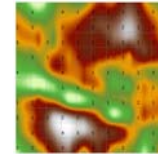
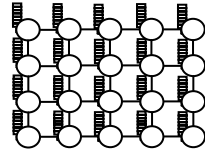


Evaluierung von SOMs

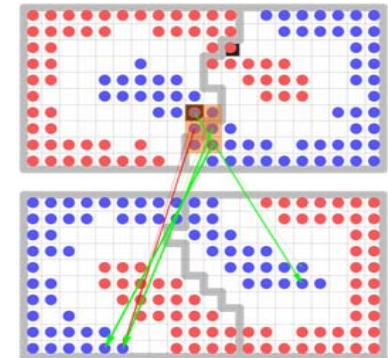
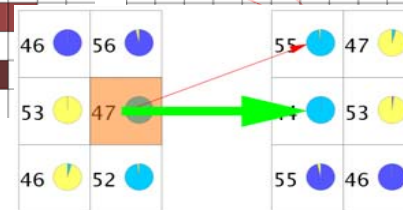
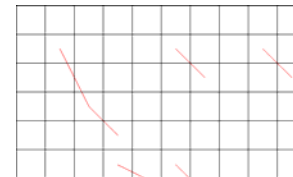
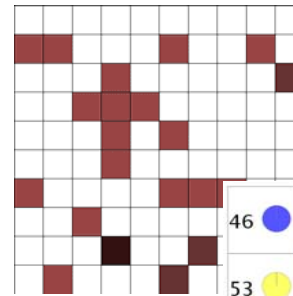
VU Selbst-Organisierende Systeme
Andreas Rauber

<http://www.ifs.tuwien.ac.at/~andi>

- SOM Basics



- Visualisierungen

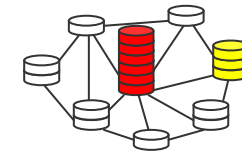
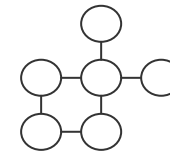
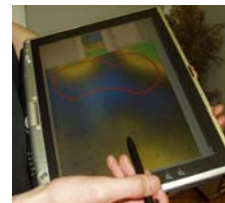


- SOM Qualitätsmaße

- SOM Comparison

- Verwandte Verfahren & Architekturen

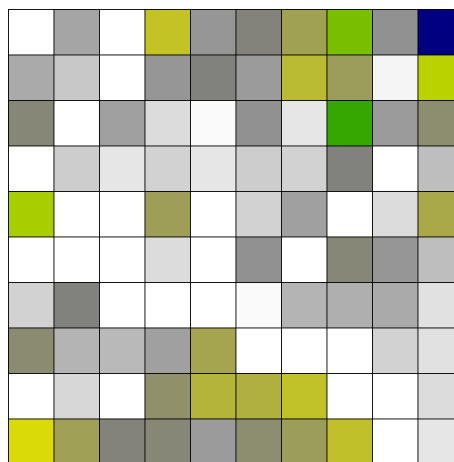
- Applikationen



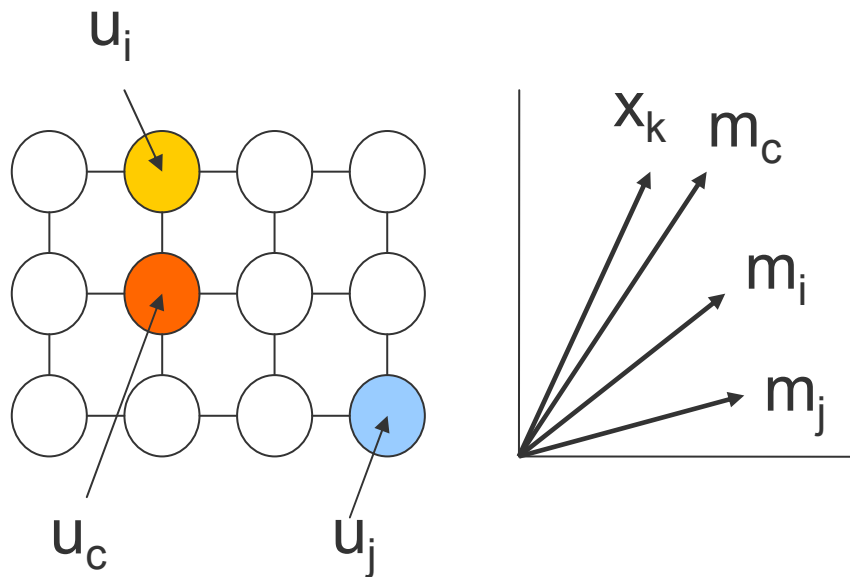
- How to evaluate the quality of a SOM?
- Compare which one is better – quantitatively, as opposed to qualitatively using the comparison visualization?
- SOM performs 2 tasks
 - vector quantization
 - vector projection / topology-preserving mapping
- Quality measures for these two characteristics
- plus: measure cluster characteristics

- Quantization Error
- Calculate distance between individual data vectors and the winner
- 4 measures
 - **qe**: quantization error for each unit: sum over all mapping distances
 - **QE**: quantization error for a map: sum over all qe
 - **mqe**: mean quantization error for each unit: mean of all mapping distances for each unit (i.e. qe / #of mapped vectors)
 - **MQE**: mean quantization error for a map: mean over all mqe's

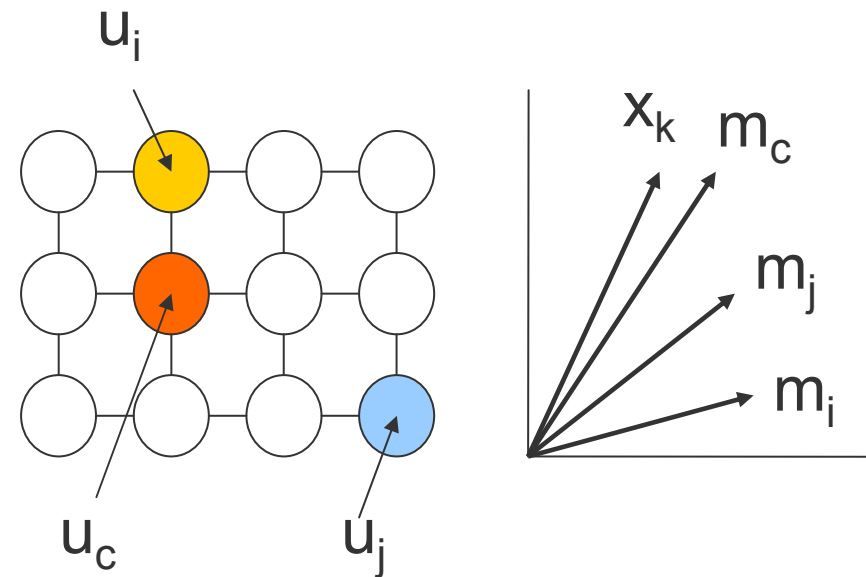
- Difference between these measures?
 - when is qe high but mqe low for a unit?
 - when do you need to compare 2 maps using the MQE rather than the QE?
 - why does it not make sense to always pick the map with the lowest QE or MQE?



- Topology preserving mapping:



(a) ordered mapping



(a) disordered mapping

- Measure, in how far (and where) topology is violated
- Principle: measure if / how often / how strongly two data items, that are close to each other in input space, are not close to each other in output space
- Several approaches:
 - visually: neighborhood graphs: long lines connect neighboring data
 - topographic product
 - intrinsic distance
 - topographic error

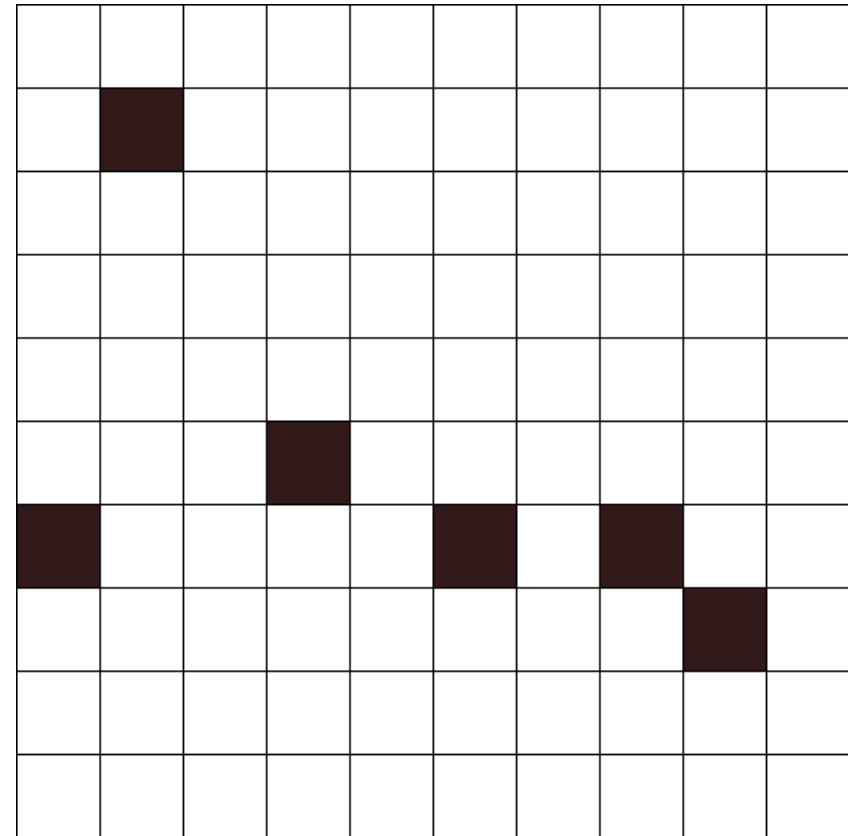
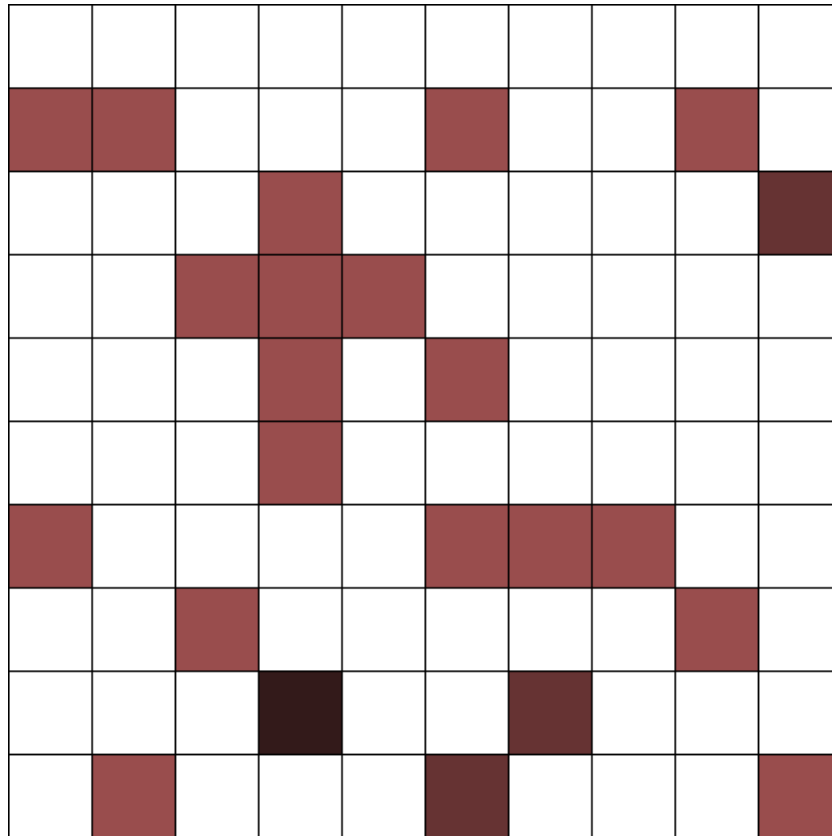
Topographic Error

- Percentage of data samples for which second-best matching unit is not adjacent (in output space) to the best-matching unit
- Usually calculated as quality measure for the entire map
- Can be visualized on a per-unit basis
- Usually uncovers topology violations
- Sometimes, such “topology violations” occur in dense cluster areas and are thus no errors

(Q: why?)

hint: dense areas on SOM stretched due to magnification factors, thus →?)

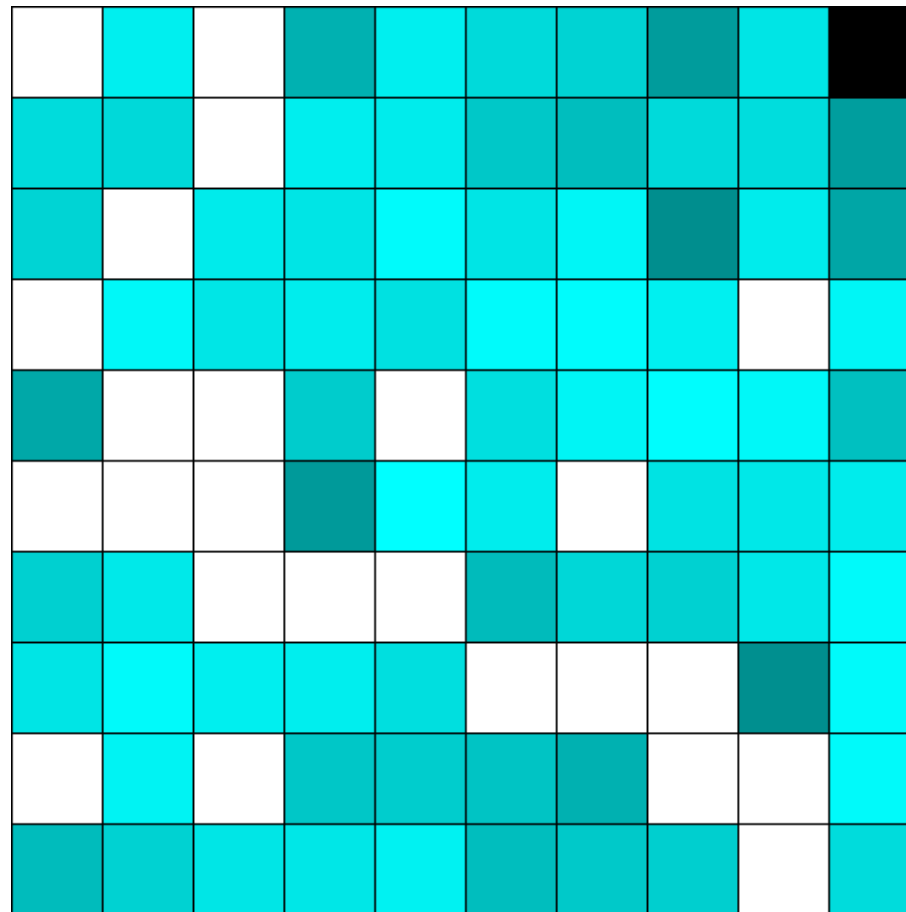
- Topographic Error: 4 and 8 units



Intrinsic Distance

- Combines aspects from Quantization Error and Topographic Error
- Compute BMU and 2nd-BMU for each data sample
- Calculate mapping distance from data to BMU in input space (quantization error)
- Add distance in input space along path (in output space via neighboring units) to 2nd-BMU
- Samuel Kaski and Krista Lagus. Comparing SelfOrganizing Maps. In: *Proceedings of the International Conference on Artificial Neural Networks (ICANN'96)*, pages 809–814. Springer, Berlin, 1996.

