

# **txtkit** - Visual text mining tool

SW/OfCD

Schoenerwissen / Office for Computational Design  
Anne Pascual & Marcus Hauer

studio@sw.ofcd.com – <http://www.sw.ofcd.com>

19. März 2004

## **Zusammenfassung**

Die Software **txtkit** ist ein kollektives Expertensystem, das die maschinelle und inhaltliche Analyse von textueller Information visualisiert. Nicht die umfassende und semantische Analyse eines Textes steht im Vordergrund, sondern die dynamische Darstellung strukturierter und numerischer Textdaten, sowie deren Transformation durch serielle Lektüren und Montagen. Statt der Anwendung statischer und vordefinierter Parameter und Wissensbasen, stehen die sich kontinuierlich ändernden Modalitäten dynamischer Wissensentwicklung im Vordergrund. Hierfür wurde ein innovatives Datenbanksystem und ein Textparsingprogramm zur Materialaufbereitung entwickelt, welches die Erschließung der Information visualisiert und den Kriterien der Relevanz und Repräsentativität anpasst. Das zweiteilige Interface der Software begünstigt zudem das Zusammenspiel von schriftsprachlicher und visueller Wissensexploration, mit dem Ziel veränderte Verfahren zur Erkenntnisgewinnung in den Blick zu bekommen.

## Inhaltsverzeichnis

<b>1 Hintergrund</b>	<b>3</b>
1.1 Zielsetzung . . . . .	3
<b>2 Material</b>	<b>4</b>
2.1 Bestimmung der textuellen Information . . . . .	4
2.2 Text im Netz . . . . .	5
2.3 Textbegriff . . . . .	6
<b>3 Parsing Verfahren</b>	<b>6</b>
3.1 Exkurs Megatext . . . . .	7
<b>4 Datenbank</b>	<b>9</b>
4.1 txtkit . . . . .	9
4.2 Kollektive Speicher . . . . .	10
<b>5 Interface</b>	<b>10</b>
5.1 Der Commandline Client txtshell . . . . .	11
5.2 UNIX as Literature . . . . .	12
5.3 Schriftbild / Schriftsysteme . . . . .	13
5.4 Schriftzeichen / Schlussweisen . . . . .	14
5.5 Bilder der Wissenschaft . . . . .	16
5.6 Monitoring Applikation txtvbot . . . . .	17
<b>6 Ergebnisse</b>	<b>18</b>
6.1 Der pool als Montage . . . . .	21
6.2 Interaktivität . . . . .	22
6.3 Gebrauchsszenario . . . . .	23
<b>7 Fazit</b>	<b>24</b>
<b>8 Credits Software</b>	<b>28</b>
<b>9 Anhang</b>	<b>29</b>

---

# 1 Hintergrund

**txtkit** wurde im Herbst 2002 als beispielhafte Applikation des Modellversuches KIT (Kunst Informatik Theorie) in Auftrag gegeben. Der in der Fächergruppe Kunst- und Medienwissenschaften an der Kunsthochschule für Medien Köln (KHM) angesiedelte Modellversuch KIT untersucht die Wechselwirkung von künstlerischer Praktik und neuen Technologien unter der Perspektive einer künstlerischen Ausweitung und Veränderung digitaler Werkzeuge.<sup>1</sup> Prof. Dr. Hans Ulrich Reck und Prof. Dr. Georg Trogemann leiten den Modellversuch, Prof. Dr. Hans Ulrich Reck war an der Konzeption von **txtkit** beteiligt und betreute die Entwicklung der Software.<sup>2</sup> Die Credits für die Software sind auf Seite 28 zu finden.

Wie im Exposé des Förderantrages formuliert, kommt der Entwicklung von Methoden einer experimentellen künstlerischen Praktik im Rahmen der Kunstausbildung eine entscheidende Aufgabe zu. Erfindung und Heuristik sind Teil einer praktischen Theorie der Künste. Ziel der Entwicklung eines solchen Prototyps ist es, veränderte Arbeitsmethoden und Einsichten über digitale Verfahren in Wissenschaft und Kunst zu bestimmen. Wie lassen sich wissenschaftliche Praxis und kulturelle Bildung mit Hilfe digitaler vernetzter Technologien abbilden und durch intelligente Transformationen bzw. apparative Synthesen visualisieren? Diese Frage stand zu Beginn der Entwicklung von **txtkit**.

*Sodass also eine Heuristik des Entwerfens, die sich als dynamisch prozessierende Imagination versteht, sich permanent in den Konstruktionsbedingungen der Programmarchitekturen und der Rechnerbewegungen bewegt.*

## 1.1 Zielsetzung

Dieser Bericht stellt im folgenden die Software **txtkit** vor und verweist parallel auf die dabei gewonnen Erkenntnisse für das Projekt Kunst - Informatik - Theorie. Im Fokus stehen hier Begriffe und Konzepte von Textualität, Schrift, Visualisierung, sowie Leser- und Autorschaft. Zunächst werden wir die Entwicklung der Software in drei Abschnitten nachzeichnen. Zu Beginn wird auf die spezifische Aufbereitung des Materials eingegangen, um dann die Entwicklung der Benutzeroberfläche auszuführen. Bevor wir den Bericht mit einer vorläufigen Bewertung von **txtkit** als Teil einer praktischen Theorie der Künste abschliessen, erläutern wir ausführlich das Gebrauchsszenario der Software und seine Konsequenzen im Hinblick auf das Verhältnis von Wissenschaft und Kunst.

Der Antrag des Modellbausteins enthält folgende Absichten:

- Das geplante Expertensystem zur Aufbereitung theoretischer, textbasierter Wissensquellen soll unabhängig von Inhalten und Formaten die Arbeit am und mit dem Text aufzeichnen.

---

<sup>1</sup>Vgl. H. U. Reck und G. Trogemann, 'Exposé zum Antrag im Modellversuchsförderprogramm Kulturelle Bildung im Medienzeitalter' (URL: <http://www.khm.de/kmw/kit/pdf/expose.pdf>).

<sup>2</sup>Erste Anwendungsbeispiele von **txtkit** finden sich in den Marginalspalten, die dort aufgeführten Inhalte wurden den ersten Gedankenmaterialien der Software entnommen, welche von Prof. Dr. Hans Ulrich Reck verfasst wurden.

- Statt statischer Strukturabbildung, stehen die Modalitäten dynamischer Wissensentwicklung im Vordergrund.

- Ein innovatives Datenbanksystem zur Materialaufbereitung, sowie die Visualisierung zur Darstellung der Transformationsprozesse werden gemeinsam mit den Verarbeitungsregeln der Software entwickelt.

Eine weitere Forderung war, eine Situation zu schaffen, in der Erkenntnisgewinnung auf Erfahrung basiert, in dem Sinne, dass sich der Prozess der Aneignung der Ideen hin zu den Objekten des Wissens der Handelnden verschiebt.

## 2 Material

### 2.1 Bestimmung der textuellen Information

*Autorschaft bleibt individuell, unverzichtbar, evident und unersetzbar, aber die Figur des Autors wird selber zum formalisierbaren und inszenierbaren Stoff einer 'Kunst durch Medien'.*

Um theoretische, textbasierte Wissensquellen adäquat aufzubereiten, muss geklärt werden, welche Eigenschaften das Material aufweist und wie es dem zu Folge verarbeitet werden kann. Im Fall von **txtkit** wurden zunächst bereits verfasste, publizierte und rezipierte kunsttheoretische Schriften verwendet, die nach Autorisierung des Autors Prof. Hans Ulrich Reck, für eine beliebige Anzahl von Nutzern, innerhalb dieses besonderen Rahmens zur Bearbeitung freigegeben wurde.

Die digital überlieferten Textdokumente, wurden auf Seiten angeordnet, welche als einzige Referenz zu anderen Verwendungszusammenhängen in Buchform dienen. Das Verfassen und Erstellen der Textdateien wurde mit Hilfe eines herkömmlichen Textverarbeitungsprogramms vorgenommen, somit liegt ein spezifischer Produktionszusammenhang vor, der sich in der Art des Schreibens und des Materialträgers widerspiegelt. Den technische Vorgang des Verfassens von Texten am Computer blenden wir jedoch weitgehend aus, um die Bedingungen des Auftauchens von Gedanken und ihrer Beurteilung freizulegen.

Unser Fokus liegt somit nicht auf den veränderten Vorgaben bei der Erstellung von Texten und Inhalten. Wir verwenden bereits vorhandene textbasierte Wissensquellen, um veränderte Methoden zur Beobachtung von Transformationsbeziehungen einzuführen. Damit steht die Bearbeitung textueller Information und nicht primär das Erstellen von zusammenhängenden Schriften in digitalen, vernetzten Darstellungsräume im Mittelpunkt unsere Aufmerksamkeit. Der unvermeidlichen Nähe zu Fragestellungen der Textkritik und Interpretation, sowie deren Aufgaben Sinn und Bedeutungen zu generieren, setzten wir einen erweiterten Lektüre begriff gegenüber.

An dieser Stelle wird kurz auf den veränderten Status von Textualität in Alltag und Wissenschaft eingegangen, um zu erläutern, woraus diese veränderte Fragestellung resultiert.

---

## 2.2 Text im Netz

Die Produktion und Veröffentlichung von Texten im Internet ist zur alltäglichen Praxis geworden. Dass sich die Art der Erstellung digitaler Dokumente, ihre Speicherung, Distribution und Zugriffsmöglichkeiten auf die Inhalte von Texten auswirkt, lässt sich an zahlreichen Schreib- und Publikationsformen im Internet beobachten. So finden sich unter den Sammelbegriffen digitale Literatur und Hypertext zahlreiche Beispiele für die Möglichkeiten vernetzten Schreibens, Publizierens und deren Wechselwirkungen auf Form und Inhalte von Literatur.<sup>3</sup> Besonders die Kombination von Text, Bild und Klang verbunden mit der Technik des Linking verheißen ein größeres Maß an Variabilität und Intervention. Diverse Formen digitaler Schriftlichkeit von Individuen und Gruppen bilden den Gegenstand zahlreicher wissenschaftlicher Untersuchungen, ebenso wie Sprachformen und Sprechweisen im Netz.

