Diplomarbeitspräsentationen der Fakultät für Informatik

Analyzing, Labeling and Interacting with SOMs for Knowledge Management

Angela Roiger

Technische Universität Wien
Institut für Softwaretechnik und interaktive Systeme
Arbeitsbereich: Information & Software Engineering Group
Betreuer: Ao.Univ.-Prof. Dr. Andreas Rauber

FAKULTÄT FÜR !INFORMATIK

Diplomstudium:
Informatik

Zielsetzung

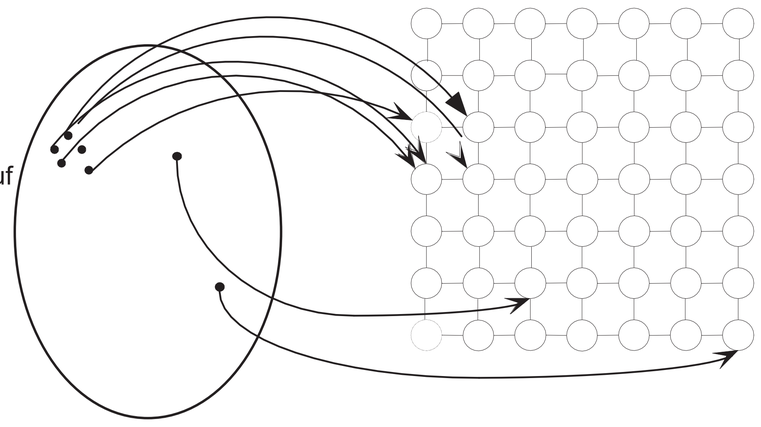
Das Ziel der Diplomarbeit ist die bestehende SOM Toolbox (Software für SOMs) durch Clustering zu erweitern und die entstehenden Cluster grafisch darstellen und zu beschriften. Insbesondere soll das Erkunden der Clusterstrukturen interaktiv möglich sein und die Beschriftungen manuell verbessert werden können.

Self-Organizing Map (SOM)

2-dimensionales neuronales Netz

Mapping von mehrdimensionalen Daten auf 2-dimensionale Fläche

Topologieerhaltend

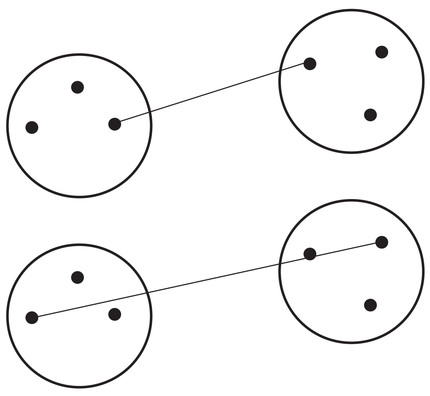


Clustering Methoden

Bottom-up: Iteratives Zusammenfügen der jeweils nähesten Cluster. Drei Methoden zur Bestimmung der "nähesten Cluster" implementiert:

Single linkage (kürzest mögliche Distanz zweier Clusterelemente)

Complete Linkage (längste mögliche Distanz zweier Clusterelemente)



Ward's Linkage (Anstieg der Varianz, wenn die beiden Cluster zu einem vereint werden)

$D = V(AB) - [V(A)+V(B)]$

Cluster darstellen

Grenzlinien einzeichnen

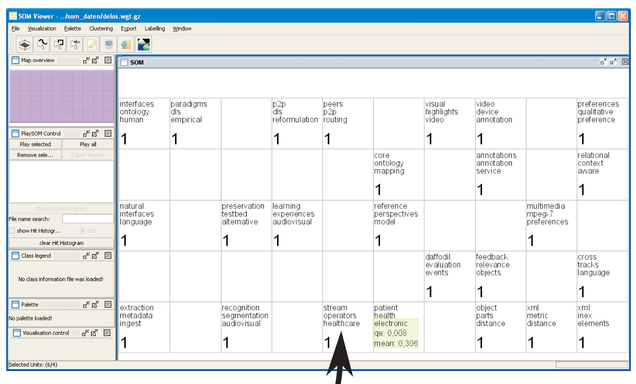
Cluster einfärben

Cluster beschriften

Text (anhand bestehender Unit Labels)

Größe (abhängig von der Kartenbreite)

Position (Schwerpunkt oder Mittelpunkt des umgebenden Rechtecks)



Unit Labels

Textdaten: Wörter, die in den enthaltenen Dokumenten am häufigsten vorkommen
Bei anderen Daten: die Bezeichnungen der Merkmale welche alle Daten auf einer Unit am ähnlichsten haben.