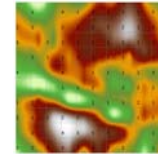
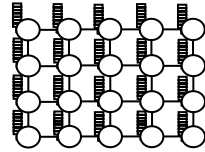
- Intrinsic Distance



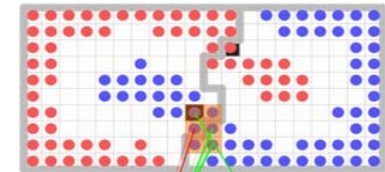
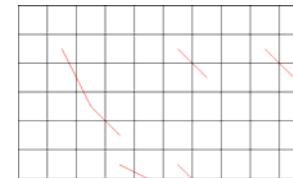
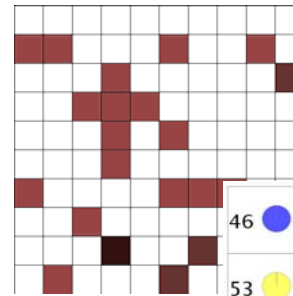


Outline

- SOM Basics

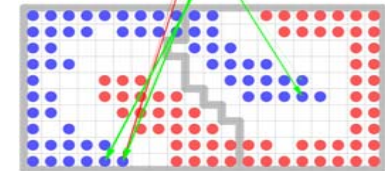
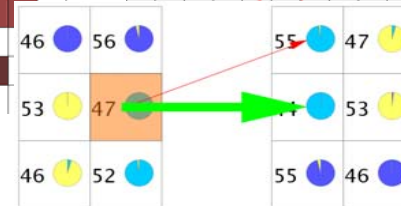


- Visualisierungen

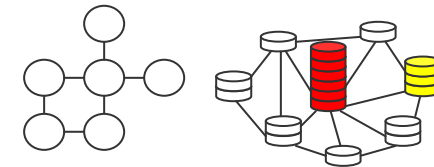


- SOM Qualitätsmaße

- SOM Comparison



- Verwandte Verfahren & Architekturen



- Applikationen



- SOM nicht deterministisch
 - Initialisierung
 - zufällige Reihenfolge der Vektorpräsentation
 - gleiche / unterschiedliche Größe
- Wie lassen sich 2 Karten vergleichen?
 - Qualitätsmaße
 - Vergleich der Abbildung
- Verfahren zur Unterstützung des Vergleichs von 2 und mehr SOMs

- Auswirkungen unterschiedlicher SOM-Parameter testen
 - Mögliche Parameter: Nachbarschaftsradius, Lernrate, Größe der Karte, Trainingsdauer, Init, Random-Seed Vektorpräsentation, etc.
- Topologieverletzungen finden und Stabilität der SOM-Projektion einschätzen
- 4 Verfahren:
 - data / cluster comparison
 - data / cluster shifts

- Rudolf Mayer, Robert Neumayer, Doris Baum, and Andreas Rauber.

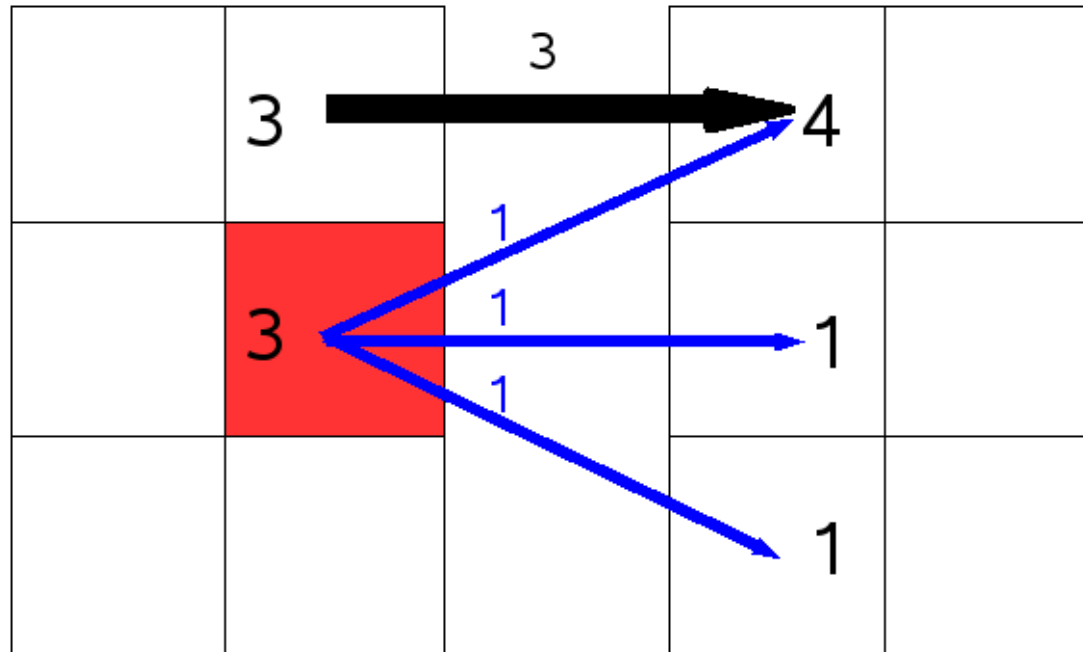
Analytic Comparison of Self-Organising Maps.

In *Proceedings of the [7th International Workshop on Self-Organizing Maps \(WSOM'09\)](#)*, St. Augustine, FL, USA, June 8 - 10 2009. LNCS 5629, pp 182-190, Springer.



SOM Comparison Visualisierung

- Vergleicht “beliebig viele” SOMs miteinander:
 - eine Haupt-SOM, die angezeigt wird und
 - eine oder mehrere Vergleichs-SOMs über die ein Durchschnitt gebildet wird
- Für jede Unit auf der Haupt-SOM: Einfärbung korrespondierend zur mittleren paarweisen Distanz der Datenvektoren im Ausgaberaum (Karte) in den anderen SOMs.

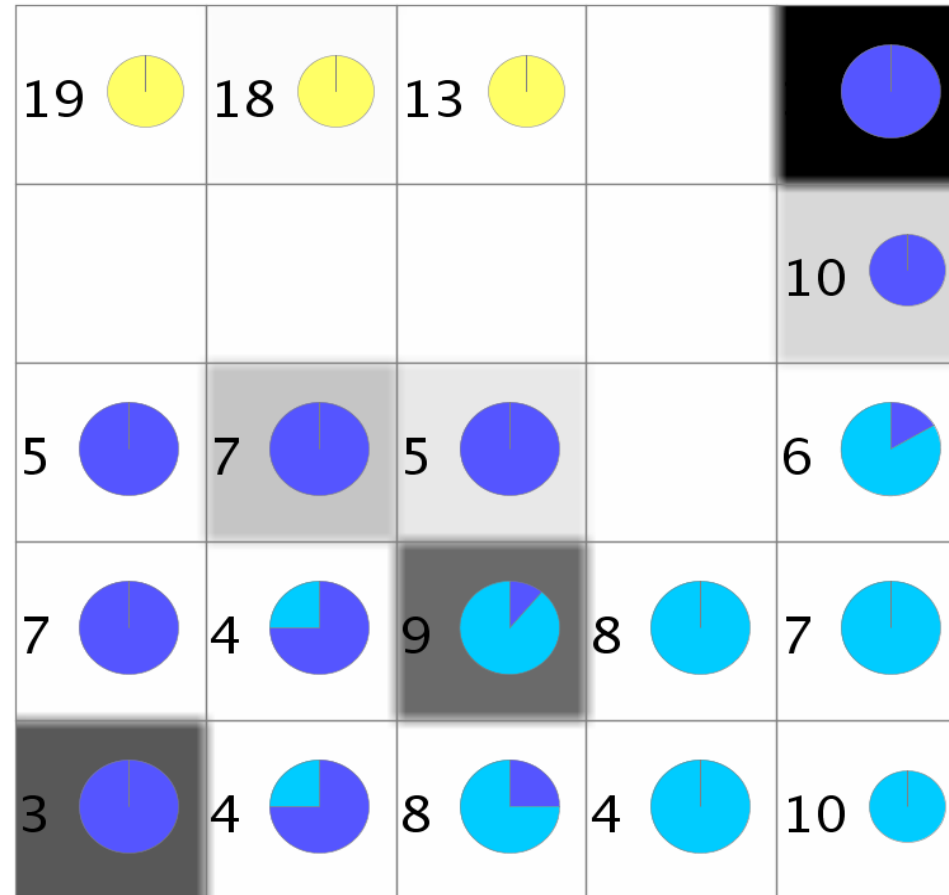


Haupt-SOM

Vergleichs-SOM

Distanzwerte: 0 und $4/3 = 1.3$

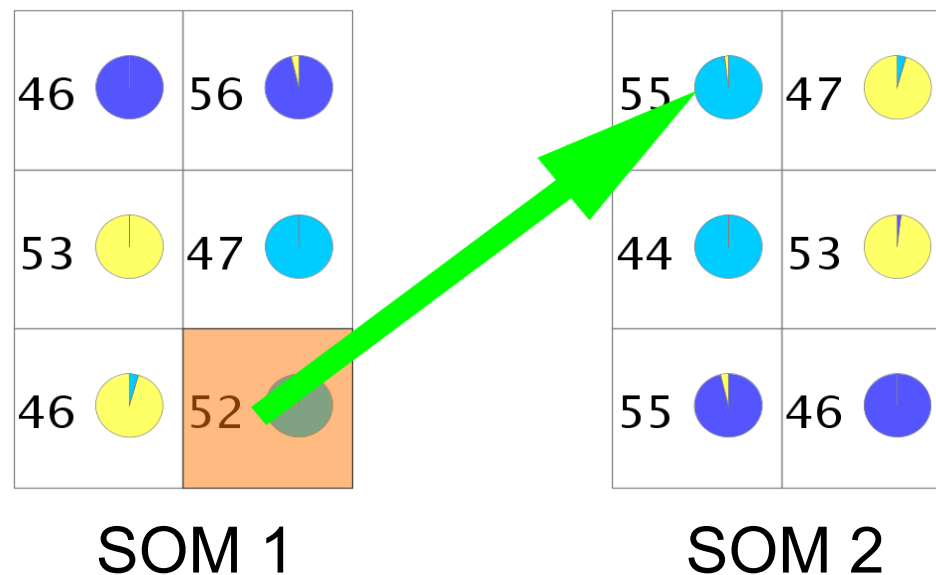
Comparison Visualisierung



Variationen:

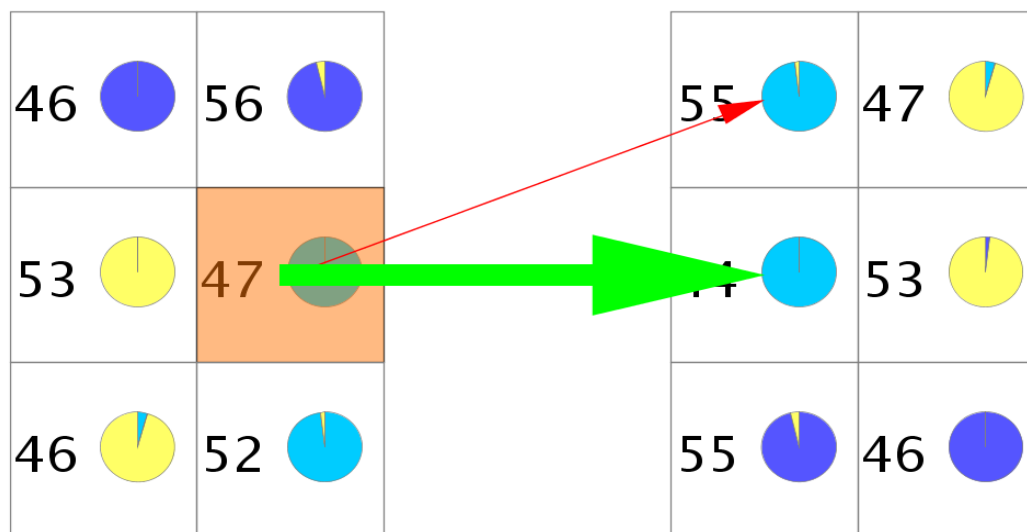
- Clusterdistanz statt euklidische Distanz:
 - Abstand zwischen Vektoren entspricht Abstand zwischen Clustern, in denen sie liegen
 - Abstand zwischen Clustern wie Single Linkage (Minimaler Abstand zwischen zwei Units der Cluster)
- Varianz statt Durchschnitt / Erwartungswert
- Schwellwert: Wenn einzelne paarweise Distanzen kleiner als Schwellwert, dann zählen sie nicht zum Durchschnitt

- Vergleicht zwei SOMs miteinander
- Zeigt für Datenvektoren von SOM 1 wo sie jeweils auf SOM 2 zu liegen kommen: „Shift“
- Bsp.: nach Neutraining feststellen, wohin sich die Daten verschoben haben