Die Geschichte des computerisierten Netztextes beginnt mit den Entwürfen Vannevar Bushs, Douglas Engelbarts und Ted Nelsons. In ihren Hypertext-Konzepten werden kulturelle und technologischen Konstruktionen unauflösbar miteinander verbunden. Vergleicht man ihren Anspruch, bisherige Fixierungen durch nicht-lineare, netzartige Strukturierungen aufzubrechen, zeigt die heutige Situation, dass ihre Vorstellungen teilweise realisiert, teilweise obsolet wurden oder sie aber gar von der Politik der Netze überholt zu sein scheinen.<sup>4</sup>

Gerade weil diese Divergenzen offen zutage treten, befassen sich auch die Sozial- und Geisteswissenschaften zunehmend mit der Umbruchsituation kultureller Kommunikationsverhältnisse. Denn mit der Entwicklung der neuen Medien seit dem ausgehenden 19. Jahrhundert zeichnen sich klar veränderte Vorgaben von Textualität ab. Ebenso wie während des Überganges vom Mittelalters zur Frühen Neuzeit, als die Verschriftlichung der Volkssprachen im 12. Jahrhundert, so wie die Erfindung des Buchdrucks im 15. Jahrhundert, die Dominanz von Textualität etablierten, erfahren wir heute eine gegenläufige Bewegung.<sup>5</sup>

Kulturwissenschaftliche Forschung begegnet dem medialen Wandel in dem sie sowohl ihren Gegenstandsbereich öffnet und erweitert, als auch ihre textwissenschaftlichen Methoden modifiziert. Verstärkt geraten Tätigkeiten des Produzierens, Herstellens, Machens und der Handlungen, der Austauschprozesse, Veränderungen und Dynamiken, der Akteure sowie kulturelle Ereignisse in den Fokus der Wissenschaft. Dies hat auch zur Folge, dass sich vor allem die Textwissenschaften von der systemischen Analyse und von den auf Überprüfbarkeit oder auf Wiederholbarkeit und Konstanz basierende Verfahren zugunsten einer Erforschung des Materials und von Handlungsvollzügen lösen.<sup>6</sup>

---

<sup>3</sup>Vgl. Roberto Simanowski, *Interfictions. Vom Schreiben im Netz* (Frankfurt a. M.: Suhrkamp, 2002).

<sup>4</sup>Vgl. Stephan Porombka, *Hypertext. Zur Kritik eines digitalen Mythos* (München, 2001).

<sup>5</sup>Vgl. Dr. Friedrich Balke, 'Programm des Kulturwissenschaftlichen Forschungskollegs Medien und kulturelle Kommunikation an der Universität zu Köln' (URL: [http://www.uni-koeln.de/inter-fak/fk-427/prog/prog\\_fk\\_intro.html](http://www.uni-koeln.de/inter-fak/fk-427/prog/prog_fk_intro.html)).

<sup>6</sup>Vgl. Prof. Dr. Erika Fischer-Lichte, 'Gesamtkonzept Kulturen des Performativen, Son-

## 2.3 Textbegriff

*Wissenschaft entscheidet sich nicht an der Frage der formalen Kohärenz oder Formatierung von Datenflüssen, sondern an der extern überprüfbaren Referenz von Aussagen auf - damit oder dafür - kompatible Gegenstände.*

Die Erstellung von Textdokumenten, sowie ihre Publikation im Internet unterscheidet sich von herkömmlichen Zusammenhängen aufgrund zweier Faktoren: Entmaterialisierung und Entäußerung.<sup>7</sup> Texte sind von ihrer ursprünglichen Formgebung, Buch, Papier, Layout und den entsprechenden Zugriffsmöglichkeit (Erwerb, Bibliothek) entbunden. Die damit verbundenen Konventionen werden zwar zunächst aufgehoben, sie sind aber weiter wirksam. Gerade weil ein Großteil unserer Kulturtechniken auf diesen Konventionen beruht, verstellen diese all zu leicht den Blick für nicht-intentionale und diskursive Aspekte wissenschaftlicher Symbolproduktion.

Beide Faktoren - Entmaterialisierung und Entäußerung der bisherigen Form und Struktur von Texten - verändern das Verhältnis von Autor und Text, aber auch des Lesers zum Text, ebenso wie die Beziehung von Texten untereinander. Zum einen transformiert die Entmaterialisierung die Textgestalt zugunsten dynamischer Gebrauchsmodi und zum anderen bedeutet die Entäußerung der Textgestalt eine scheinbare Überwindung bisherigen Einschränkungen der Autorschaft und der Zugriffsquellen.

Um sich diesem veränderten Textbegriff nähern zu können, wäre sowohl eine Untersuchung seiner Materialität oder aber die Erforschung der kulturellen Implikationen, *Kultur als Text*, *Text als Kultur* von Interesse. Wir möchten uns stattdessen auf die Verfahren explorativer Erkenntnisgewinnung konzentrieren, welche den herkömmlichen Textbegriff überschreiten. Die Software **txtkit** dient uns dabei als technisches System, das neue Techniken zur Umdeutung materialer und kultureller Bedingungen zur Verfügung stellt. Im Folgenden wird der systemische Aufbau der **txtkit** Schnittstelle erläutert werden.

## 3 Parsing Verfahren

Der Aufbau der Software sieht vor, dass deutsche und fremdsprachige Textdokumente (im Unicode utf-8 Format) mittels eines eigens programmierten Parsers eingelesen und in ihre Einzelteile zerlegt werden. Hierbei kommen reguläre Ausdrücke (Regular expression bzw. regexp/regex) zum Einsatz, mit denen sich (Unter-)Mengen von Zeichenketten beschreiben lassen. Diese Ausdrücke werden verwendet, um bestimmte Muster zu suchen und dann eine spezifische Aktion auszuführen.

Die geparsten Texte liegen damit nicht mehr als geschlossene, lineare Schriftbilder vor, sondern als einzelne Datensätze, die in einer Datenbank abgelegt und indexikalisiert werden. Durch eine quantitative Erhebung werden alle textuellen

derforschungsbereich FU Berlin' (URL: [http://www.sfb-performativ.de/seiten/frame\\_impressum.html](http://www.sfb-performativ.de/seiten/frame_impressum.html)).

<sup>7</sup>Vgl. Ioannis Kanellos, 'Texte et intertexte, concepts de médiation entre les nouvelles technologies et le requisit d'une assistance à la compréhension', in: Eric Sadin (Hrsg.), *éc/artS 3* (Paris, 2002).

---

Informationen numerisch erfasst, Schlüsselwörter bestimmt und geordnet. Keywords sind in Sätzen und diese wiederum in Paragraphen als Einheiten abrufbar bzw. lokalisiert. Durch diese Organisationsform ändert sich die Beschaffenheit des ursprünglichen Textmaterials wesentlich, da es nicht länger als zusammenhängendes Gebilde vorliegt, kein Anfang und kein Ende hat. Seine durch den Verfasser angelegte Form wurde vollständig aufgelöst und kann nicht mehr mit den herkömmlichen Wahrnehmungs- und Denkmustern erfasst werden. Die textuelle Information enthält keine Fußnoten, Seitenzahlen, Inhaltsverzeichnis oder andere *Gesten der Wissenschaftlichkeit*.<sup>8</sup> Das Parsing ist eine integrale Funktion der Applikation, da es zum einen die Grundlage für die maschinelle Analyse bildet und zum anderen wiederholt und variabel ausgeführt werden kann, so dass beliebig viele Texte mit **txtkit** verwendet werden können.

Zunächst möchten wir kurz darlegen, warum wir uns für eine rein quantitative Erfassung der Daten entschlossen und von einer semantischen Analyse des Materials abgesehen haben.

### 3.1 Exkurs Megatext

Eines der vielversprechendsten Projekte zur Visualisierung semantischer Textfelder wurde an der Universität Göteborg im Bereich der Computer-Linguistik realisiert. Das Projekt Megatext wurde von Maria Hakansson und Maria Lindqvist entwickelt<sup>9</sup> und soll an dieser Stelle exemplarisch die Vor- und Nachteile semantischer Textanalyse deutlich machen.

Hakansson und Lindqvist planten die Visualisierung grosser Textmengen mit Hilfe computerlinguistischer Verfahren. Hierzu verwendeten sie Texte, die in einer Datenbank abgelegt und bereits mit Zusätzen zur Identifizierung semantischer Ausdrücke versehen wurden. Als Wissensbase verwendeten Hakansson und Lindqvist die Reuters Datenbank, die zu jedem Wortlaut einen Verweis auf sein entsprechendes Bedeutungsrepertoire enthält, das dann mittels XML Code beschrieben wird. Es gibt unterschiedliche Datenbanken dieser Art, die tagtäglich mit neuen Datensätzen gefüttert werden, und bei der Entwicklung von NLP (Natural Language Processing) Verfahren eingesetzt werden.

NLP Verfahren überprüfen die Wahrscheinlichkeit von Bedeutungen. Durch den Vergleich von Einträgen z.B. der Reuters Datenbank wird während des Einlesens oder Parsings grosser Textmengen mittels NLP überprüft, welche Bedeutung des Wortlauts die wahrscheinlichste ist. Dieses Vorgehen basiert auf dem Prinzip semantischer Gleichartigkeit. Wörtern wird die entsprechende Bedeutung

---

<sup>8</sup>Michael Kahn bestimmt in seinem Aufsatz die Fussnote als jenen Gegenstand an dem Diskursanalyse und Buchgeschichte, Literaturwissenschaft und Wissenschaftsgeschichte miteinander korrespondieren können. Michael Cahn, 'Die Rhetorik der Wissenschaft im Medium der Typographie. Zum Beispiel die Fussnote', in: Hans-Jörg Rheinberger u.a. (Hrsg.), *Räume des Wissens. Repräsentation, Codierung, Spur* (Berlin, 1997).

