

Web Mining: Übung 2






Lösungsvorschlag



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Aufgabe 1:

Einteilung der Testdaten

- ▶ Verwendeter Testdatensatz: Vier Universitäten
`www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/data/`
- ▶ Testdaten sind in Unterordnern sortiert:
Kategorie/Ort_der_Erhebung/VolleURL
- ▶ Beispiel:
 - ▼   > course
 - ▼   > cornell
 -  `http://cs.cornell.edu/Info/Courses/Current/CS415/CS414.html`
- ▶ Zur einfacheren Verarbeitung soll die Ordnerhierarchie entfallen und die Informationen in den Dateinamen übernommen werden.

Aufgabe 1:

Bash-Skript zur Neusortierung



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Das Namensschema der Dateien soll lauten:
TRAIN|TEST ##### Kategorie ##### VolleURL
 Trenner Trenner