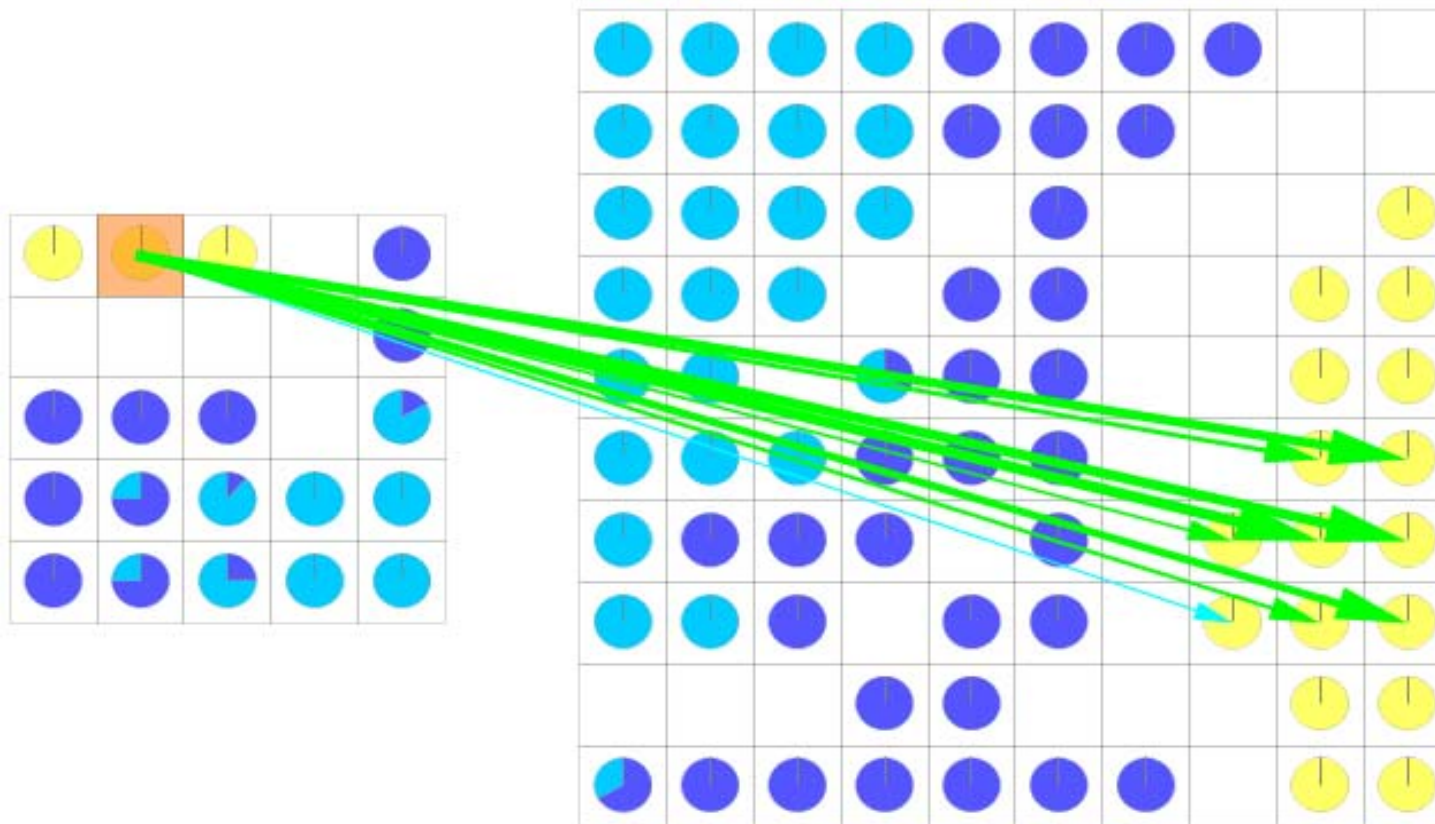


- Beim Vergleich der Positionen eines Daten-vektors werden die Datenvektor-Nachbarn auf der SOM betrachtet:
 - sind viele der Nachbarn in SOM 1 auch Nachbarn in SOM 2: stable shift
 - kommt ein Vektor in SOM 2 in einer anderen Umgebung zu liegen als seine Nachbarn aus SOM 1: outlier shift
- Größe der Nachbarschaft und Anzahl der nötigen Nachbarn einstellbar
- Erlaubt Aussagen über die Stabilität des Clusterings

- Grün: stable; Rot: outlier; Nachbarschaftsgröße: 0
- Dicke der Pfeile proportional zur Anzahl der gleich bleibenden Nachbarn

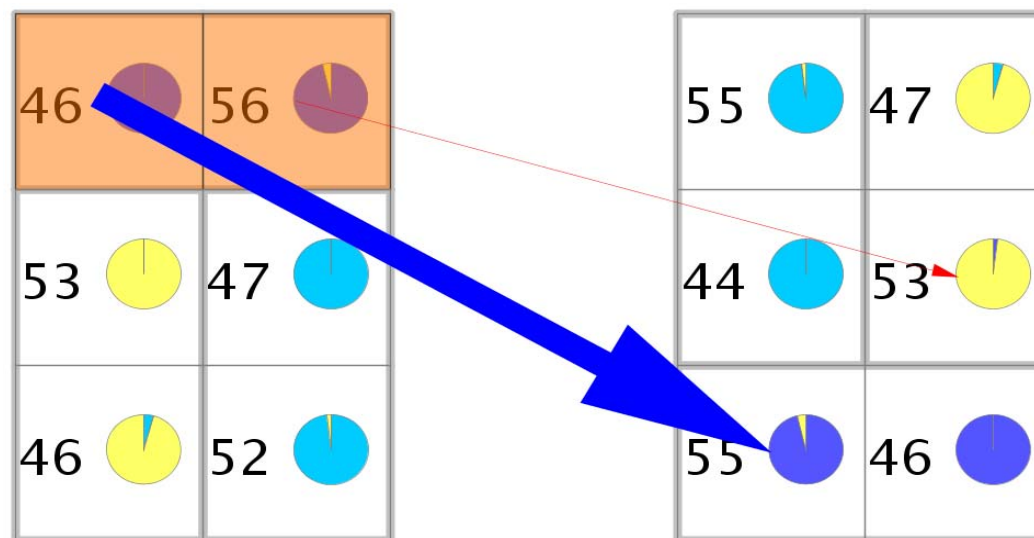


- Data Shifts: kleine Karte auf große Karte

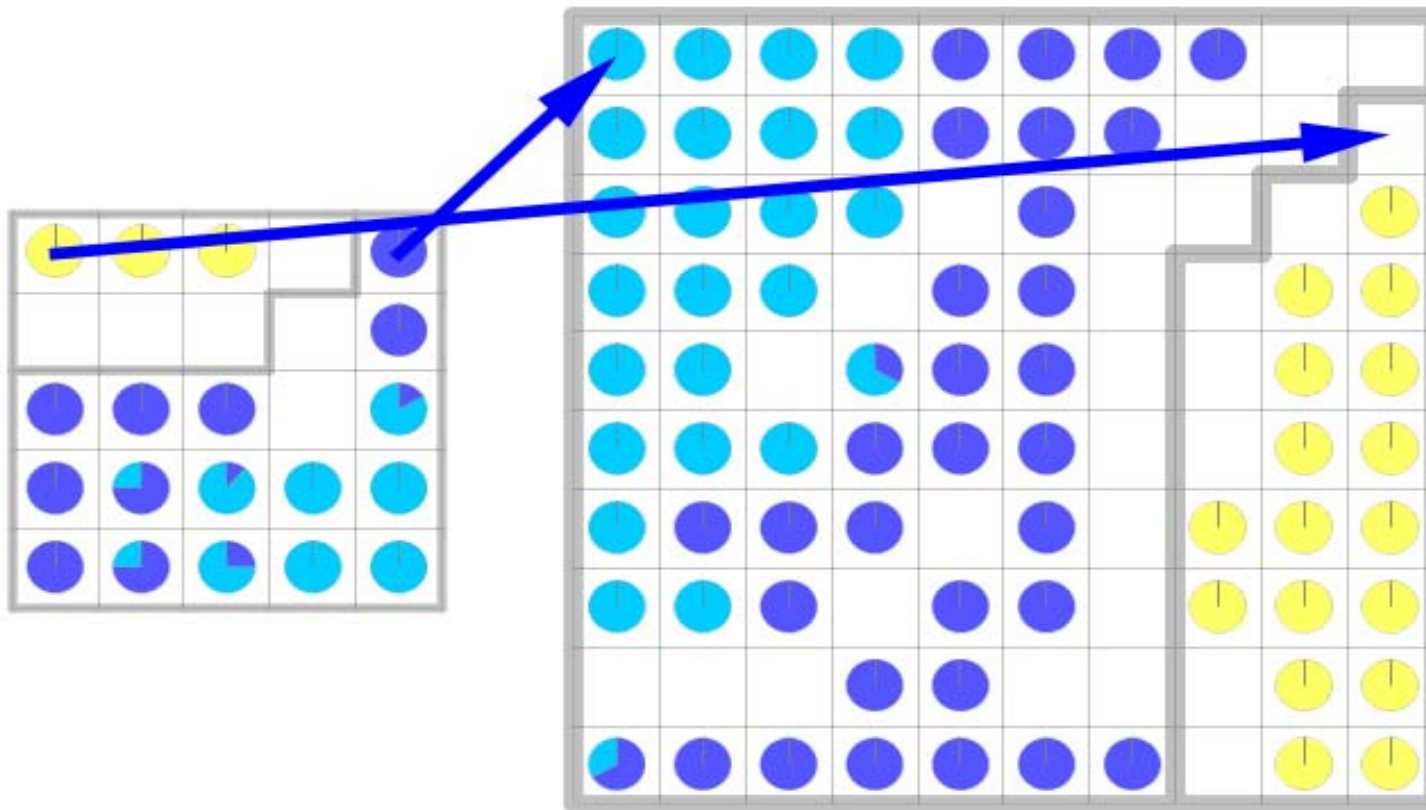


- Ist ein Datenvektor auf SOM 2 im gleichen Cluster gelandet wie in SOM 1?
- Clustering auf beiden SOMs mit einstellbarer Anzahl von Clustern
- Cluster in SOM 1 werden mit Clustern in SOM 2 identifiziert (nach Übereinstimmung von Datenvektoren in beiden Clustern)
- Liegt ein Datenvektor in miteinander identifizierten Clustern: stable shift; sonst: outlier

- Blauer Pfeil: miteinander identifizierte Cluster
- Dicke der Pfeile proportional zur Übereinstimmung zwischen den beiden Clustern

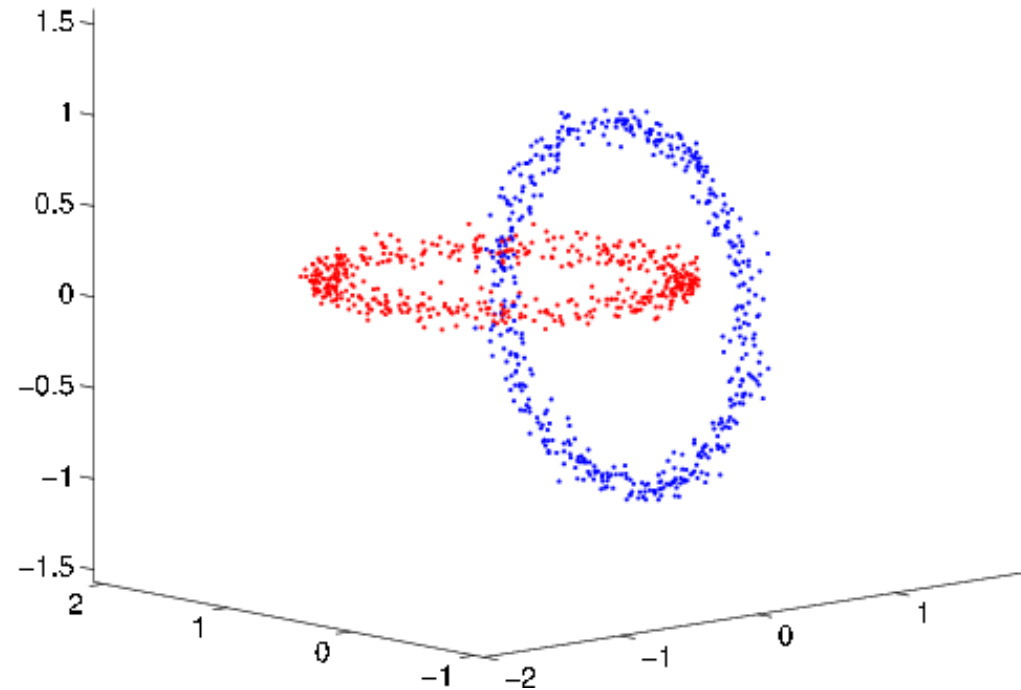


- Cluster Shifts: kleine Karte auf große Karte
- Feststellen, wo Cluster auf neuer Karte liegen
- Besonders bei fehlender Klassen-Information hilfreich!

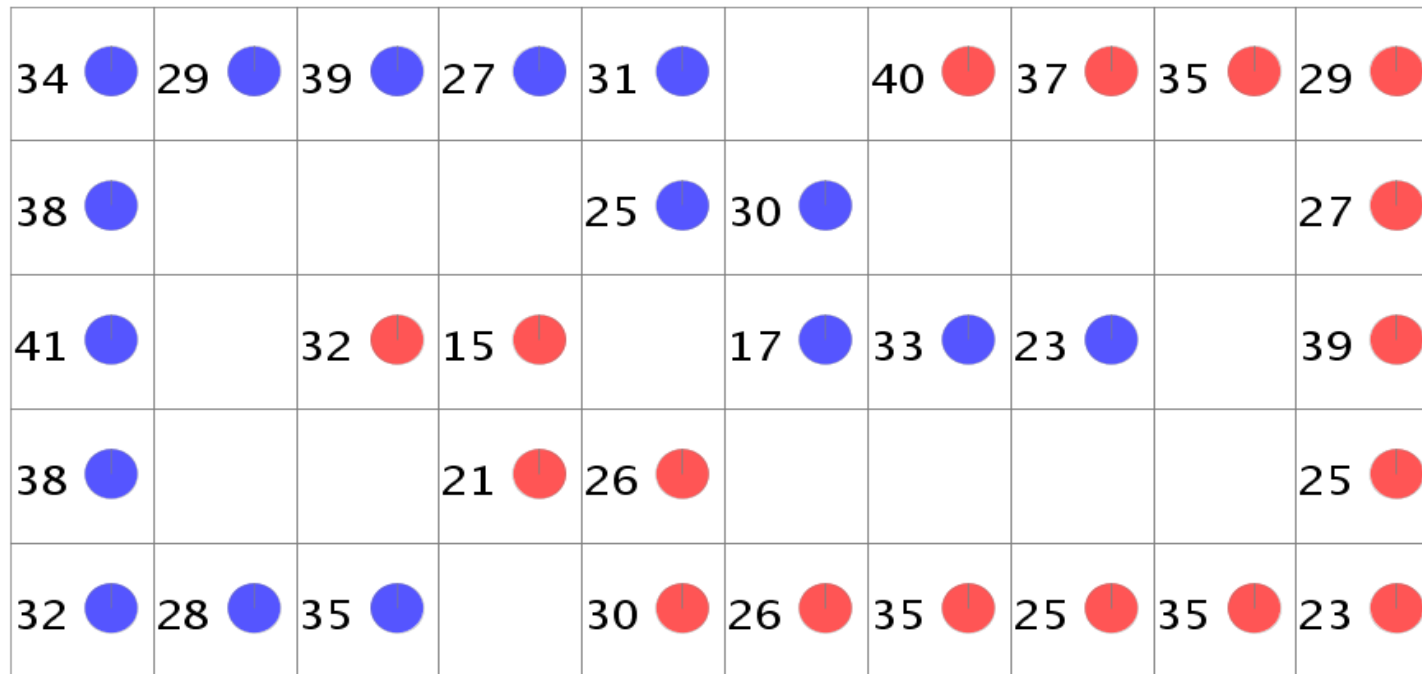


verschränkte Ringe

- 3-dimensionale Daten
- Projektion auf 2D-Grid
- Topologie-verletzung!

