



Die Forschungsrichtung **Datenbionik** verwendet aus der Informationsverarbeitung in biologischen Systemen entlehnte Prinzipien für den Zweck der technischen Datenverarbeitung. Hierbei ist die **Selbstorganisation** ein zentrales Funktionsprinzip, welches in der belebten wie unbelebten Natur zu finden ist. In Systemen mit vielen kooperierenden elementaren Prozessen ist damit unter geeigneten Bedingungen die Bildung **emergenter Strukturen** verbunden.

### Selbstorganisation

Selbstorganisation ist die Fähigkeit eines Systems sich ohne gerichtete Vorgaben von außen zu ordnen. Überträgt man die Idee in den Bereich der Datenverarbeitung, so bedeutet dies, dass Daten nicht durch Vorgaben organisiert werden, sondern sich durch eine geeignete Computertechnik selbstständig organisieren. Datensätze sollen sich selbstständig zu zusammengehörigen Gruppen zusammenfinden. Weiterhin sollten solche Gruppen eine die wesentlichen Merkmale der Gruppe charakterisierende Beschreibung entwickeln.

### Emergenz

Emergenz bedeutet das Auftauchen einer übergeordneten Struktur, hier die Gruppierung, aus der Kooperation vieler elementarer Prozesse. Emergente Strukturen beschreiben das System der elementaren Prozesse auf einem neuen, übergeordneten Niveau. Ein populäres Beispiel für eine solche emergente Struktur ist das Auftauchen einer „La-Ola“-Welle die aus der Selbstorganisation vieler Menschen entstehen kann.

### Datenbionisches Knowledge Discovery

Die in der Arbeitsgruppe Datenbionik (Prof. Ultsch) entwickelten datenbionischen Verfahren haben ihren Schwerpunkt auf der Wissensgewinnung aus Datensammlungen (Knowledge Discovery in Databases). Hierbei geht es darum, in Datensammlungen bislang unbekanntes und nützliches Wissen zu entdecken. Die Emergenz von Strukturen ist hierbei nur ein erster Schritt. Ein weiterer Schwerpunkt der Werkzeuge ist die Erzeugung von verständlichem Wissen, das für Menschen direkt nutzbar ist.

## Selbstorganisierende Verfahren der Datenbionik

### U-Sphere

Die Selbstorganisation rezeptiver Felder im menschlichen Gehirn stand Pate bei der Entwicklung des U-Sphere Verfahrens. Die hochdimensionalen Datensätze werden hierbei, analog den Sinnesreizen auf den sensorischen Hirnkarten, mittels Selbstorganisation auf einer inselähnlichen Landschaft platziert, die sich auf einem Globus befindet. Diese Abbildung wird als U-Sphere bezeichnet und hat die Eigenschaft, wichtige Strukturen des Datenraumes derart zu visualisieren, dass eine für den

Menschen möglichst natürliche Darstellung entsteht. Die Anschaulichkeit der entstehenden Darstellung ermöglicht auch ungeübten Betrachtern die Strukturen des an sich unanschaulichen hochdimensionalen Datenraumes direkt zu erfassen. Die Information über Gleichheit oder Verschiedenheit von Datensätzen ist durch eine Höhenstruktur repräsentiert. Ein hoher Berg auf der U-Sphere bedeutet dabei einen deutlichen Unterschied in den hochdimensionalen Datensätzen, ein Tal hingegen eine Gruppe von ähnlichen Datensätzen.

Prof. Dr. Alfred Ultsch  
Fabian Mörchen  
Dirk Malorny  
Thorsten Graf  
Klaus Zaeper

FB Mathematik und Informatik  
Philipps-Universität Marburg  
Arbeitsgruppe Datenbionik  
Hans-Meerwein-Straße  
D-35032 Marburg  
Telefon +49(0)6421/282-2185  
Telefax +49(0)6421/282-89 02  
ultsch@informatik.uni-marburg.de  
[www.mathematik.uni-marburg.de/~nkisec](http://www.mathematik.uni-marburg.de/~nkisec)

## Anwendungsbeispiele

### Aktienanalyse



Die Selbstorganisation von 8.000 an den US Börsen notierten Unternehmen führte zu einer neuartigen Segmentierung des Aktienmarktes. Aus dem daraus gewonnenen Wissen konnte auf die Wahrscheinlichkeit eines Kursgewinnes von Aktien geschlossen werden.