<sup>9</sup>M. Hakansson und M. Lindqvist, 'Megatext. Visualising Semantic Relations', Diplomarbeit, Computational Linguistics, University of Göteborg (2002), (URL: [http://www.cling.gu.se/~c18mhaka/MegaText\\_compressed.pdf](http://www.cling.gu.se/~c18mhaka/MegaText_compressed.pdf)).

zugeordnet, je nachdem welche Aussage ihnen zuvor in welchem Kontext zugeschrieben wurde. Derartige Regeln der semantische Gleichartigkeit kommen bislang vor allem bei Suchmaschinen zum Einsatz. Hier kann der Suchbegriff des Nutzers mit verwandeten Begriffen in Verbindung gesetzt und erweitert werden.<sup>10</sup>

Ein Vorteil von NLP ist sicherlich die Expansion des Suchfeldes für den Nutzer, was mit dem Ausdruck unscharfe Suche gemeint ist. Allerdings mussten wir bei Tests von NLP Datenbanken feststellen, dass Ähnlichkeiten nur selten übereinstimmen bzw. sehr von den Anwendungsgebieten der Datenbank abhängen. Die Grenzen eines Bedeutungsrepertoires werden vor allem durch den Publikationskontext festgelegt. Da die Datenbanksammlungen meist Artikel und Beiträge als Quellen nutzen, wird das Vokabular unterschiedlicher Fachsprachen und Quellen vermischt. Wie sich zeigt, kommt es dabei statt zu einer Verfeinerung zu grossen Ungenauigkeiten, was die Variation und die Abstufung von Bedeutungsinhalten betrifft.

Dieses Problem resultiert sicher aus der Tatsache, dass trotz wachsender Datenbanksammlung die Auswahl des verfügbaren Wortschatzes im Vergleich zu unserem Sprachgebrauch viel zu klein ausfällt. Im Zuge der Etablierung vernetzter lexikalischer Sammlungen kann dieses Missverhältnis sicherlich kontinuierlich abgebaut werden. Ein weiterer Nachteil liegt darin, dass die Mehrzahl der Systeme Eigennamen und Spezialwörter nicht erkennt. Wir prüften die Anwendbarkeit anhand von Beispieltexen, die mittels NLP jedoch nicht hinreichend ausgelesen werden konnten, da eine Vielzahl der verwendeten Ausdrücke nicht erkannt wurden, also noch nicht Eingang in die Datenbanksammlungen gefunden hatten.

Das Projekt Megatext nutzt die Reuters Datenbank als Wissensbasis, um sie dann mit einer mathematischen Lernmethode LSI/LSA visuell auszugeben. LSI/LSA gleicht der in **txtkit** realisierten Aufbereitung insofern, als dass auch **txtkit** die Wörter zählt und entsprechend der Häufigkeit ihres Vorkommens als Schlüsselwörter bestimmt und ordnet. Diese statistische und quantitative Analyse wurde von uns allerdings in mehrerer Hinsicht im späteren Verlauf erweitert, vor allem aufgrund der Inbezugnahme der Handlungen des Anwenders.

Hakansson und Lindqvist erstellen im Folgenden konzeptionelle Indizes von Texten, um semantische Felder anzulegen, welche Wörter und Dokumente im Hinblick auf ihre Bedeutungsnahe organisieren. Die Visualisierung gibt Aufschluss über die Struktur von Dokumenten, am prägnantesten dadurch, dass Kategorisierungen farblich angezeigt werden.

Ein längerer Vergleich mit Megatext erscheint uns an dieser Stelle nicht weiter notwendig, da der Ansatz zur Darstellung semantischer Textfelder sich von dem unsrigen wesentlich durch zwei Aspekte unterscheidet. Zum einen findet hier keine Rückkopplung von Aufbereitung und Visualisierung bzw. Nutzer statt und zum anderen wird auf bereits vorgegebene Bedeutungszuweisungen zurück-

---

<sup>10</sup>Im Unterschied dazu vgl. U. Ecos Begriff des Codes als Wahrscheinlichkeitssystem Joachim Huber, *Urbane Topologie. Architektur der randlosen Stadt* (Weimar, 2002), S. 241.

---

gegriffen, was den Ausgang der Versuchsanordnung in unseren Augen stark einschränkt. Zudem wird der dynamischen Entwicklung von Bedeutungen kaum Rechnung getragen, sondern vor allem die effektivere und eindeutige Zugriffnahme von Texten verfolgt. Besonders letzteres würden wir anhand von **txtkit** zugunsten sich kontinuierlich ändernden Modalitäten dynamischer Wissensentwicklung in Frage stellen. Um die technischen Grundlagen, die dies ermöglichen sollen, näher auszuleuchten, gehen wir nun auf das Datenbanksystem von **txtkit** ein, um die von uns gewählte Kombination von maschineller und inhaltlicher Analyse erläutern zu können.

## 4 Datenbank

Betrachtet man die Speicherung des geschriebenen Wortes in das Organisationsmodell Datenbank, ändern sich die Wahrnehmungsbedingungen. Lev Manovich beschreibt die *Datenbank* als symbolische Form und sieht in ihrer breiten, vermehrten Nutzung einen Wandel kultureller Ausdrucksformen.<sup>11</sup> Denn eine Sammlung einzelner, gleichwertiger Datenobjekte weist keine zeitliche, inhaltliche oder formale Entwicklung auf. Erst wenn die Datenbank eine hierarchische, vernetzte, relationale oder eine objektorientierte Struktur aufweist, fügen sich die Objekte in ein komplexes Gebilde ein. Im Unterschied zu narrativen Organisationsmodellen fehlen dramaturgische Strukturen, die in manchen Hypertextanwendungen durch strukturelle Darstellung ersetzt wurden, ohne jedoch auf die hierarchische oder evolutionären Aspekte zu verzichten.

Datenbanken stellen nicht per se ein dynamisches Modell zur Datenverwaltung dar. Die Ein- und Ausgabe der Datensätze basiert meist auf statischen und vordefinierten Parametern. Man kann sagen, dass die Operationalität wesentlich von der Struktur solcher Wissensbasen abhängig ist, d.h. es handelt sich dabei um ein deterministisches System: es kann nur das abgefragt werden, was bereits vorliegt, in unterschiedlicher Komplexität oder als Variationen.

### 4.1 txtkit

Durch das Parsing Verfahren wurden die Textdaten zerlegt, welche nun in einer Datenbank erfasst wurden. Diese sichert zwar den Wortlaut, führt aber nicht zu einer originären Fassung zurück. Unser Entwurf sieht vor, dass die Datensätze kollektiv bearbeitet werden können, hierfür müssen diese in einer Datenbank auf einem zentralen Server abgelegt werden, und können so via Internet von mehreren Nutzern gleichzeitig genutzt werden. Diese Infrastruktur führten wir vor allem aus dem Grund ein, um neben den Textdaten, die Interventionen aller Nutzer protokollieren zu können.

In Anlehnung an adaptive Erkennungstechnologien zeichnet die Datenbank die Handlungen der Anwender auf, welche textuelle Information sie auf welchem

---

<sup>11</sup>Vgl. Lev Manovich, *The Language of New Media* (Cambridge, Mass., 2001), S. 218ff.

*Ganz abgesehen davon - ein weiteres Tabuthema in der theoriepoetischen Medieneuphorie - dass zentralisierte Datenbanken, die wesentliches Grundlagenwissen der avancierten Forschung, zum Beispiel in der Biochemie, aufbereitet haben, in keiner Weise allgemein oder öffentlich zugänglich sind.*

*Kollektive Autorschaft als eine Form der Kooperation verwandelt den Produzenten in den Zeugen, Urheberschaft in Kommentar, Erfindung in Kritik.*

Weg erschliessen. Auf diese Weise werden die inhaltlichen Analysen der Nutzer unmittelbar mit den Textdaten verknüpft, was zur Folge hat, dass die zu Beginn angelegte statistische Erhebung verfeinert wird und damit die maschinelle Analyse im weiteren Verlauf entscheidend erweitert.

Die Anzahl der verzeichneten Gebrauchsspuren wächst durch die Lektüren, gleichzeitig steigen die Relevanzwerte, die sich im Unterschied zu vordefinierten Parametern und Wissensbasen den unterschiedlichen Methoden und Modellen für die Erkennung von Zusammenhängen dynamisch anpassen.

## 4.2 Kollektive Speicher

*Die Tatsache, dass ohne Medien, ohne zwischengeschaltete Technologien, Speicherapparate, Sende- und Empfangsvorrichtungen, diese Kommunikation nicht funktioniert, bedeutet, dass neben den Codes der Signalübermittlung vor allem die Gesten, Rhetoriken und alle orientierenden Formatierungen von Erfahrungen über 'Objektivität' entscheiden.*

txtkit verwendet die Datenbank also nicht allein zur Speicherung des Textmaterial, sondern vor allem auch zur Protokollierung der Gebrauchsspuren. Hans-Jörg Rheinberger hat vielfach darauf hingewiesen, dass die Sicherung naturwissenschaftlicher Daten in Form von Datenbanken, wie es beispielsweise die Aufzeichnungen von Laborergebnissen sind, nicht zuletzt ihrer Verkollektivierung dient. Auf diese Weise wird suggeriert, dass wissenschaftliche Sachverhalte interpersonal verglichen und ausgewertet werden können. Die Aufzeichnung und Verwaltung von Daten sind Elemente wissenschaftlicher Darstellung und Inszenierungsweisen.<sup>12</sup> Dabei bestimmen die technischen, symbolischen und medialen Bedingungen des jeweiligen Experimentalsystems die Bedeutungszuschreibung, die der Gesamtheit der Daten im Nachhinein widerfährt, wesentlich mit.