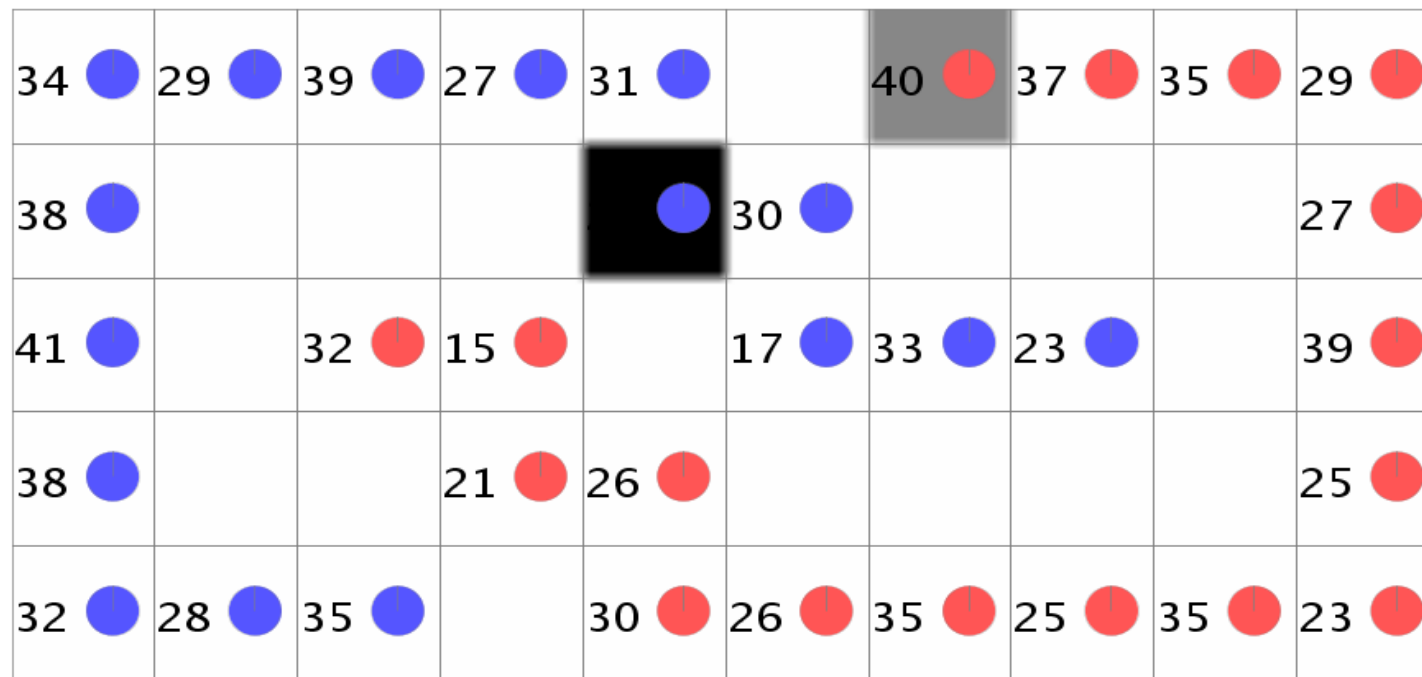


SOM Comparison



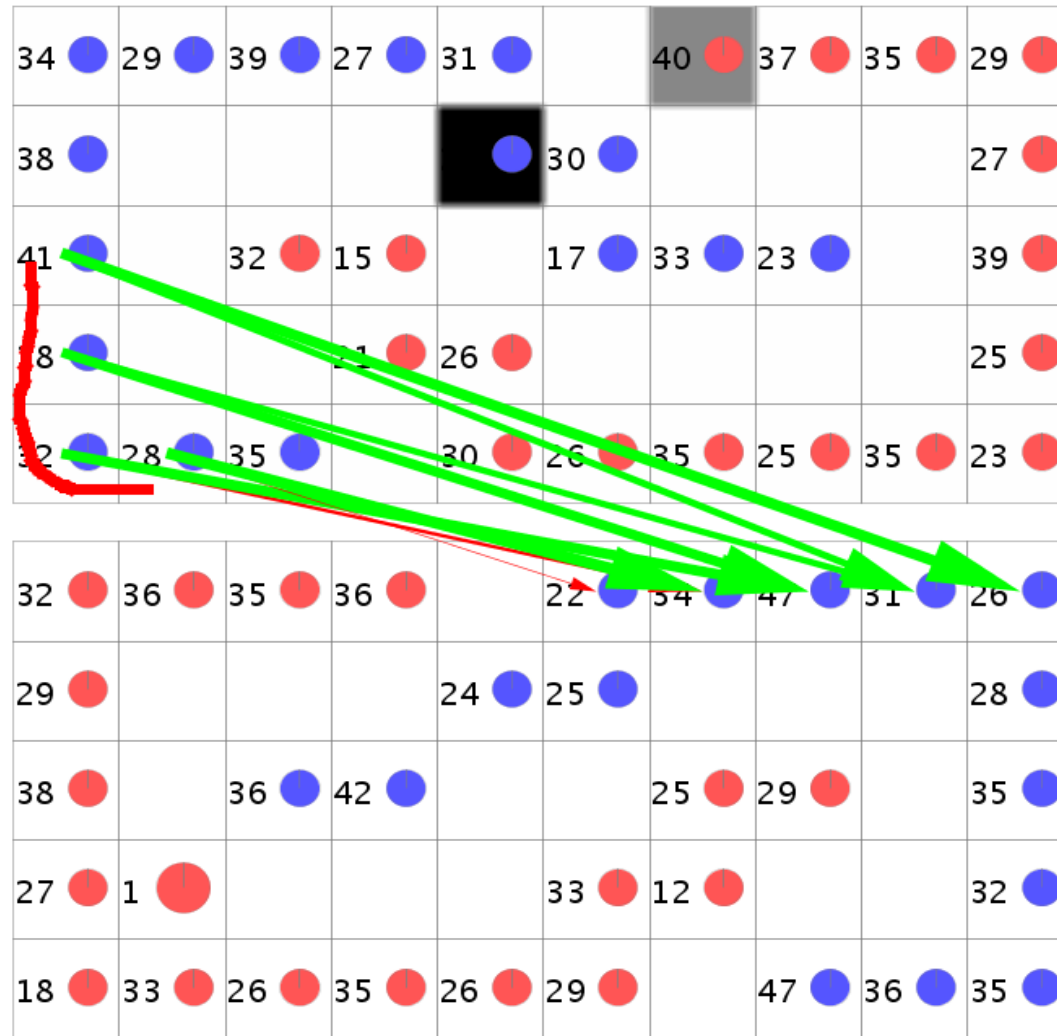
beliebige SOM: Ringe aufgebrochen

Comparison Visualisierung

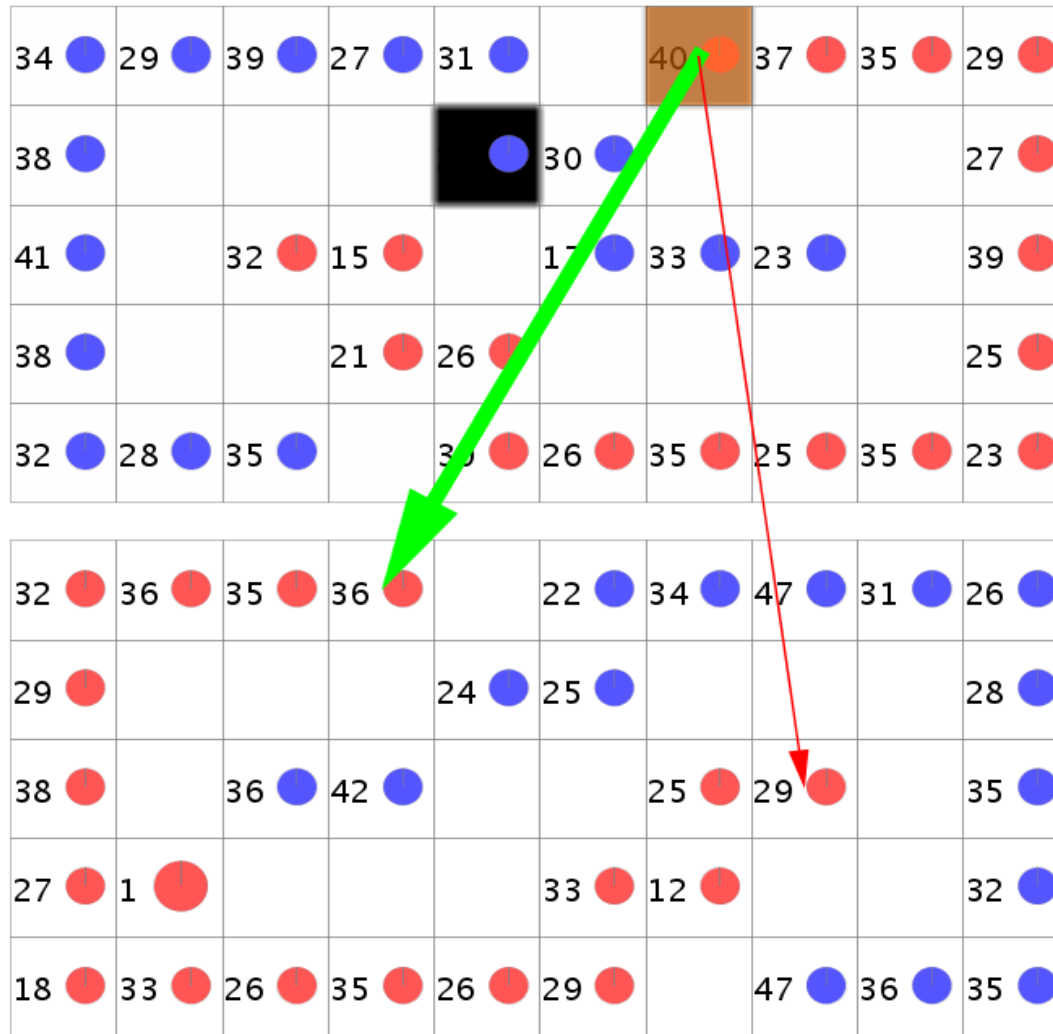


Schwellwert von 3: Nur paarweise Distanzen größer als 3 zählen → große Distanzen stärker gewichtet