Bearbeiten der Beschriftungen

Text (auch mehrzeilig)

Schriftgröße

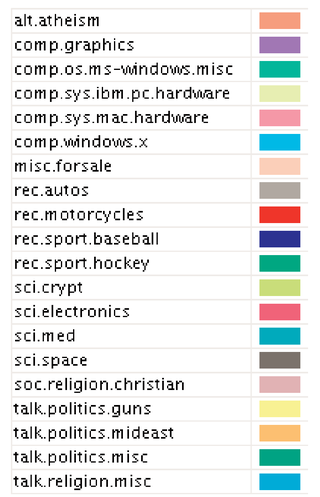
Platzierung (mittels drag and drop)

Farbe (hervorheben / abheben vom Hintergrund)

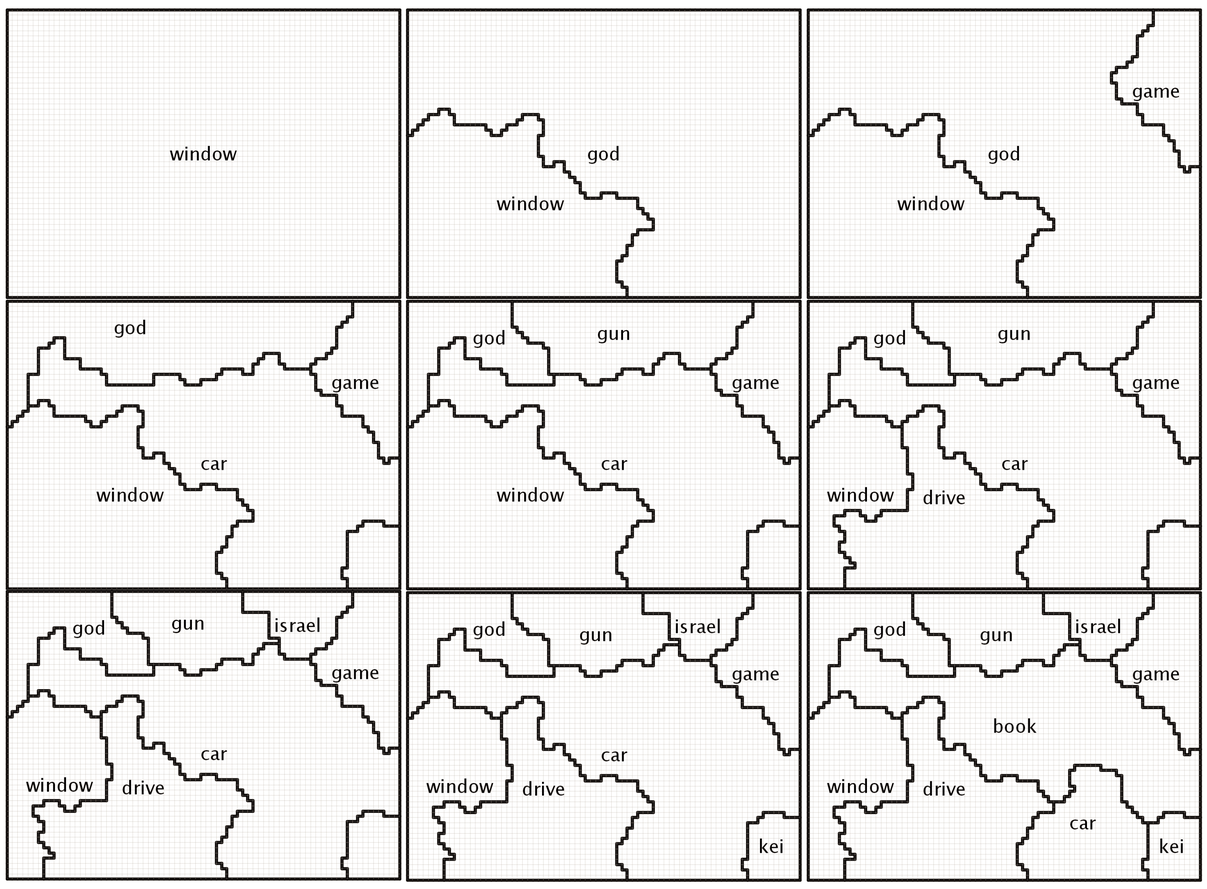
Rotation

nicht anzeigen

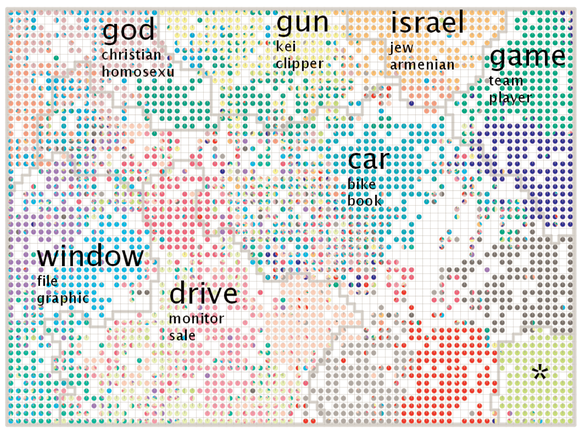
SOM von je 1.000 Postings aus 20 Newsgroups (Klasseninformation)