Auf die Problematik naturwissenschaftlicher Evidenzherstellung kann hier nicht ausführlicher eingegangen werden, allerdings entsteht durch die Verwendung einer Datenbank zur Speicherung schriftlicher Äußerungen ein interessantes Spannungsfeld von wissenschaftlichen Methodiken. In unserem Zusammenhang gilt es herauszustellen, inwieweit sich dieser Teil der Versuchsanordnung auf die Erkenntnisgewinnung auswirkt und ob Rheinbergers Feststellung zur Diskursivität wissenschaftlicher Praxis nicht unter ganz anderen Vorsätzen Eingang in geisteswissenschaftliche Zusammenhänge finden kann. Es bleibt weiter zu beobachten, welchen Einfluss die Effekte experimenteller Verfahren auf die Wissensgenerierung in geisteswissenschaftlichen Kontext haben.

## 5 Interface

Um eine Rückkopplung von maschineller und inhaltlicher Analyse zu erreichen, bzw. sie gemeinsam mit den immanenten Qualitäten explorativer Erkenntnisgewinnung in den Blick zu bekommen, stellt txtkit ein zweiteiliges Interface zur Verfügung.

Das Interface teilt sich in zwei unterschiedliche aber mit einander verbundene Module. Ein Modul der Commandline Client `txtshell`, ist für die schrift-

---

<sup>12</sup>Vgl. Hans-Jörg Rheinberger, *Experimentalsysteme und epistemische Dinge. Eine Geschichte der Proteinsynthese im Reagenzglas* (Göttingen, 2001).

---

sprachliche Ein- und Ausgabe reserviert, der zweite Teil der Software `txtvbot` – visualisiert die Ergebnisse der maschinellen und inhaltlichen Analyse.

In der freien Koordination zwischen `txtshell` und `txtvbot` geben verbale wie non-verbale Zeichen Aufschluss über die textuelle Information und lenken die Schlussweisen des Lesers. Die Visualisierung verzichtet auf strukturelle Darstellungen z.B: einer Gesamtansicht des Textes, wie es Hypertextdiagramme zumeist sind. Statt eine allumfassenden Übersicht von Verknüpfungen zu schaffen, geht es darum, Darstellungsmittel zu entwickeln, die Erfahrung als Bestandteil der Erkenntnisgewinnung sichtbar machen und die Wechselwirkung von bereits verfasstem Wissen und noch nicht fixiertem Wissen in den Blick zu bekommen.

## 5.1 Der Commandline Client `txtshell`

Um die vermeintliche Nichthintergebarkeit von Schrift<sup>13</sup> und die Referenz zur Buchform zu vermeiden, suchten wir eine Form, die den Unterschied von Zeichen und Schrift kenntlich macht und die dem Computer inhärenten Nutzungszusammenhang entspringt.

Der `txtshell` Client, welcher innerhalb einer Unix-Shell läuft, teilt sich in drei Bereiche: Status, Ausgabe/GUI und Commandline. Die Commandline dient, wie ihr Name bereits suggeriert, der Eingabe von Befehlen, bzw. Operationen, welche in unserem Fall, dem Nutzer ermöglichen, den Text zu durchforsten, zu lesen und zu bearbeiten.

Ein derartiges Eingabefenster dient in herkömmlichen Nutzungszusammenhängen der Steuerung aller wesentlichen Abfragen der Dateien, sowie die Verwaltung der Schreib-, Lese- und Ausführungsrechte. Hierzu werden Kommandos eingegeben mit denen der Anwender die angezeigten Ergebnisse bearbeiten kann. Die erste Shell dieser Art wurde 1979 für das UNIX Betriebssystem entwickelt und hat ihre ursprüngliche Gestalt seitdem nicht wesentlich verändert. Weite Verbreitung fand eine weitere Variante der Shell, C shell, oder auch `csh` genannt. Die Kommandos der C Shell ähneln den Angaben, die für die Programmiersprache C verwendet werden.

Der Shell-Client wird automatisch beim Starten von `txtkit` ausgeführt, allerdings besteht auch immer die Möglichkeit dieses Modul separat in der Shell, plattformunabhängig auszuführen. Entscheidet sich der Nutzer für den selbstgesteuerten Lesemodus, und nicht für einen semi-automatisierten (basierend auf einem vorher gespeicherten `pool`), kann er durch die Eingabe von `goto` oder der entsprechenden Kurzform `g` kombiniert mit Begriffen oder numerischen Angaben den Text durchforsten. Ist die Anfrage erfolgreich, erscheinen alle Stellen, Orte an denen der eben geschriebene Begriff auftaucht bzw. ähnliche oder verwandte Begriffe. Eine `goto` Abfrage zeigt eine begrenzte Anzahl numerierter Satzfragmente an, die den eingegebenen Begriff enthalten. Um die angezeigten Gedanken einzusehen, wählt der Anwender mittels des Befehls `read` (oder

*'Écriture' und 'inscription' treten an die Stelle der Schrift als des bisherigen Mediums des Mediums, nämlich die Sprache: Es dominiert nun eine Permanenz des (Ein-)Schreibens.*

---

<sup>13</sup>Vgl. H.-J. Rheinberger, B. Wahrig-Schmidt und M. Hagner, 'Räume des Wissens. Repräsentation, Codierung, Spur.' (1997), S. 8.

r) einen Satz aus. Um nun den kompletten Paragraphen, die größte zusammenhängende Texteinheit von **txtkit** einzusehen, also den Kontext des gerade gefundenen Satzes, gibt man `readp` bzw. `rp` plus die Nummer des Ergebnisses an. Nun befinden wir uns im Paragraph.

Alle vorhergehend aufgeführten Befehle können ebenfalls in dem, für ein intuitives und ergonomischeres Bewegen im Text - implementierten `ncurses` Interface - durch einfache Tastaturbefehle (wie z.B. `links`, `rechts`, `enter`) direkt an den Ergebnissen durchgeführt werden. Um den eingangs erwähnten `pool` zu benutzen wird ein neuer Befehl zum Selektieren von bestimmten Sätzen bzw. Paragraphen ausgeführt (`select` bzw. `s`). Diese Auswahl wird im `pool` gespeichert und kann dann per spezieller Funktionen neu arrangiert werden. Bereits existierende Pools können Neueinsteigern als Orientierungshilfe verwenden. Per Taste `g` können diese eine Suchanfrage mit einem ausgewählten Satz starten und erhalten eine dazu nach Relevanz sortierte Ergebnisliste. Die Ansicht des `pools` kann, wie eben beschrieben, während einer Lektüre jederzeit aufrufen werden, um die Segmente zu ordnen, zu verschieben und zu löschen. Dennoch wird diese Form des Editing nicht explizit markiert oder durch graphische Zeichen kenntlich gemacht. Der `pool` kann jederzeit als Textdatei gespeichert werden.

## 5.2 UNIX as Literature

Mit Blick auf die Nutzungsgeschichte des Desktopcomputers sei bemerkt, dass die Handhabung einer Shell die mögliche, aber leidige Unterscheidung von User und Experte widerspiegelt. Denn mit der Erfindung mausklickbarer Bildschirm-Piktogramme in den 70er Jahren durch Alan Kay konnte der Nutzer den Computer bedienen, ohne ihn weitgehend verstehen zu müssen, genauer ohne ihn programmieren zu können. Die grafische Nutzung wurde mehr und mehr von der schriftsprachlichen Programmierung abgekoppelt, was zur Folge hatte, dass sich aufgrund dieser unterschiedlichen Steuerungskompetenzen entsprechende Nutzergruppen herausbildeten.<sup>14</sup>

*Das hat natürlich eine Grenze bei Medien, deren Systemarchitektur für Nicht-Eingeweihte vollkommen undurchdringlich geworden ist und wegen deren rigider interner Programmierung, die zugleich eine externe Dominanz von Eliten jenseits der Kunst etabliert, Dilettantismus schiere Ohnmacht und lächerliche Wehklage wird.*

Diese graduelle Abstufung von Fachwissen ist in unserem Zusammenhang von besonderem Interesse, da sich an ihr weitere Schlüsse bezüglich der medialen Dimension von Software festmachen lassen. Florian Cramer weist darauf hin, dass die Unterscheidung zwischen grafischer und schriftsprachlicher Benutzung den Status des Anwenders wesentlich bestimmt. Während mit Hilfe einer Shell bereits Anweisungen, wie Schleifen oder Zuweisungen ausgeführt werden können, welche den Zugriff auf das gesamte Dateisystem ermöglichen, beschränkt sich die Bedienung einer graphischen Benutzeroberfläche auf die Manipulation einer standardisierten Organisationsstruktur des jeweiligen Betriebssystems. Zudem können mit der Kommandozeilensprache einer Shell viele Probleme gelöst werden, ohne eine weitere Programmiersprache zu benutzen.

<sup>14</sup>Florian Cramer, 'Exe.cut[up]able Statements: The Insistence of Code', in: *Code. The Language of our Time, Ars Electronica 2003* (Osterildern-Ruit, 2003), S. 108f.