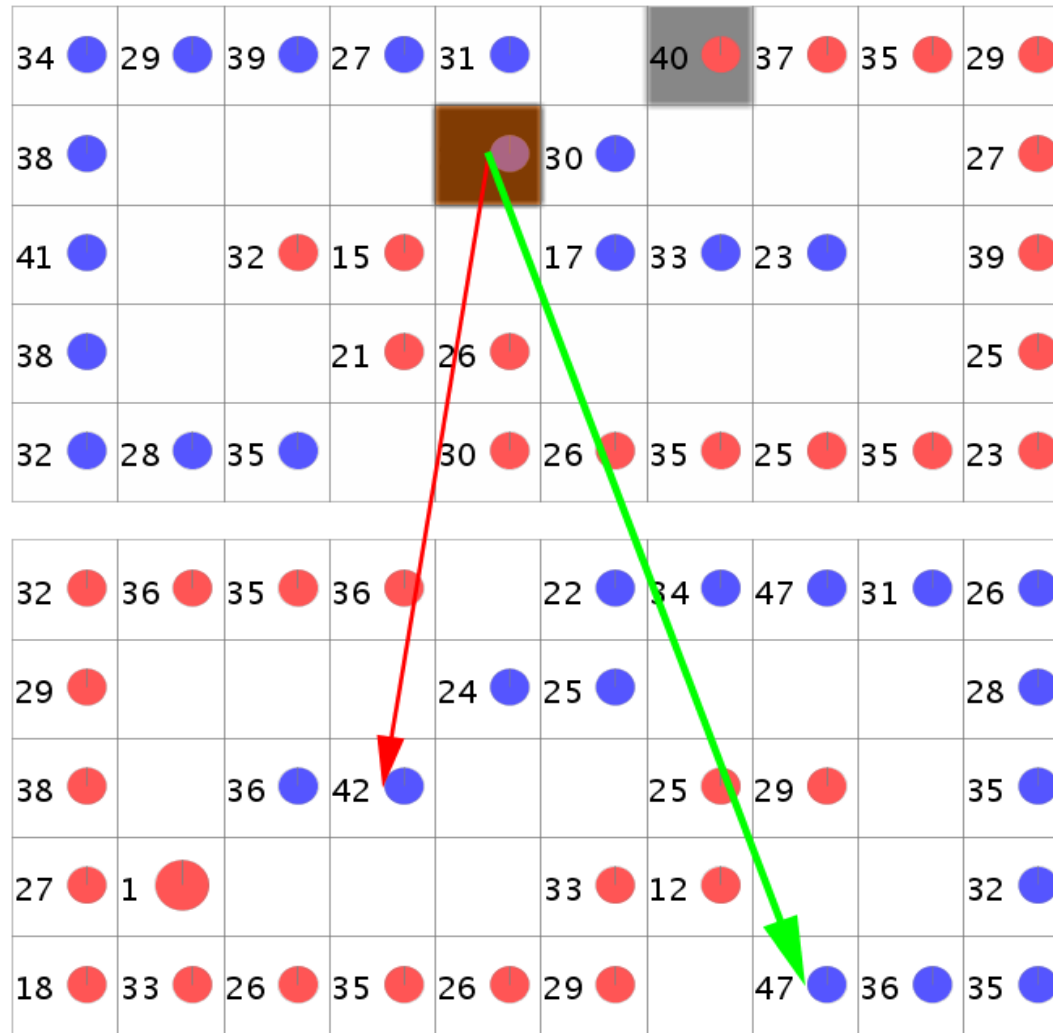
Data Shifts Visualisierung



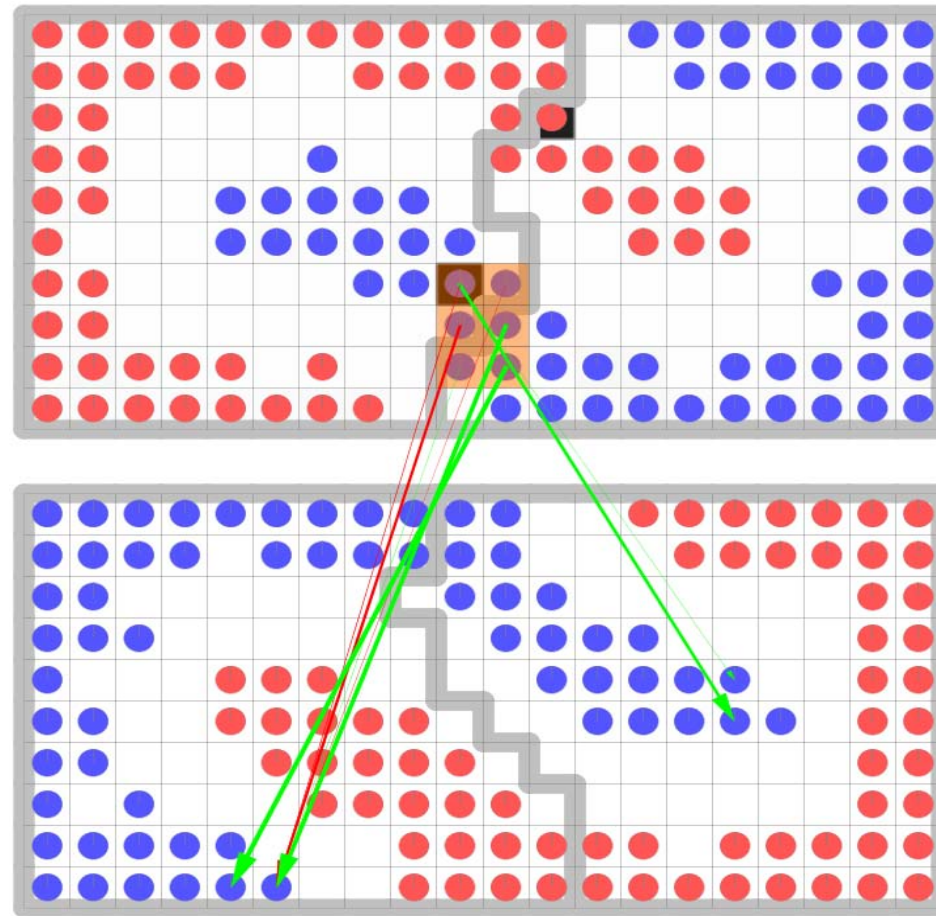
Data Shifts Visualisierung



Data Shifts Visualisierung



Cluster Shifts Visualisierung







Anwendungen

- Prozess-Monitoring
- Explorative Datenanalyse
- Text-Mining: SOMLib
- Musikanalyse: SOMeJB



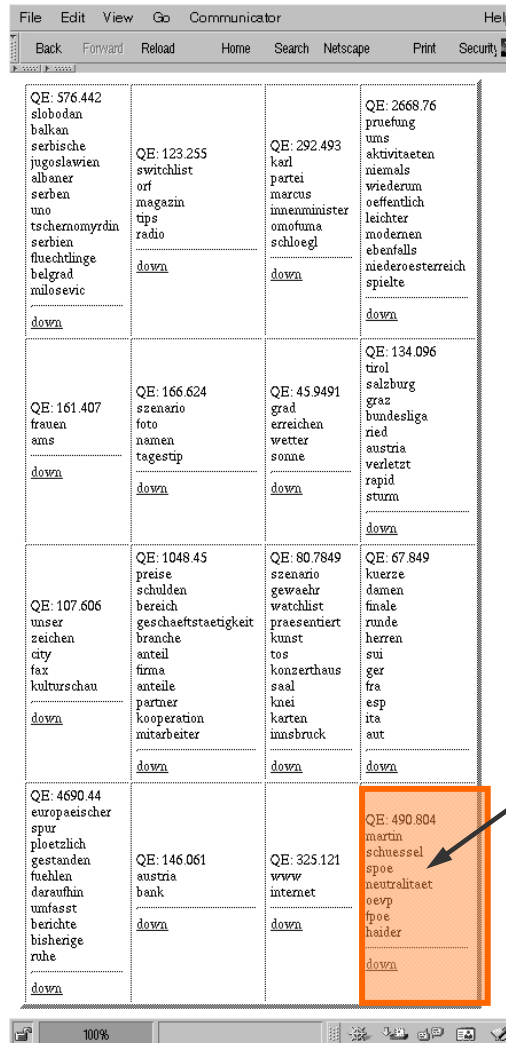
Growing Hierarchical SOM.....

■ Dokumentensammlung

- Nachrichtenartikel aus dem "Standard"
(2. Quartal 1999)
- 11.627 Artikel
- 3.799 Worte dienen zur Beschreibung der Artikel

Growing Hierarchical SOM

- oberste Ebene

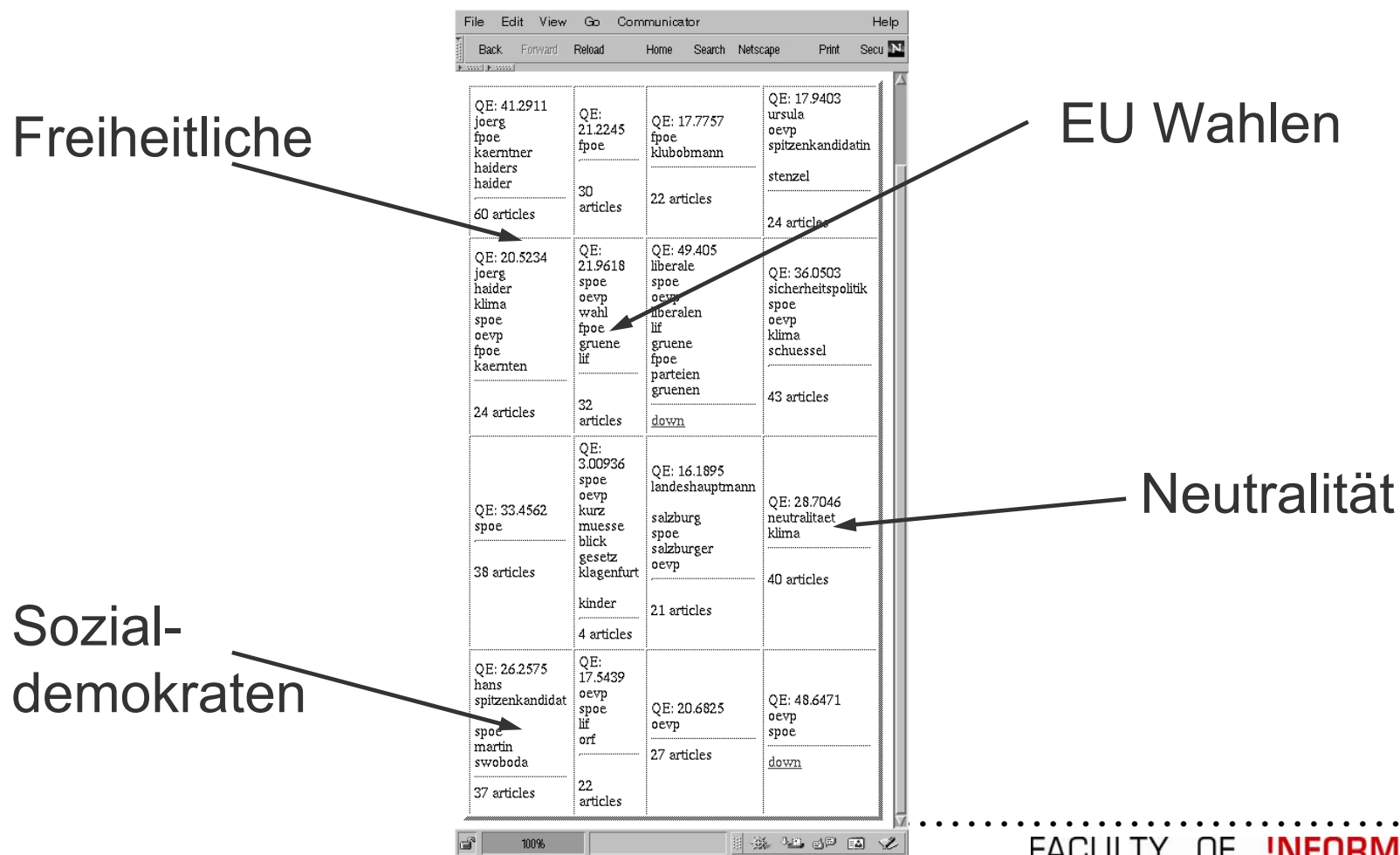


<p>QE: 576.442 slobodan balkan serbische jugoslawien albaner serben uno tschernomyrdin serbien fluechtlinge belgrad mllosevic</p> <p>down</p>	<p>QE: 123.255 switchlist orf magazin tips radio</p> <p>down</p>	<p>QE: 292.493 karl partei marcus innenminister omofuna schloegl</p> <p>down</p>	<p>QE: 2668.76 pruefung ums aktivitaeten niemals wiederum oeffentlich leichter modernen ebenfalls niederosterreich spiele</p> <p>down</p>
<p>QE: 161.407 frauen ams</p> <p>down</p>	<p>QE: 166.624 szenario foto namen tagestip</p> <p>down</p>	<p>QE: 45.9491 grad erreichen wetter sonne</p> <p>down</p>	<p>QE: 134.096 tirol salzburg graz bundesliga ried austria verletzt rapid sturm</p> <p>down</p>
<p>QE: 107.606 unser zeichen city fax kulturschau</p> <p>down</p>	<p>QE: 1048.45 preise schulden bereich geschaefstaetigkeit branche anteil firma anteile partner kooperation mitarbeiter</p> <p>down</p>	<p>QE: 80.7849 szenario gewaehr watchlist praesentiert kunst tos konzertHaus saal knei karten innsbruck</p> <p>down</p>	<p>QE: 67.849 kuerze damen finale runde herren sui ger fra esp ita aut</p> <p>down</p>
<p>QE: 4690.44 europaeischer spur ploetlich gestanden fuehlen daraufhin umfasst berichte bisherige ruhe</p> <p>down</p>	<p>QE: 146.061 austria bank</p> <p>down</p>	<p>QE: 325.121 www internet</p> <p>down</p>	<p>QE: 490.804 martin schuessel spoe neutralitaet oevp ipoe haider</p> <p>down</p>