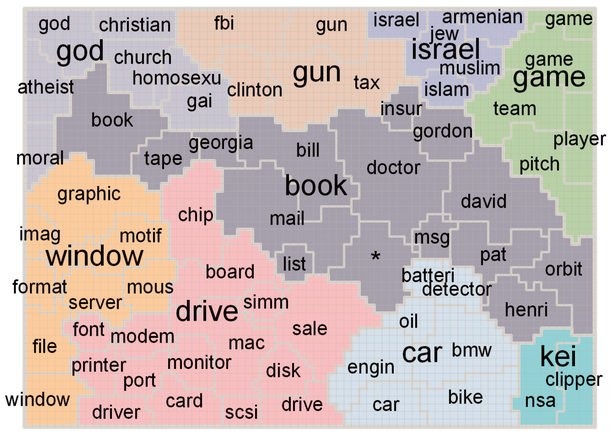
Farbzuordnung zu Klassen



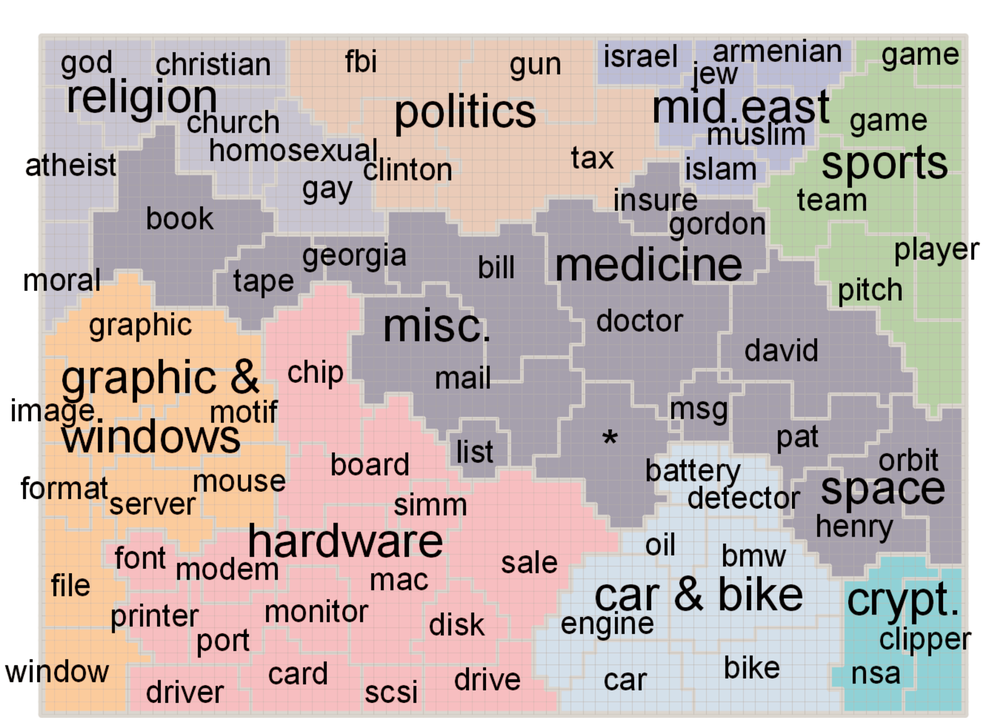
9 Clustering Schritte (in umgekehrter Reihenfolge) mit Beschriftung



7 Cluster mit je 3 Beschriftungen
+ Klasseninformation



9 große und 67 kleine Cluster



Editierte Karte

Kontakt: angela@roiger.at