---

Die Offenheit der schriftsprachlichen Bedienung innerhalb einer Shell lässt sich selbstverständlich nicht auf alle inzwischen heterogenen Gebrauchszusammenhänge des Computers übertragen. Dennoch steht die Shell für die Tatsache, dass Computersoftware aus Schrift gefertigtes Werkzeug ist und dass Software die Trennung von Werkzeug und Schrift, wie sie die mechanische Schreibmaschine noch inkorporiert, endgültig aufhebt.<sup>15</sup> Zudem ruft die Kommandozeilensprache einer Shell in Erinnerung, dass Programmierung keine vom System losgelöste Anwendung ist, sondern eine Erweiterung der alltäglichen Nutzung, welche die dem Computer genuine Funktion Daten und ihre Prozessierung austauschbar zu machen, garantiert.

### 5.3 Schriftbild / Schriftsysteme

Die zur Lektüre des Textmaterials verwendeten Befehle ähneln Sprechakten, da sie eine Zustandsänderung bewirken und in der Regel an eine Instanz gerichtet sind, in unserem Fall an die Datenbank. Die Befehle sind keine inhaltsleeren Begriffe, denn ihre Bedeutung hat unterschiedliche Wirkungskraft. Der Nutzer startet mit dem Wortlaut `goto` eine Anweisung an die Datenbank, die von dem Programm **txtkit** ausgeführt wird. Die Bedeutung des Wortlautes `goto` übersetzt der Nutzer als Anweisung *Gehe zu....* Diese Bedeutung unterscheidet sich von der dadurch ausgelösten Ausführung der Anweisung durch das Programm.

Gabriele Gramelsberger hat in ihrer Studie zur "Digitalen Schrift", die Entkoppelung von Schrift und Sprache hin zum operativ-formalen Zeichengebrauch ausführlich skizziert.<sup>16</sup> In unserem Zusammenhang sei nur insoweit auf die unterschiedlichen Eigenschaften dieser Symbolsysteme hingewiesen, als dass die Commandline eben jene Grenze markiert, an der natursprachliches Textmaterial mittels formalisierten Handlungsanweisungen als schriftliche Instruktionen, bearbeitet werden kann.

Schrift wird in der traditionellen Sprachphilosophie als ein aus der Sprache abgeleitetes Symbolsystem verstanden, das als Abbild der Rede fungiert.<sup>17</sup> Das von uns digitalisierte Textmaterial setzt sich aus eben dieser Art von Schriftzeichen zusammen.<sup>18</sup>

Im Fall des Befehls `goto` werden Schrift und Sprache allerdings entkoppelt, die Schriftzeichen werden als operative Zeichen verwendet. Das wird möglich,

*Es ist ein prinzipielles Phänomen, dass Bilder nicht in dem Sinne Fragen stellen, Befehle erteilen oder Aufforderungen negieren können wie wortsprachliche Propositionen.*

---

<sup>15</sup>Cramer(Anm. 14), S. 108f.

<sup>16</sup>Vgl. Gabriele Gramelsberger, 'Semiotik und Simulation: Fortführung der Schrift ins Dynamische. Entwurf einer Symboltheorie der numerischen Simulation und ihrer Visualisierung', Dissertation, FU Berlin (Berlin, 2002), (URL: <http://darwin.inf.fu-berlin.de/2002/118/>).

<sup>17</sup>Vgl. a. a. O.

<sup>18</sup>David Smalls "The Talmud Project" ist ein praktisches Beispiel dafür, wie durch neue Displaytechnologien der zweidimensionale Zeichencharakter der Schrift aufgelöst werden kann. David Small, 'The Talmud Project: Rethinking the book', Dissertation, School of Architecture and Planning (MIT Media Lab, 1999), (URL: <http://acg.media.mit.edu/projects/theses.html>) – Zugriff am 2001.

indem die Zeichenfolge `goto` nicht länger allein auf einen von ihr bezeichneten Gegenstand referiert, sondern auf sich selbst, respektive auf ihre eigene Form und die damit verbundene Handlungsanweisung verweisen. `goto` konstituiert seine Funktion oder Bedeutung, nämlich eine von uns festgelegte Vorschrift, welche die Ausführung expliziter Regeln zur Folge hat. Symbolschriften bzw. Kalkülsysteme bedeuten eine Entkoppelung der Zeichen oder Symbole von den Gegenständen, hin zu den Verfahren, die sie konturieren, strukturieren und zugleich Objekte generieren.<sup>19</sup>

Sybille Krämer weist in ihrem Aufsatz "Textualität, Visualität und Episteme" darauf hin, dass Kulturtechniken, wie es schriftliche Zeichensysteme sind, nicht nur für den Erwerb, sondern auch für die Begründung von Wissen bedeutsam werden.<sup>20</sup> Sie macht das mediale Potential von Schriftbildlichkeit an der Tatsache fest, dass die Schrift ein hybrides Medium verkörpert, in welchem das Sagen und das Zeigen, das Diskursive und das Ikonische sich verschränken und wechselseitig zuarbeiten.

Die Entwicklung der Shell dient bei der Betrachtung des Computers als Schriftmedium dazu die Differenz von Schrift/Sprache versus Zeichen/Operation in den Blick zu bekommen und thematisiert, wie formalisiertes und technisch mediatisiertes Tun Eingang in Denkprozesse<sup>21</sup> findet.

#### 5.4 Schriftzeichen / Schlussweisen

Die Eingabe der Befehle wird in Echtzeit angezeigt, ebenso wie die Ausgabe der Ergebnisse. Die Befehle haben unterschiedliche Ausgabeformate. Eine `goto` Abfrage zeigt eine begrenzte Anzahl numerierter Satzfragmente an, die den eingegebenen Begriff enthalten. Um den vollständigen Satz, die kleinste gedankliche Einheit in **txtkit** lesen zu können, muss der Nutzer den Befehl `READ` in Kombination mit der Angabe der gewählten Satznummer eingeben. Erst nachdem der vollständige Satz erschienen und gelesen worden ist, kann die umschliessende gedankliche Umgebung des Satzes mit Hilfe des Befehls `Read-Paragraph` eingesehen werden. Nach Auftauchen des Satzes bzw. Paragraphen können die einzelnen Zeilen durch Betätigen der Pfeiltasten durchsprungen werden. Diese Funktion ermöglicht dem Nutzer weitere Schlüsselwörter aus den Textsegmenten zu selektieren und ihrem Vorkommen nachzugehen.

Mit diesem Vorgehen verkehrt sich das Verfahren der Verschriftlichung in sein Gegenteil. Entlastet die Verschriftlichung den Nutzer von memorativen Zwängen, geht es hier eben nicht um eine Technologie des Vergessens-Könnens,

---

<sup>19</sup>Vgl. Sybille Krämer, 'Kalküle als Repräsentation. Zur Genese des operativen Symbolismus', in: Hans-Jörg Rheinberger u.a. (Hrsg.), *Räume des Wissens. Repräsentation, Codierung, Spur* (Berlin, 1997).

<sup>20</sup>Vgl. ders., 'Textualität, Visualität und Episteme. über ihren Zusammenhang in der frühen Neuzeit', in: Stefan Rieger Renate Lachmann (Hrsg.), *Text und Wissen, Technologische und anthropologische Aspekte* (Tübingen, 2003).

<sup>21</sup>Vgl. ders., 'Kalküle als Repräsentation. Zur Genese des operativen Symbolismus'(Anm. 19).

---

sondern um die des Memorierens, der Aktualisierung zeitlich verschobener Gedankenprozesse. Durch die jederzeit mögliche Aktualisierung löst sich die Schrift von ihrem Übermittler. Das Auftauchen der Wörter in der Shell stellt die Bedingung der Möglichkeit des Funktionierens von Sprache dar. Denn die lose gekoppelte Menge sich unentwegt neu generierende Menge von Wahrnehmungen fordert dem Nutzer eine entscheidende Handlung ab, auf der Wissenserwerb im Wesentlichen beruht: das Unterscheiden oder Schliessen.

Indem der Nutzer seine Beobachtungen unterscheidet und sich für bestimmte Operationen entscheidet, hat er den Dingen, die er auf der Shell liest, eine Bezeichnung gegeben. "Nichts und niemand kann etwas bezeichnen, ohne es zu unterscheiden. Nichts und niemand kann Unterscheidungen einführen, ohne irgendeine Art der Bezeichnungsleistung vollzogen zu haben. Dies ist nicht dies ohne das, und das ist nicht das ohne dies."<sup>22</sup> Beobachtungen sind auf Dedikationen angewiesen und fungieren daher als dezidierte Operation.

In Peirce semiotischer Theorie der Kognition nimmt der Mechanismus des Unterscheidens eine besondere Stellung ein. In ihm vollzieht sich eine der drei Arten des schlußfolgernden Denkens, welche uns durch die Betrachtung dessen, was wir bereits wissen, etwas anderes herausfinden lassen, was wir nicht wissen<sup>23</sup>. Neben der Induktion und der Deduktion führt Peirce als dritte Art des Schlußfolgerns die Abduktion ein, ihr fällt im Hinblick auf den Wissenserwerb eine besondere Rolle zu.

Der Akt des Lesens geht darin auf Unterscheidungen vorzunehmen und zu schliessen, um anschliessend zu einer weiteren Aussage zu gelangen. Dieser Vorgang ist nichts anderes als der Prozess des Findens bzw. Erfindens einer Hypothese (Abduktion), um den Aussagen innerhalb eines Bezugsrahmens (Codierungsregel) in unserem Fall "Kunst als Medientheorie", Bedeutung zuzuweisen. Die Abduktion ist also das Prinzip, das zu rekonstruieren erlaubt, wie der Nutzer mittels situativer, spontaner und momentaner Hypothesenbildung (in Form einer Mini-Theorie bzw. einer Regel- bzw. Gesetzeshypothese) konstruktive und konzeptuelle Ordnungen erzeugt.