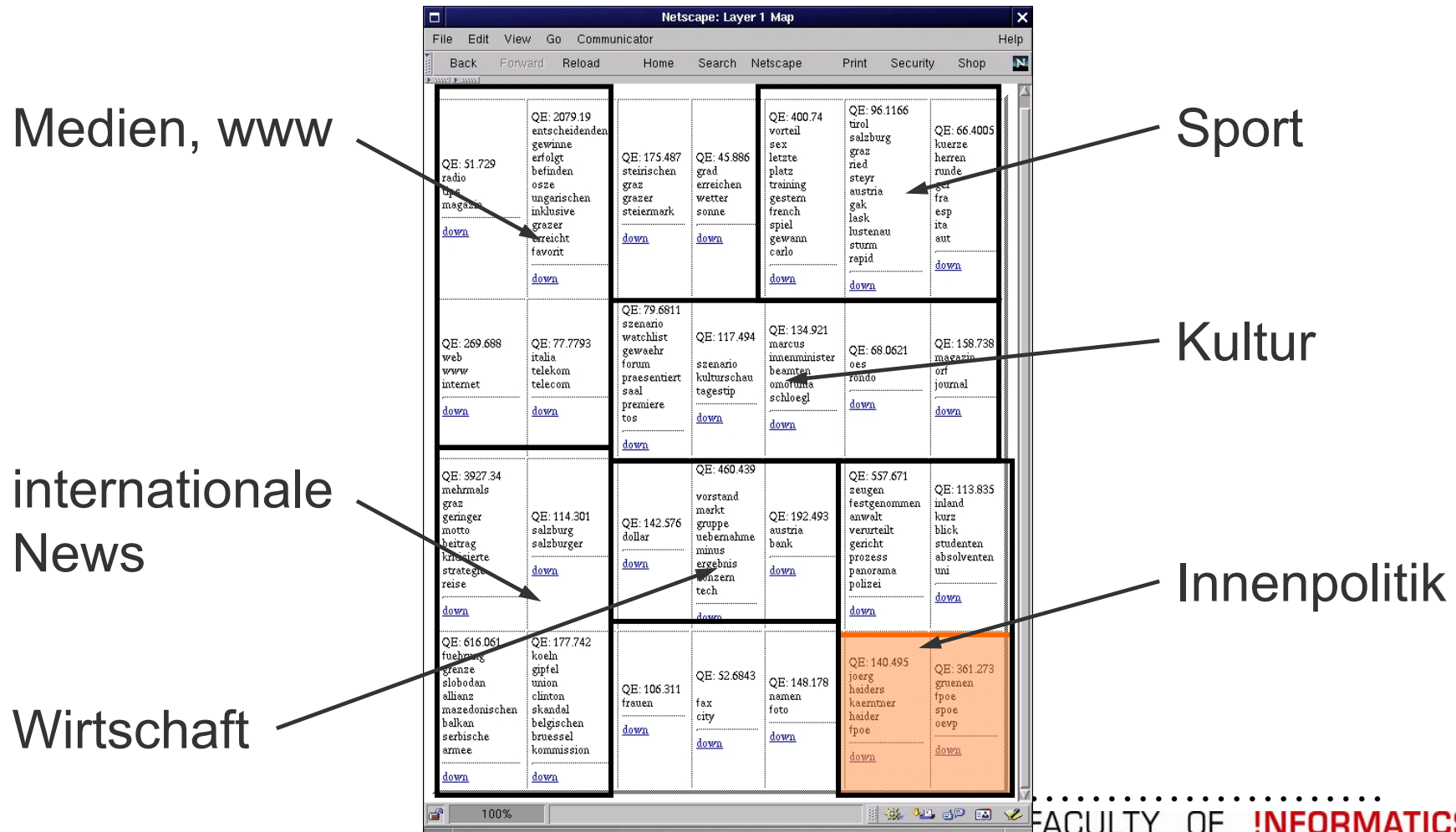
Innenpolitik

Growing Hierarchical SOM

- Karte der 2. Ebene

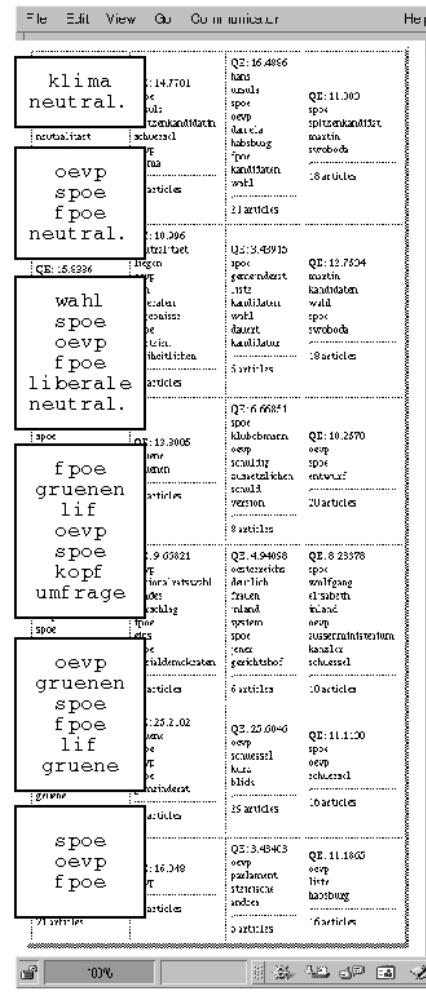
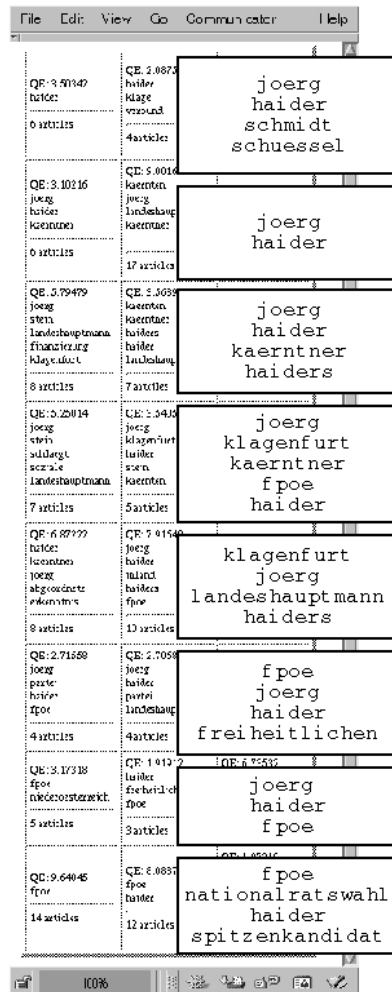