Eine diesbezügliche Prognose erwies sich auch bei den derzeit generell schlechten Märkten als erfolgreich. ▶▶▶ U-Sphere Aktien

### DNA Microarrays



Bei einer Analyse von DNA Microarray Daten war es möglich, in einem Datensatz mit ca. 4000 Gendaten die für eine Krankheitsprognose relevantesten zu extrahieren. Die Reduktion führte zu 19 besonders relevanten Genen. Erst diese enorme Reduzierung der Komplexität ermöglicht eine weitergehende Untersuchung der Gene auf ihre genaue Funktion. Erste Untersuchungen lassen auf ein neues Verständnis von Tumorwachstum bei einer verbreiteten Art von Tumoren bei Kindern (Neuroblastome) schließen.

▶▶▶ U-Sphere Gendaten

### Kundensegmentierung



Die Anwendung auf das Kundenverhalten einer Mobilfunk Gesellschaft ermöglichte eine Klassifizierung der Kunden in verschiedene Gruppen. Hierin konnte neues Wissen über das Verhalten von Kundensegmenten gewonnen werden. Von besonderem Interesse waren die Kunden, die mit hoher Wahrscheinlichkeit bald den Vertrag kündigen.

▶▶▶ U-Sphere Mobilfunk

### Internistische Labordiagnosen



Ein Beispiel aus der Medizin ist die Diagnose von Krankheiten wie Borreliose oder Meningitis anhand einer Analyse der Rückenmarksflüssigkeit (Liquor). Die größte Gruppe (rot) die sich in der U-Sphere herauskristallisiert, repräsentiert die nicht infizierten Patienten. Die anderen Gruppen stellen zum Teil noch nicht bekannte Unterformen der Infektion dar.

▶▶▶ U-Sphere Liquor Diagnostik

## Databionic Artificial Life

Das Databionic Artificial Life stellt einen alternativen Ansatz der Datenbionik dar, der sich weniger an der Organisation sensorischer Karten als vielmehr an den Verhaltensweisen von Lebewesen in einem Kollektiv orientiert. Die künstlichen Lebewesen (Artificial Life Forms oder hier auch als DataBots bezeichnet) repräsentieren dabei einen Datensatz und haben ein bestimmtes Verhaltensmuster, nach dem sie mit benachbarten DataBots interagieren. Die zentrale Idee hierbei ist, dass sich kollektive Verhaltensmuster entwickeln, die mit strukturellen Eigenschaften eines hochdimensionalen Eingaberaumes übereinstimmen.

## Wissensgewinnung

Neben der selbstorganisierenden Darstellung von hochdimensionalen Daten ist die Gewinnung von neuem und verständlichem Wissen ein wichtiges Ziel. Hierzu wurden Algorithmen entwickelt, die die in der U-Sphere sichtbaren Strukturen abstrakt beschreiben können. Die Verständlichkeit der entstanden Beschreibungen, sowie die Erzielung eines möglichst hohen Abstraktionsgrades stehen im Vordergrund. In der Regel sind die hierbei entstehenden Wissens Elemente dazu geeignet, einem Experten des Gebietes einen Einblick in die im System erkannten Strukturen zu geben. An dieser Stelle kann ein Experte entscheiden ob das mit unseren Methoden gewonnene Wissen neue, bislang unbekannte Zusammenhänge enthält.

## Databionic

Databionic is about transferring the concepts of self organization and emergence from biological systems to data processing. The Research Group DataBionics (Prof. Ultsch) is developing methods for data analysis that show emergent behavior through self organization.

**Self organization** is the ability of a system to show high level structure by cooperation of many elementary processes. **Emergence** is the appearance of high level structure that is impossible to see from the elementary processes.

The ESOM (Emergent Self Organizing Map) adaptively adjusts to structures in a high dimensional space and gives a low dimensional projection that preserves the topology of the input space.

The central idea of the Artificial Life method is that a large number of artificial life forms simulated in a computer show collective behavioral patterns that correspond to structural features in a high dimensional input space.

The U-Sphere Method shows the structural information from the self organizing algorithms in form of a 3D landscape that offers a direct insight into the structures of the high dimensional input data. Data points inside a valley form a cluster.

The U-Spheres shown here are taken from applications to stock analysis, gene analysis, customer relationship management and medical diagnosis. An important goal is to extract knowledge from the data. For this purpose algorithms were developed that convert the visible structures of the U-Sphere into rules that represent the knowledge on a high abstraction level. Experts can evaluate these rules to decide whether new knowledge has been found.