Jedesmal wenn der Nutzer eine Operation in der Shell vornimmt, hat er zuvor eine nicht gegebene Idee eingeführt und weist den ungeordneten Aussagen eine Bedeutung zu.<sup>24</sup> Diese Einführung und Einschreibung von Ideen vollzieht sich bei **txtkit** vor allem im Medium der Wahrnehmung, die durch beobachtbare Aufzeichnung gegeben ist.

Die Shell allein kann dieses prozessierende System der Differenzierung nicht ausreichend in den Blick bekommen. Hierzu entwickeln wir ein von dem Schriftsystem unabhängiges Aufzeichnungswerkzeug, mit dem Ziel weitere Unterscheidungen zu generieren.

---

<sup>22</sup>Peter Fuchs, 'Die Signatur des Bewusstseins', in: Jörg Huber (Hrsg.), *Interventionen 12. Person/Schauplatz*, 123 - 137 (Wien/New York, 2003), S. 126.

<sup>23</sup>Vgl. Charles Sanders Peirce, *Schriften zum Pragmatismus und Pragmatizismus* (Frankfurt a.M., 1991), S. 152.

<sup>24</sup>Vgl. a. a. O., S. 333.

Obwohl oder gerade weil sich Lesen im Medium der Schrift vollzieht, fügen wir eine weitere Ebene der Repräsentation ein, die ebenso um die symbolische als auch um mentale und kognitive Vorgänge des Unterscheidens und Schliessen kreisen. Durch die Einführung visueller Aufzeichnungsmethoden erhoffen wir uns zudem die unterschiedlichen Perspektiven von interessegeleiteter Wahrnehmung, in den Blick zu bekommen.

## 5.5 Bilder der Wissenschaft

Das für **txtvbot** entwickelte Notationssystem verwendet abgeleitete Formen einer Helix. Wir entwickelten diese dreidimensionale Struktur, weil sie in unseren Augen, sowohl die Eigenschaften der textuellen Information, als auch die Handlungen der Nutzer adäquat abbilden kann. Ausserdem sollte eine Analogie zu zweidimensionalen Speicherträgern wie Buchseite, Tafel oder Papier zu vermeiden.

Vielen Nutzern wird die Form der Zirkelschnecke vor allem an das Modell der DNA-Struktur erinnern, das Anfang der 50er Jahre von James Watson und Francis Crick entworfen wurde. Mit Hilfe eines räumlichen Drahtgitter-Modells formulierten die beiden Wissenschaftler die Hypothese der molekularen Struktur der Chromosomen. Die ästhetischen Merkmale ihres Modells waren für die Wirkungskraft und Etablierung im wissenschaftlichen Diskurs sicherlich genauso ausschlaggebend, wie seine wissenschaftliche Exaktheit.<sup>25</sup>

Vergleicht man heutige Wissenschaftsberichte über die Genetik fällt auf, dass das Thema der Erbinformation und des genetischen Codes vermehrt mit Konzepten des Textes, des Buches, des Alphabets und der Information in Beziehung gebracht wird. Dass im Fall der Genetik vermehrt die Textualisierung des Organischen verfolgt wird, zeigt einerseits die Wandelbarkeit wissenschaftlicher Bedeutungskonzepte und andererseits, dass die Formierung und Regulierung der Diskurse nicht zuletzt auf der Aneignung fremder Sprachkonzepte basiert. Wie Sigrid Weigel in ihrem Aufsatz "Der Text der Genetik" deutlich macht, sind experimentelle Operationen und theoretische Hypothesenbildung immer schon mit dem metaphorischen Diskurs verbunden.

Auch in unserem Zusammenhang sollen die Kriterien der Entwicklung eines Repräsentationsmodells nicht verschwiegen werden. Die Verwendung einer Helix im Rahmen von **txtkit** zielt auf eine dynamische Darstellung, die sich am ehesten in kulturelle Konzepte des Organischen oder Lebendigen festmachen lassen. Trotzdem besteht die praktische Leistung von **txtkit** darin, die Helix Struktur funktional umzubauen.

Genauso zeigt die zunehmende Popularisierung des Textparadigma von genetischen Codes, dass die Produktion wissenschaftlicher Symbolräume und Bedeutungssysteme aus einer langen Kette von Übersetzungen kultureller Konnotationen und Metaphern besteht, welche die Grenzen der Wissenschaften, aber auch

---

<sup>25</sup>Vgl. Reichle.

---

die der Künste verschieben und überschreiten. Um diese komplexen Herstellungs- und Transformationsprozesse in den Blick zu bekommen, untersuchen die textbasierten Wissenschaften die Diskursgeschichte (natur)wissenschaftlicher Modelle. Im Unterschied dazu markiert **txtkit** im Rahmen einer praktischen Theorie der Künste eben diesen Weg von den Dingen zu den Worten oder zu den Bildern und umgekehrt.

## 5.6 Monitoring Applikation txtvbot

Der `txtvbot` zeichnet die Ereignisse, die innerhalb des `txtshell` stattfinden, unmittelbar auf. So wird jede Eingabe graphisch repräsentiert, die visuellen Eigenschaften geben Aufschluss sowohl über statistische, als auch inhaltliche Parameter. Die Darstellung des Befehls `goto` erscheint als Fächer, dessen Anzahl der Strahlen mit der Anzahl der Ergebnisse übereinstimmt, und zusätzlich durch die entsprechende Größe roter Polygone die jeweiligen Relevanzen einzelner Ergebnisse anzeigt.

Wählt der Anwender mittels des Befehls `read` (oder `r`) einen Satz aus, wird eine Helix visualisiert, deren Radius und Länge mit der Länge des jeweiligen Satzes korrespondiert. Zudem lassen sich daran die Schlüsselwörter des Satzes ablesen, sowie von welchem `goto` Strahl, das Ergebnis ausgewählt wurde. Auch hier wird die Stärke der Relevanz durch rote Polygone angezeigt.

Die Ansicht eines Paragraphen unterscheidet sich von der eines Satzes zwar nicht wesentlich in der Form, aber proportional: die Skalierung seiner numerischen Daten läßt den Paragraphen kleiner und proportional kürzer als den Satz dargestellt, was einer Massstabsverschiebung gleicht: der Satz erscheint wie durch ein Mikroskop *vergrößert* und damit detailreicher als der Paragraph. Somit rückt die Menge der Information eines Paragraphen in den Hintergrund, nicht zuletzt weil auch die Schlüsselwörter nicht angezeigt werden. Stattdessen ist die Länge der Sätze durch kleine Markierungen an der Helix gekennzeichnet, um den *Rythmus* eines Paragraphen zu vermitteln, also wie die Argumentation innerhalb eines Paragraphen verläuft. Selbstverständlich wird auch die Selektion eines Paragraphen gekennzeichnet: dieser färbt sich dann blau.

Am Anfang einer jeden `readp`-Helix tritt frontal eine weitere rotfarbene Helix hervor, die dem Anwender anzeigt, welche Schlüsselwörter andere Nutzer gelesen haben, die ebenfalls den zuvor angezeigten Satz gelesen haben. Auf diese Weise entstehen inhaltliche Querverbindungen zu anderen Lektüren.

Wird ein neuer Befehl zur Selektion von bestimmten Sätzen bzw. Paragraphen eingeführt (`select`, schliesst sich die entsprechende Helix durch eine transparente, blaue Fläche).

Der visuellen Anordnung liegt eine zeitliche Ordnung zugrunde: alle Ereignisse werden nacheinander aufgezeichnet, gleich einem Lesepfad, der auf eine Gesamtsicht oder eine strukturelle Darstellung verzichtet. Hiermit soll die explorative Erkenntnisgewinnung gefördert werden, bzw. die Wechselwirkung mit der maschinellen Analyse unterstützen. Letzteres liefert wie oben erläutert die

gesammelten, kollektiven Gebrauchsspuren und verhindert so, dass die Visualisierung als simple reaktive Grafik verkannt wird.

Neben der visuellen Repräsentation der statistischen bzw. numerische Größen und der Notation der Ereignisse, liefert der txtvbot zusätzliche Informationen. Jedesmal, wenn der Leser an einer Stelle verharrt, liefert der txtvbot alternative Textstellen, die im Zusammenhang mit bisherigen Anfragen des Nutzer stehen. Somit kommt es zu einer permanenten Durchdringung von maschineller und inhaltlicher Analyse, wobei Leser und Software unterschiedliche Verfahren einsetzen.

Ein wichtiger Bestandteil der explorativen Lektüren ist die Beobachtung des txtvbot. Die drei-dimensionale Grafik rotiert kontinuierlich um die eigene Achse, so dass immer andere Details der Visualisierung in den Vordergrund geraten. Ausserdem kann von einer Seiten-Ansicht auf eine Frontal-Ansicht manuell gewechselt werden. Eine stufenlose Verschiebung der Helix ermöglicht es, diese näher oder ferner zu rücken, bzw. von vorn nach hinten oder umgekehrt zu durchfahren. So geraten unterschiedliche Begriffe, Konstellationen und Relevanzen in den Fokus. Dennoch entscheidet nicht zuletzt der Betrachter, welche Valenzmuster er situationsgebunden anwendet.

*Die Benennbarkeit der Dinge und das semiotische Universum von sukzessiven Bedeutungsveränderungen durch Montage schaffen vielgliedrige Bezüge zwischen Bild und Text, Text und Text, Bild und Begriff, Begriff und Text, Bild und Bild: Zeitsprünge eröffnen Bedeutungsräume.*

Hier wiederholt sich die logische Figur der Peirceschen Abduktion erneut, in diesem Fall findet der kreative Prozess der Hypothesenbildung auf non-verbaler Ebene statt. Wie im Zusammenhang mit schriftsprachlichen Zeichen, verlangt der txtvbot dem Leser eine selektive Beobachtung ab, um den jeweils bestmöglichen Schluss aus der visuellen Darstellung zu ziehen. Dabei spielt es eine entscheidende Rolle, dass die von uns festgelegten grafischen Parameter trotz statistischer und numerischer Faktoren, dem Leser zu Beginn undefinierbar erscheinen. Genau dieses Wahrnehmungsstadium ist eine schnelle und unkontrollierbare Phase der Wissensproduktion. Erst durch die regelmäßige Beobachtung, durch wiederholte Anwendungen durch Testen und Versuchen werden Muster und Abweichungen erkannt, erwachsen dem spontanen, hypothetischen Schliessen und seinen partiellen und vorläufigen Deutungen mehr und mehr nachvollziehbare Korrelationen.