- oberste Ebene

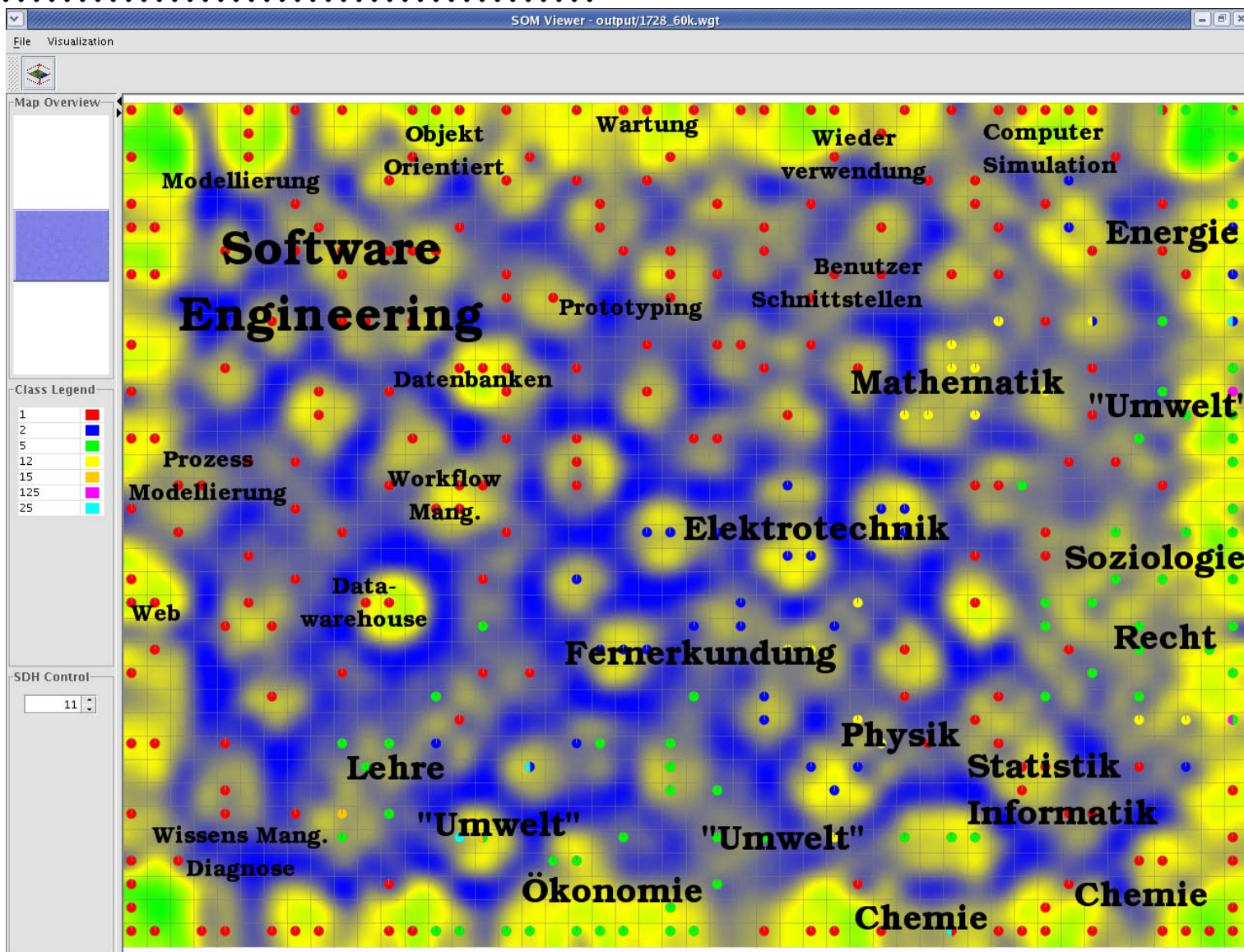


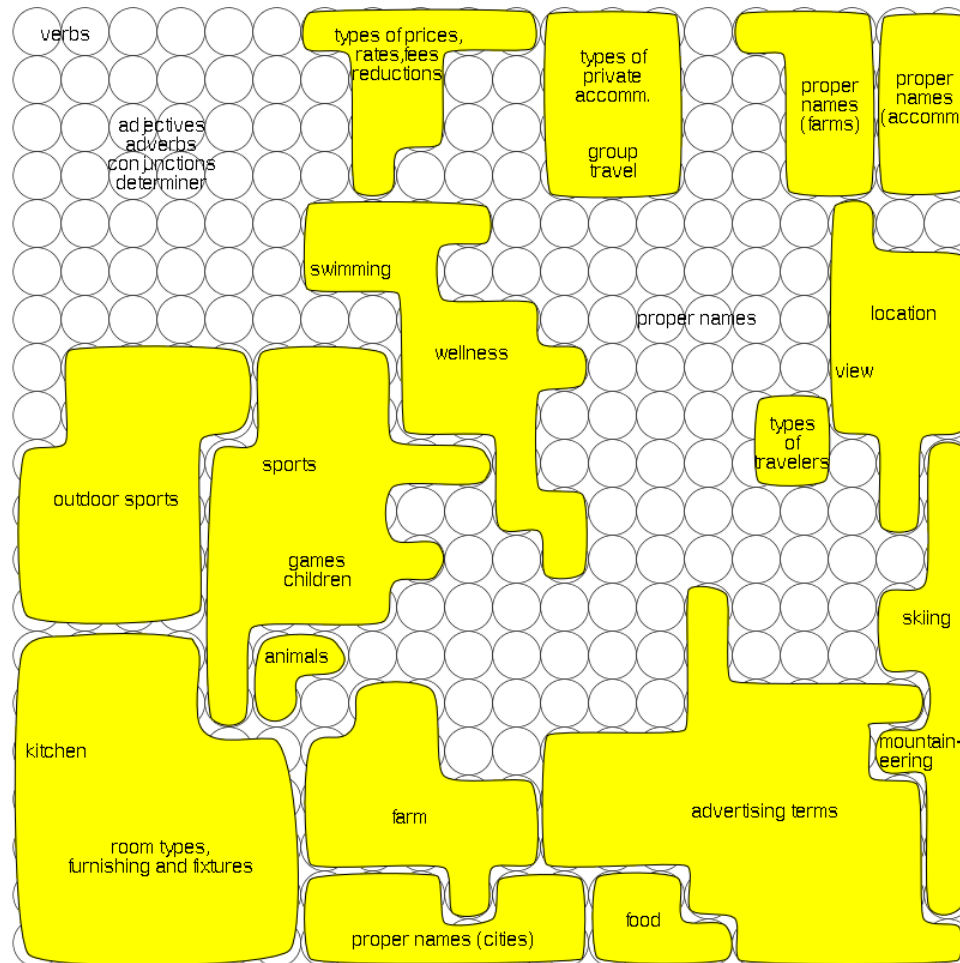
Growing Hierarchical SOM

- benachbarte Karten der 2. Ebene

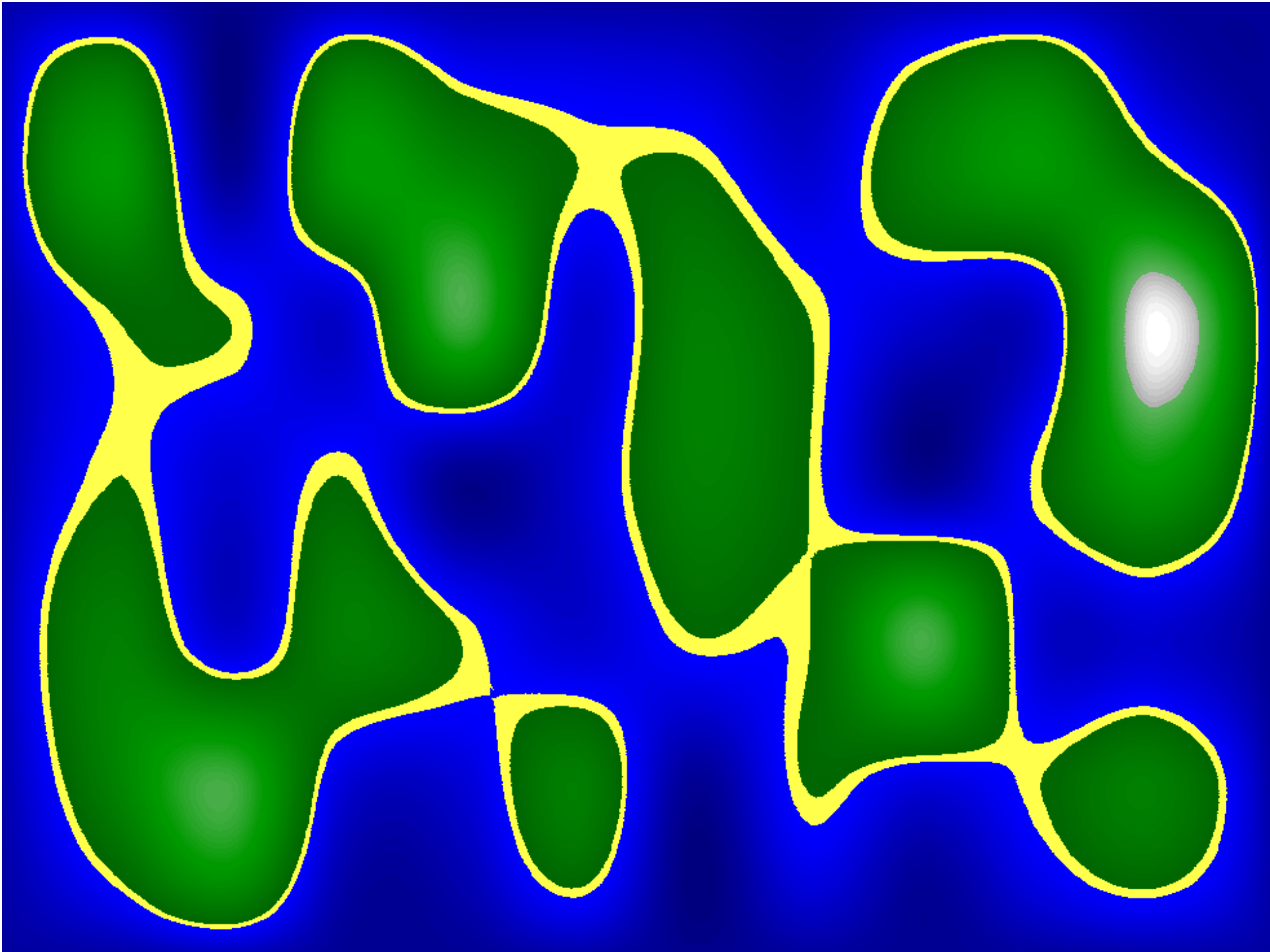


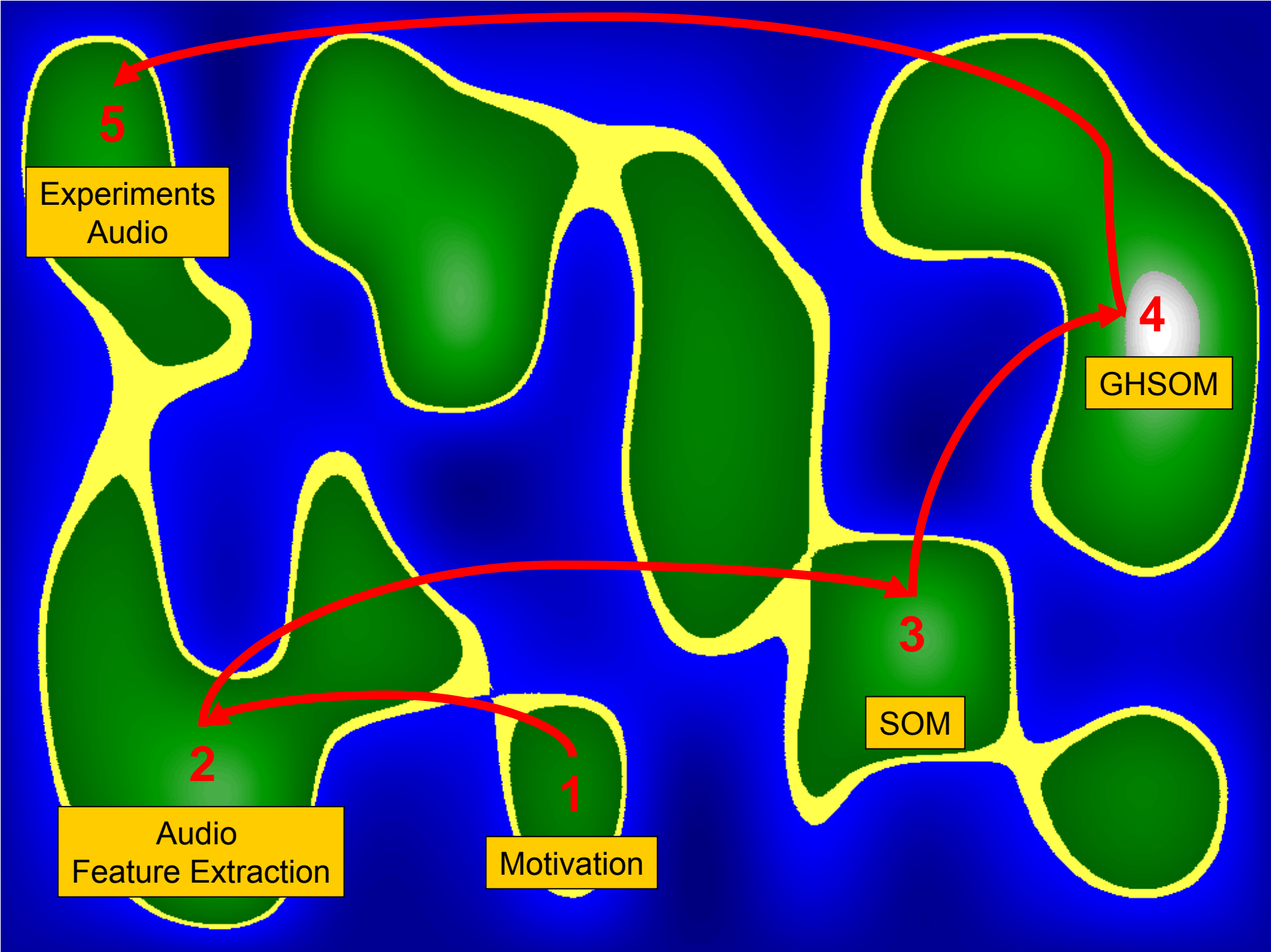
TU DOK Pilot Study





kochnische	bad
wanne	stockbett
sofa	doppelzimmern
badewanne	duche
waschraum	schlafraeume
doppelbett	zimmerausstattung
schlafmoeglichkeiten	dreibettzimmer
hotelzimmer	wohnschlafrum
essraum	schlafzimmer
kochecke	zimmer
duschen	fliesswasser
kinderzimmer	einbettzimmer
schlafraum	komfortzimmer
wohnschlafzimmer	doppelschlafzimmer





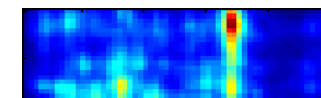
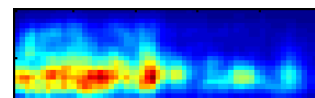


Music Case Studies

- Cluster music by perceived similarity ("genre")

- Music Features:

- analyzing frequency spectra
- Rhythm Patterns:
amplitude modulation in different frequency bands
psycho-acoustic transformations
1.440-dimensional vectors per song
- statistical spectrum descriptors (SDD),
Marsyas features,...



- Prototypes:

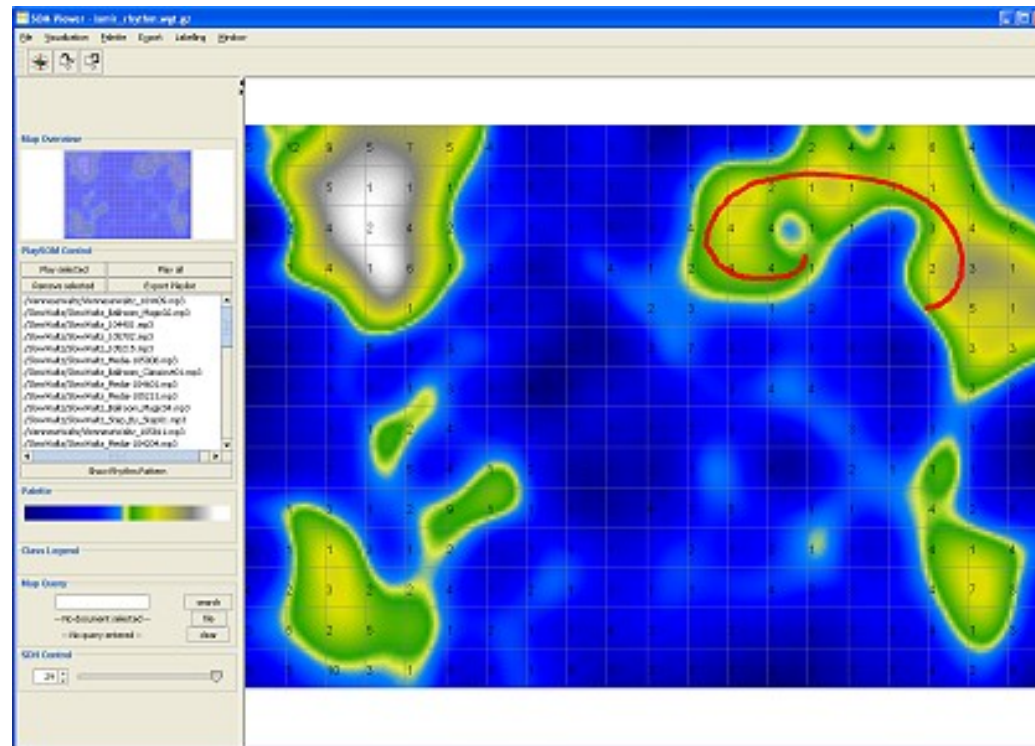
- PlaySOM for Desktop PC's
- PocketSOMPlayer for PDA's



PlaySOM - Playlist Selection

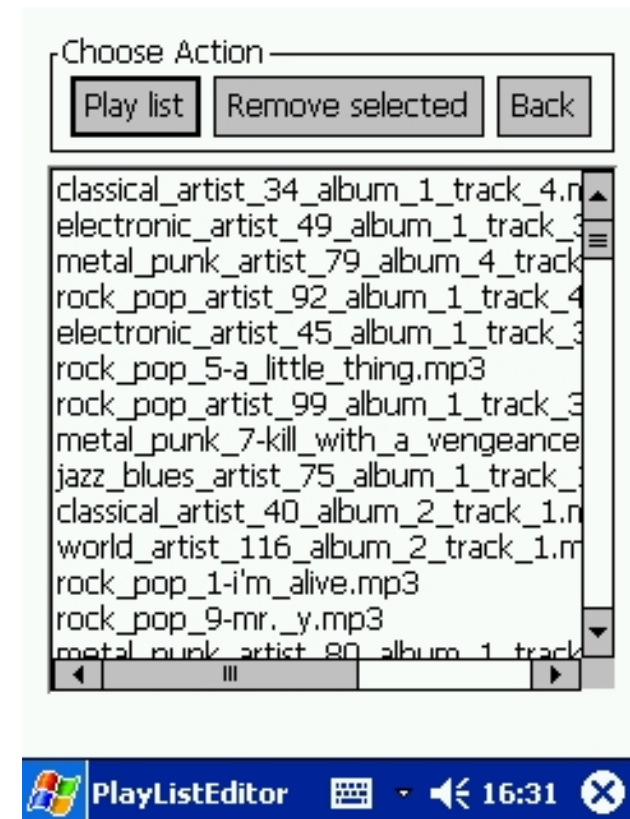


- Organizing Music
- Creating Playlists

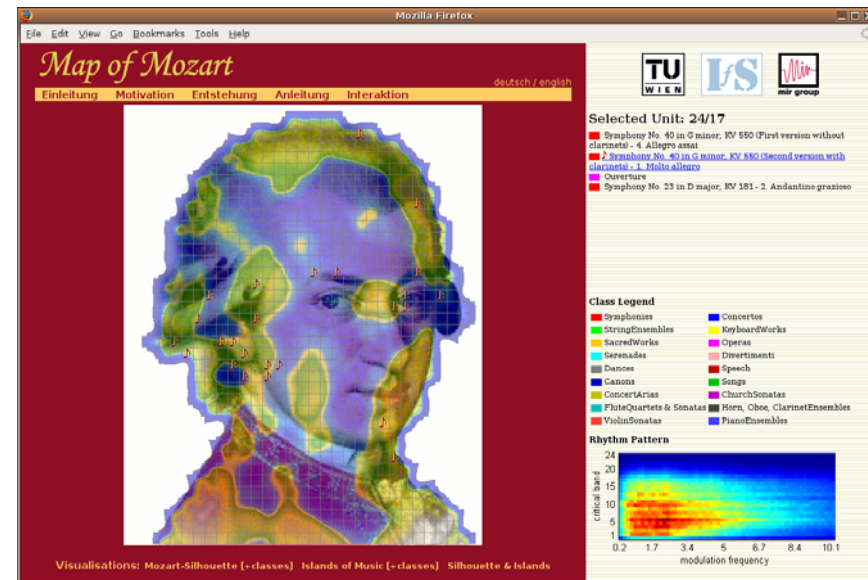
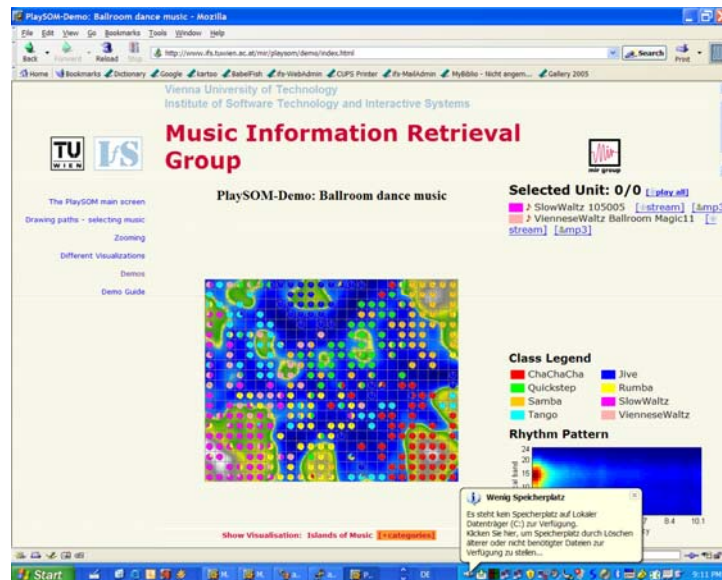


- Application for mobile devices
- Streaming audio
- Remote control



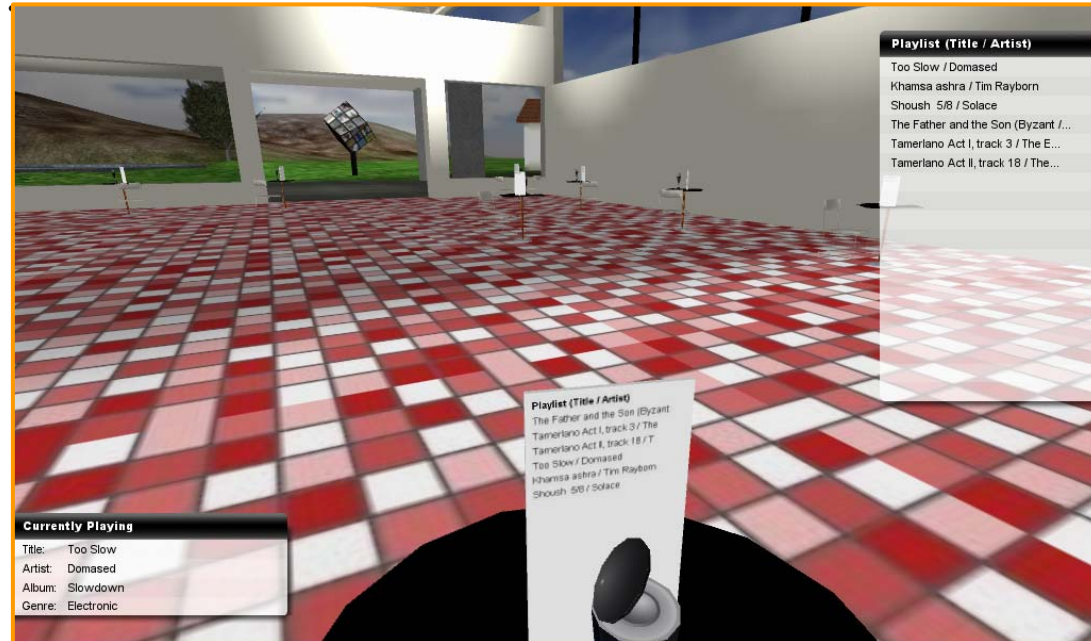


- Web-based interface
- Reduced functionality



- SOM organizes music by sound similarity
- forms baseline for room set-up
 - real-life & virtual
- Coffee shop, tables, each table plays its music
tables in a zone play similar music
- Get your coffee and choose a table where the music is to your liking (if there's one free there...)

3D Music Worlds



<http://ispaces.ec3.at/muscle.php>



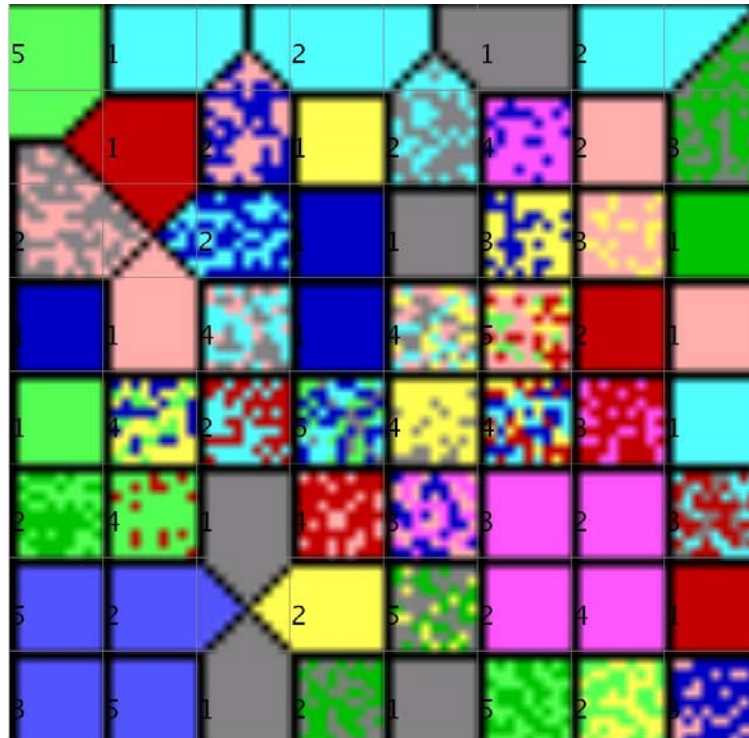
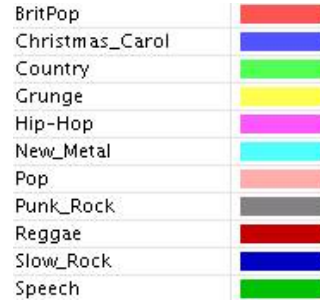


Comparing Multiple SOM Views.....

