste

Web Mining Übung

- Programmierung
 - beliebige Sprache
 - besonders geeignet: Skriptsprachen
 - Python, Perl, Ruby, Groovy
 - aber auch Java, **Javascript**, etc.
 - Benutzung von Libraries erlaubt
 - doch Aufgabenstellung ist zu beachten
 - „implementieren“ heißt nicht „verwenden“
 - einige auf Homepage aufgezeigt
- Weitere Tools
 - matplotlib, jfree, gnuplot für Plots
 - r-project für statistische Berechnungen und Grafiken
 - graphviz fürs Zeichnen von Graphen
 - etc. etc. etc.

Web Mining Übung

- Betreuung
 - Eneldo eneldo@ke.tu-darmstadt.de
 - inhaltliche Frage, Übungsbetrieb, Übung
 - Sprechstunde: Dienstag 16:00-17:00
 - Clemens doerrhoefer.c@googlemail.com
 - Übung, Umsetzung
 - Forum