## 6 Ergebnisse

Obwohl die Textkit Helix auf numerischen Daten basiert, hat die Aufzeichnung der Signale und Daten nicht die Evidenzherstellung im Sinne von Ausschliesslichkeit zum Ziel. Das selbst bereits technisch erzeugte Modell der Helix, fungiert als symbolischer Darstellungsraum der zunächst und vor allem eine heuristische Funktion hat. Es besteht aber keine Identität zwischen sinnlich wahrnehmbarem Darstellungsobjekt, als Teil eines Symbolsystems und dem Gedanken oder der schriftlichen Fixierung, dem Objektwissen. Wir haben keinesfalls die Absicht naturwissenschaftliche Erkenntnisstrategien zu übernehmen, sondern wir möchten die Variabilität dieser Konstruktionen aufzuzeigen.

---

Betrachten man den Wissenserwerb im Kontext experimenteller Wissenschaften fällt auf, dass neben der Bestimmung der Verfahren und ihrer apparative Organisation, der Einsatz bildgebender Verfahren eine wesentliche Aufgabe in der Begründung von Wissen zu kommt. So setzt sich eine Versuchsanordnung aus praktischen Operationen, meist Verfahren der Trennung und Verbindung, zusammen, die eine Reaktion auslösen. Diese Ereignisse geben erst in der Wiedergabe ihrer Beobachtung Auskunft über den Sachverhalt und ermöglichen so seine Bewertung.

Wie oben beschrieben übernimmt **txtkit** die Zerlegung des Materials, stellt die Methoden zur Bearbeitung der Textsegmente zur Verfügung und gibt dem Nutzer nun die Visualisierung der Operationen und Ereignisse als Werkzeug zur Beobachtung an die Hand.

**txtkit** bricht mit der Evidenz und Totalität schriftbasierter Quellen als einzig relevante materielle und ideelle Fixierung von Wissen. Die Schrift wird ihrem gewöhnlichen Gebrauchszusammenhang entzogen und in Verbindung mit einem technisch erzeugten Darstellungsverfahren Teil eines auf Erfahrung begründeten Wissens.

Damit wird die Realität der Schrift insoweit in Frage gestellt, als das sie den Zwängen ihres Gebrauchs entbunden wird und die Bedingungen und Modalitäten zur Generierung von Sinn und Bedeutung (Versinnlichungen von Bedeutung, Repräsentation) in neuen apparativen Zusammenschlüssen verschoben werden. Die Auslegung von Schrift mündet nicht in einer statischen Festschreibung, sie ereignet sich als prozessuale und offene Denkbewegung, welche eben nicht allein auf der Linearität von Schrift basiert, sondern durch die Beobachtung dynamischer Abbildungsverfahren beeinflusst wird.

Die in **txtkit** realisierte Versuchsanordnung ermöglicht dem Nutzer selber Verfahren zur Bearbeitung des Materials zu entwickeln. Beobachtung, Spurensuche und Editing vollziehen sich in einem apparativen Zusammenhang, der offen läßt, was geschieht. Der Nutzer greift in das Geschehen ein, ohne vorhersehen zu können, wie das Ergebnis ausfallen wird.

Erst das bildgebende Verfahren der Helix bildet die Bewegungen des Nutzers ab und spiegelt weitere Ebenen der Bedeutung wider. Der Nutzer kann die argumentativen Verbindungen zwischen den einzelnen Textaussagen herleiten, seine Lektüre wird durch die Sicht auf die Dinge gelenkt und fortgeführt. Dies geschieht vor allem immer dann, wenn der Nutzer Unterscheidungen vornimmt, Bezeichnungen installiert, die jedoch immer nur vorläufig sind, da das Ende der Lektüre nicht vorher gedacht werden kann. Auf Experiment und Beobachtung folgt eine Deutung, die im Rahmen von **txtkit** nicht nur nie endgültig ist, da sich der Blick auf das Gesamte verschliesst und die Verbindungen zwischen Aussagen Bedeutungen konturieren.

Die Qualitäten experimenteller Versuchsanordnungen kommen im Rahmen von **txtkit** auf vielfältige Weise zum Ausdruck. Die **txtkit** Helix wird aus gesammelten numerischen Erhebungen von Material und Lektüren generiert. Zum einen werden die Phänomene und deren Beobachtung in der jeweiligen Gestalt

der Helix verkoppelt. Der Betrachterstandpunkt von **txtkit** hat keinen allumfassenden Blick auf die Sammlung und Strukturierung der Daten. Im Gegenteil, der Nutzer kann erst durch die kontinuierliche Eingabe von Befehlen Textstücke freigraben. Durch die Beobachtung der Helix, die immer nur Satzeinheiten oder aber Paragraphen anzeigt, werden Aussagefelder generiert, freigelegt und neu montiert. Dieses Vorgehen unterscheidet sich beispielsweise von Hypertextdiagrammen, die mittels Baumdiagrammen komplexe Strukturen eines Textkonvoluts in hierarchischen und kausale Verknüpfungen anzeigen. Solche Metasprachen suggerieren dem Nutzer über eine Gesamtsicht des Textes und über seine internen Verweis- oder Verknüpfungsstrukturen zu verfügen.

Im Fall von **txtkit** verweigern sowohl die Visualisierung als auch die Shell eine zentrale, dem Material übergeordnete Position. Beide Bereiche weisen durch inhaltliche und zeitliche Brüche auf, welche durch die Aussagen zerstreut werden. Funktion des Werkzeuges **txtkit** ist nicht die Rekonstruktion eines Textes mit Anfang und Ende, sondern die Erschließung eines Datenraums<sup>26</sup>. Die Visualisierung als Helix übernimmt dabei die zentrale Aufgabe das Ordnen, Sammeln zu verzeichnen.

*Ein Zugriff auf technologische Medienmaterialitäten ist nicht möglich ohne Aufbau des Interfaces, d. h. einer Kontaktmaterie zu einem Betrachter, die zugleich ein Schnitt durch die Dispositive ist, in denen sich die Aktualzustände der angestrebten und verfolgten Experimente aufzeichnen lassen.*

Im Unterschied zu naturwissenschaftlichen Versuchsanordnungen dient die Beobachtung und Deutung der Visualisierung jedoch nicht dazu Gesetzmäßigkeiten auszumachen. Stattdessen wird die Faktizität des Materials durch den situativen Charakter gefärbt. Die Tatsache, dass Aussagen im Laufe einer Lektüre vermehrt auftauchen können, zeigt auch, dass statt einer Vereinheitlichung der Aussagen der Modus ihres jeweiligen Erscheinens in den Mittelpunkt treten. Denn erst bei der Vervielfältigung serieller Lektüren treten schließlich die Singularitäten und Muster hervor.<sup>27</sup> Baumdiagramme bleiben mit ihren strukturalen Abbildung des Textes der Dominanz des Textes verhaftet. Selbst die grafischen Attribute, Linien und Verbindungen dienen dazu, das Schriftbild zu einem zwar komplexen, aber vereinheitlichten Bild zu verbinden.

Die **txtkit** Helix stellt an sich bereits eine Übersetzungsleistung dar. Diese oben beschriebene Transformation ereignet sich mit Hilfe der von uns zuvor festgelegten Regeln zur grafischen Darstellung. Die quantitativen Symbole fungieren keineswegs ausschliessend/exklusiv, sondern sie geben dem Nutzer ein selbst zu organisierendes Modell an die Hand. Das bedeutet, dass das Wissen des Nutzers integrativer Bestandteil der Lektüre mittels **txtkit** wird. Im Unterschied zu naturwissenschaftlichen Modellen geht es nicht um exklusive Formen der Repräsentation, sondern darum die Entstehung, den Wandel, und die Widersprüche von Aussagefelder zu zeigen und dies ohne Rücksicht auf Fragen der empirischen Begründbarkeit in Erwägung zu ziehen.

<sup>26</sup>Vgl. M. Serres und N. Farouki, *Thesaurus der exakten Wissenschaften* (Frankfurt a. M., 2001), S. 252.

<sup>27</sup>Vgl. Zur Methode von Michel Foucaults Archäologie des Wissens Wolf Kittler, 'Thermodynamik und Guerilla. Zur Methode von Michel Foucaults Archäologie des Wissens', *Trajekte. Newsletter des Zentrums für Literaturforschung Berlin* 4 (2002), S. 17.

---

## 6.1 Der pool als Montage

Obwohl die Software die dynamischen und prozessualen Eigenschaften der Aneignung von Theorie im Blick hat, haben wir doch ein Ergebnis im Sinne eines singulären Produktes eingeführt: die Montage der gelesenen bzw. ausgewählten Textsegmente als Fließtext kann am Ende einer jeden Lesesession gespeichert und ausgedruckt werden. Nun könnte die Erwartung geweckt werden, dass das als Einzelteile in der Datenbank abgelegte Textmaterial durch die Lektüre als variable Rekonstruktion im pool vorliegt. Unterschiedliche Lesarten würden entsprechende Textvarianten generieren und damit dem Prinzip der argumentativen Geschlossenheit folgen. Dies ist jedoch nicht der Fall. Die Sammlung von Textsegmenten ersetzt keinesfalls herkömmliche Textquellen und deren Funktion. Auch ist der pool kein Beispiel für die Anwendung der Cut-Up Methode.<sup>28</sup> Statt Text und Sinn zu produzieren, protokolliert der pool die Handlungen des Lesers. Im Unterschied zu der Helix geschieht dies im pool allerdings nicht graphisch oder mit Hilfe der Befehle, die Sätze oder Paragraphen erscheinen vielmehr in ihrer temporalen Abfolge.