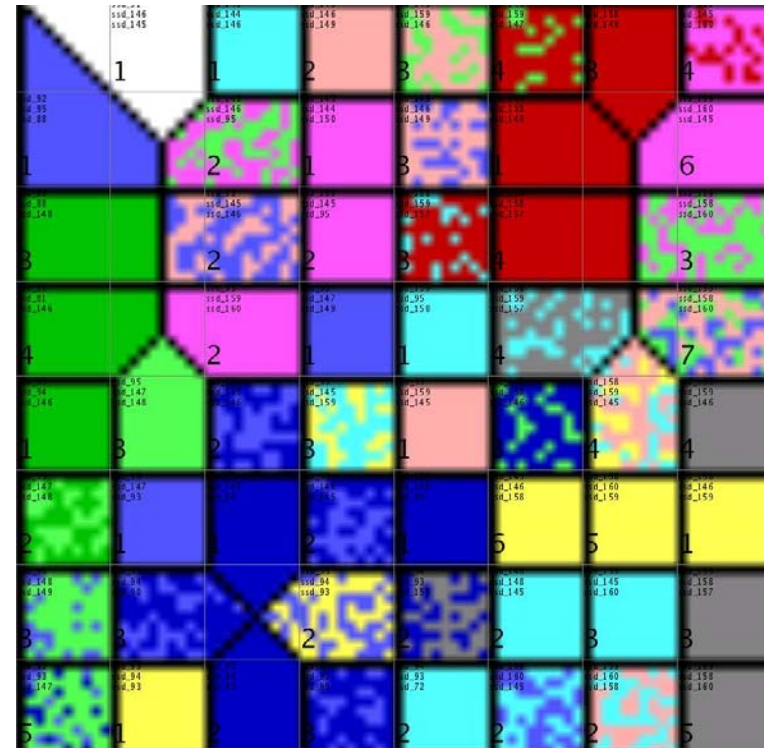
- Parallel corpus, indexed by song lyrics and music
- Clustering on a SOM for analysis
 - Lyrics SOM
 - Music SOM
- Analysis of cluster structure on both
- Class visualization based on genre labels

BritPop	
Christmas_Carol	
Country	
Grunge	
Hip-Hop	
New_Metal	
Pop	
Punk_Rock	
Reggae	
Slow_Rock	
Speech	

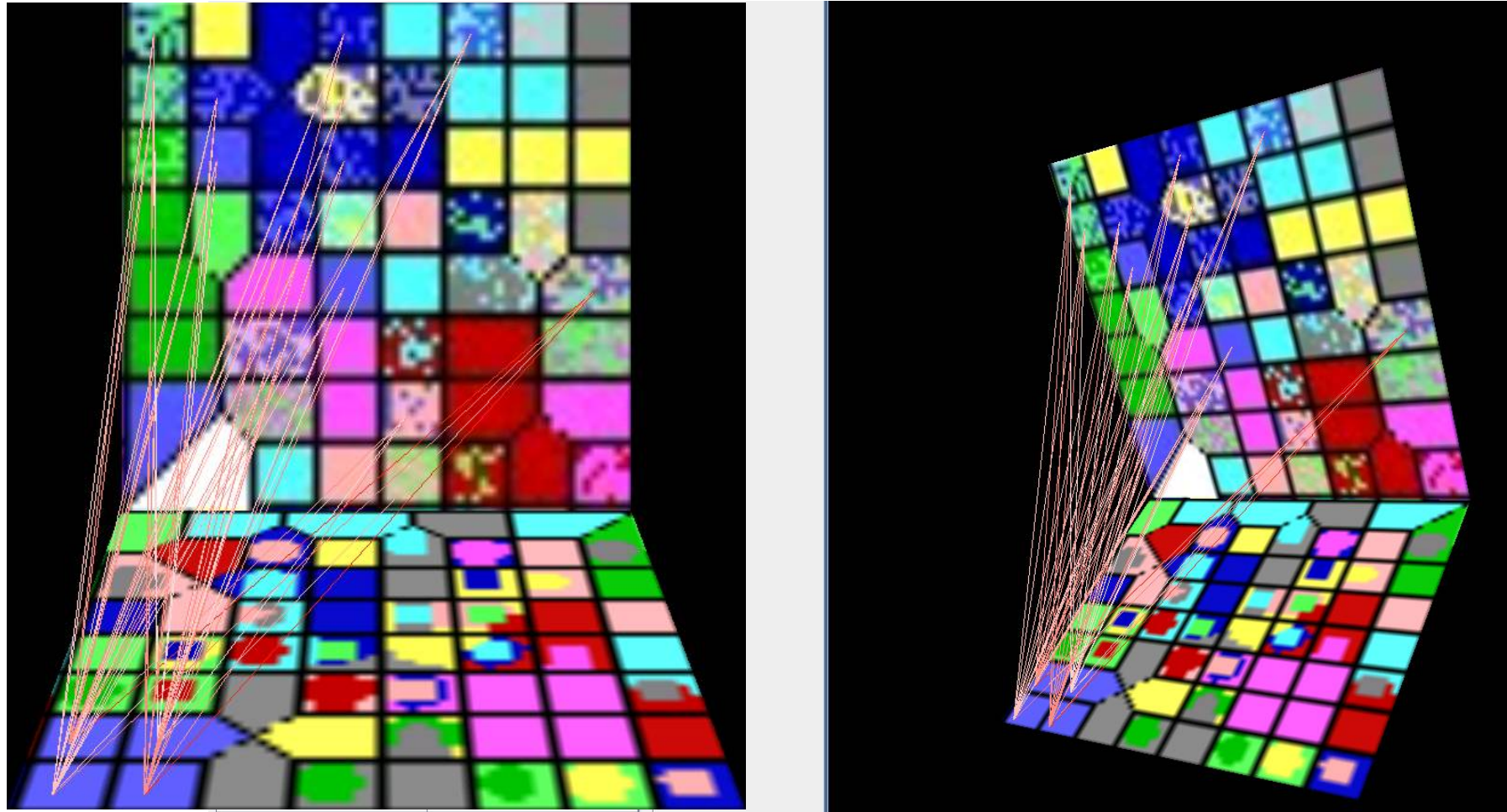
Text and Audio



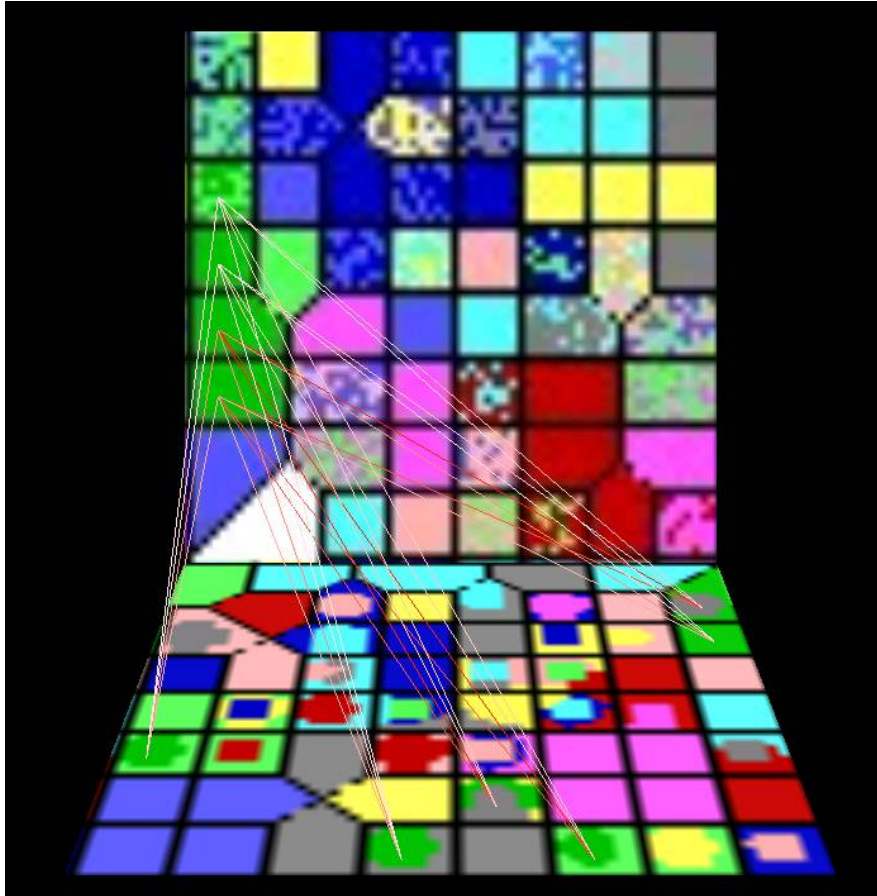
Lyrics SOM



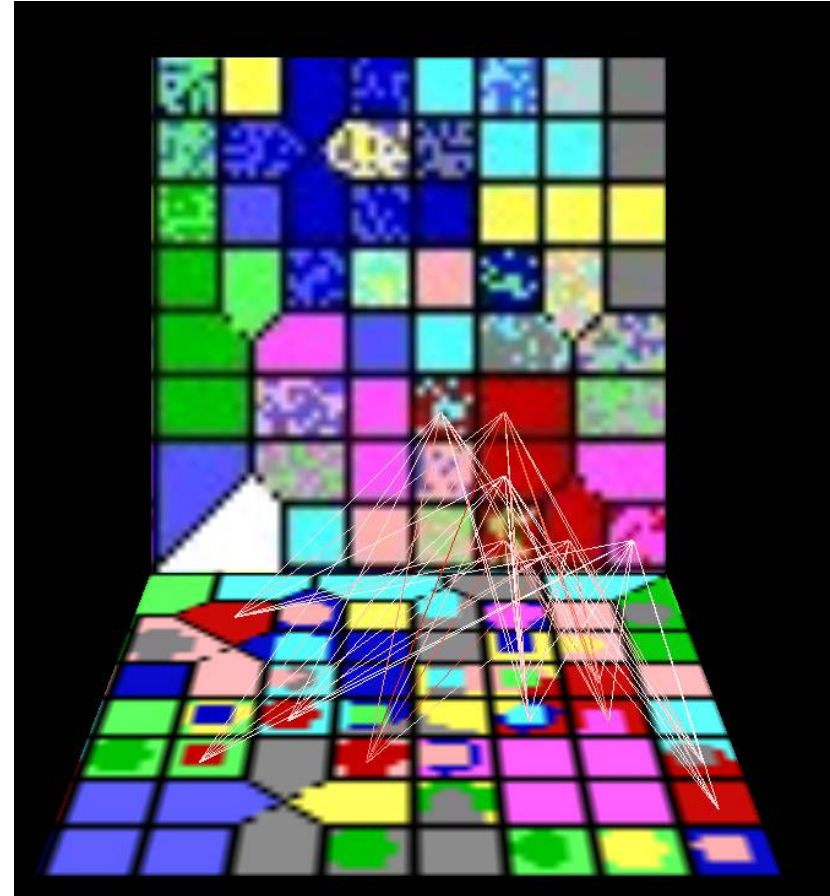
Music SOM



Christmas songs



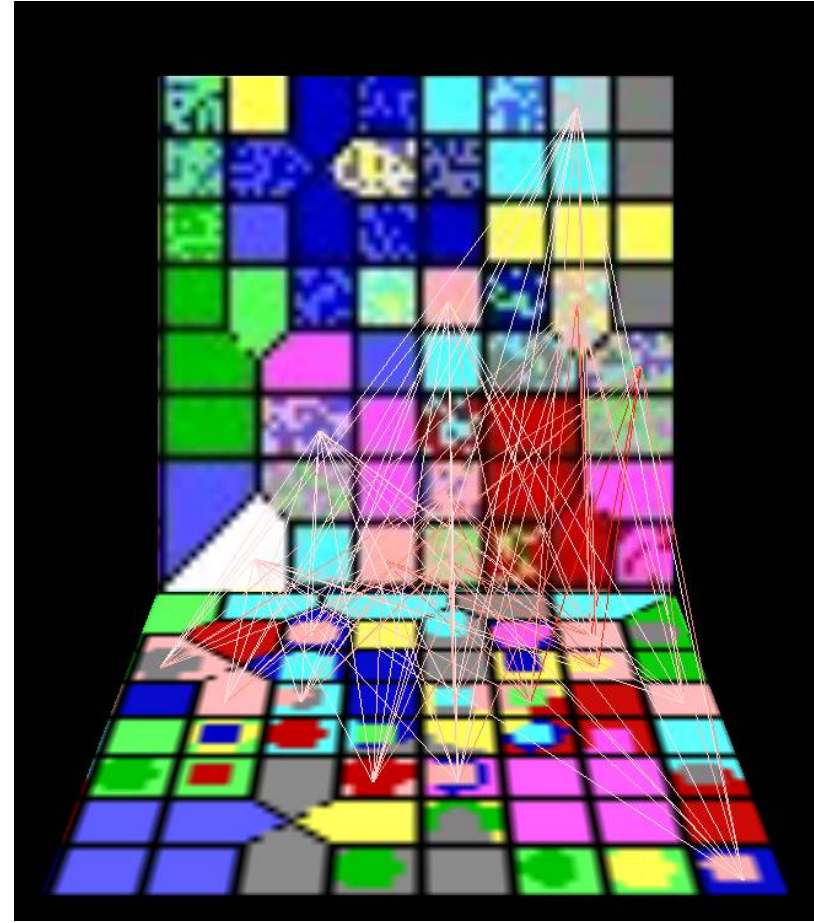
Speech



Reggae



Hip-Hop



Pop



Aktuelle

Forschungsaktivitäten

- Qualitätsmaße für SOMs
- Vergleich mehrerer SOMs
- Stabilität der Cluster-Topologie
- Semi-supervised SOMs