Da der Eindruck vermieden werden soll, dass es sich bei dem pool-Fliesstext um eine individuell erstellte Version des Textes handelt, der eine eigene, neue Argumentationslinie verfolgt und damit wie ein vermeintliches Autorenprodukt funktioniert, haben wir keinerlei Formatierungs- und Kommentarmöglichkeiten an die Hand gegeben. Der Leser kann die Ansicht des pool während seiner Lektüre jederzeit aufrufen, die Segmente ordnen, verschieben und löschen. Dennoch wird diese Form des Editing nicht explizit markiert oder durch graphische Zeichen kenntlich gemacht.

Es ist klar, dass die montierten Textstücke inhaltliche Lücken und Sprünge vermitteln, und somit den Status des Unfertigen haben. Dieser Mangel verhindert allerdings, die Textmontage als eigentlichen Zweck zu verfolgen und die Funktion der Software als Schreibwerkzeug zu verkennen. Dadurch dass die Anzeige des Pools nicht formatiert ist und der Leser diese auch nicht vornehmen kann, sollen die üblichen Regeln des Schreibens und des Editierens erst gar nicht zum Einsatz kommen und suggerieren, dass die Software Texte bzw. Textvarianten erzeugt. Ebenso wie eine Wegbeschreibung nicht das Abbild einer Landschaft ist oder als Gesamtansicht eines Areals dient, fungiert der pool allein als Protokoll der vorgenommenen Entscheidungen, welche den Lesepfad gelenkt haben. Die Ausgabe der selektierten Sätze und Paragraphen dokumentieren den Fluss des Handlungsverlaufs. Visuell spiegelt sich dieser in dem fortlaufenden Textbild wieder, tatsächlich demonstriert aber die Bedeutungsebene der Sätze, dass der Fluss der Gedanken, Hypothesen und Entscheidungen nicht die Stringenz eines verfassten Textes zum Ziel hat, sondern als unbearbeitetes Rohmaterial vorliegt.

Genau wie schon der Lektürepfad des txtvbots, erfüllt der pool seine Funktion erst, wenn die Verbindungen, die zwischen den Textsegmenten ge-

*Auch der stetige Wandel der gesprochenen wie geschriebenen Sprache scheint es nicht zu erlauben, in einem unwandelnbaren textlichen Code Informationen niederzulegen.*

---

<sup>28</sup>Vgl. N. Wardrip-Fruin und N. Montfort, *The New Media Reader* (Cambridge, Mass., 2003), S. 89.

legt wurden, als solche auch nachvollziehbar erscheinen. Paradoxien und Widersprüche sind dabei nur eine von vielen möglichen sprachlichen Indizien für die unterschiedliche Aussage- und Wirkungskraft der Gedanken, die nun zusammengefügt werden können. An das was anhand der Helix als visuelle Konstruktion zeitgleich beobachtet werden kann, nämlich der Effekt- bzw. Ereignischarakter von Bedeutungen<sup>29</sup> und ihren Anschlüssen, erinnert die schriftbasierte Montage nur nachträglich. Wie erfolgreich die Verschiebungen, die während der Lektüre stattgefunden und die Montage generiert haben, nachvollzogen werden können, hängt ebenso von dem Grad der Wahrscheinlichkeit als auch dem strategischen Muster der Lektüre ab.

*Der Mechanismus der Selektion ist natürlich ideologisch und arbiträr. Interessant ist aber, dass der Prozess der Valorisierung, auf dem jede Auswahl aus kulturellen Leistungen beruht, selber im Zustand reiner Potentialität verbleibt und in ein lang andauerndes, für menschliches Zeitempfinden unvorstellbares Warten gezwungen ist.*

Bereits von anderen Lesern erstellte pools können als Ausgangspunkt oder Wegbeschreibung eingesehen und verwendet werden. Hierbei steht die vermeintliche Zufälligkeit der jeweiligen Fundstücke im Zusammenhang mit der Perspektive des einzelnen Lesers. Erst durch den möglichen aber nicht notwendigen Nachvollzug von Entscheidungen, warum ein anderer Leser diesen Satz und nicht jenen gelesen oder selektiert hat, treten die individuellen Verfahren und Methoden der Bewertung hervor. Anders formuliert: gerade weil die Generierung von Bedeutung und Aussagen in einer Vielzahl möglicher und subjektiver Entscheidung aufgeht, geraten die allgemein nachvollziehbaren dem Text übergeordneten Muster in den Blick des Lesers: das Wissen über das Wissen und wie es gebildet wird. Die gesammelten Textfragmente funktionieren als entsprechenden Attraktoren, welche auf den Verlauf, den Rhythmus, die Richtungswechsel hinweisen.

Zusammenfassend formuliert haben wir den pool als eine weitere Darstellungsebene des Leseprozesses eingeführt, nicht zuletzt mit dem Ziel, die in der visuellen Konstruktion der Helix begonnene Arbeit des Lesers der Selektion in eine schriftsprachliche Montage zu übertragen.

## 6.2 Interaktivität

Die Funktion des Pools bringt uns zu der Frage, welche Grenzen aber auch Qualitäten die Software als Handlungsraum aufweist. Nachdem wir Interface und Pool im Hinblick auf die Bedienung und deren Auswirkungen für die Generierung von Wissen beschrieben haben, möchten wir an dieser Stelle das Potential der Software vor dem Hintergrund technischer Standardisierung des In- und Outputs des Systems hinterfragen, was unter dem Schlagwort Interaktivität verbucht werden kann.

Wie bereits oben beschrieben zeichnet sich die Visualisierung als unmittelbare und direkt nachvollziehbare Reaktion auf die Handlung des Users aus. Nun liegt die Besonderheit dieses graphischen Outputs darin begründet, dass die visuellen Eigenschaften zum einen durch die statistischen Erhebungen des Textmaterial bestimmt werden, zum anderen durch die Anwender modifiziert

<sup>29</sup>Vgl. Zum Modell des Lesens bei Paul Ricoeur Huber(Anm. 10), S. 55.

---

werden und somit nicht aus einer begrenzten Anzahl von visuellen Icons oder zeitlich begrenzter Darstellungsparameter, die sich womöglich identisch wiederholen könnten, bestehen. Die Tatsache, dass die visuelle Erscheinung numerisch variabel und sich durch Benutzung kontinuierlich verändert, macht deutlich, dass es uns nicht um die Entwicklung eines vorgegebenen, statischem Ordnungssystems geht, sondern um die Festlegung graphischer Parameter, welche die der Situation entsprechende Notation generiert. Situation meint hier sowohl die Auswirkung der jeweilige Beschaffenheit des Materials (konkret: lange Sätze, kurze Paragraphen), als auch die spezifische Form der Lektüre, im Besonderen deren temporale Struktur (viele Sprünge, kurze Abschnitte) aber auch die Dichte der inhaltlichen Suche.

Das Verhältnis von Software und Benutzer basiert sich im wesentlichen auf der Tatsache, dass die Bearbeitung des Materials einen Mehrwert generiert, den wir als Objekte des Wissens einstufen möchten. Abgesehen von dem singulären Geschehen, seinen Folgen und Ausformungen, ist das Resultat der Lektüre, dass die eingegeben Befehle, Entscheidungsfolgen etc. das System entscheidungslogisch verfeinern. Denn die Auswirkungen der Relevanzen machen das Vorgehen nicht voraussagbar, aber sie wirken entscheidend bei der Erzeugung von Begriffsfelder und Kontexten mit. Es gibt in dem Sinne auch keine Bewertung oder Hierarchie der Standpunkte, da alle Interventionen berücksichtigt werden und erst in der Erzeugung von relevanten Spuren<sup>30</sup> treten Valenzen als Lerneffekte auf.<sup>31</sup> Genau diese non-verbale Ebene der Kenntnisse kann ein System ohne die Steuerung durch die Anwender nur schwer liefern.

Tatsächlich greift der *txtvbot* die temporale Struktur der Lektüre auf, um die Ereignishaftigkeit und Aktualisierungen deutlich in den Gebrauch mit einzubeziehen. Dies geschieht nicht allein um die formale Erfahrbarkeit von Interaktivität zu erreichen, um "das Faktum, dass ich etwas am Objekt in Gang bringe - (zu erreichen)", wie Hans Ulrich Reck erläutert, "sondern (...) (damit) das 'Ich' des Betrachters selber zum Werkstoff der Installierung des Interface wird. Ich nehme eine Umwelt wahr, in die ich gleichzeitig eingeschrieben werde. Deshalb ist die Inszenierung", in unserem Fall das Interface der Applikation, "als eine wiederholende und explikative Figur von entscheidender Bedeutung."<sup>32</sup>

### 6.3 Gebrauchsszenario

Die Software ermöglicht dem Nutzer individuellen Lektürepfade zu folgen, argumentative Verbindungen zwischen den einzelnen Textaussagen herzuleiten, oder umgekehrt, Widersprüche und Paradoxien hervorzubringen. Eben diese vorläufigen Bedeutungen, die in einem ergebnisorientierten Verfahren verloren gingen, bleiben hier Teil einer zeitlich organisierten Struktur, welche keine formalen

---

<sup>30</sup>Vgl. *Spuren sind nicht nur Hinterlassenschaften von Ereignissen und Handlungen, sondern selbst Aktionen, im Französischen als Verb tracer genannt.* Huber(Anm. 10), S. 98.

<sup>31</sup>Vgl. Lernmodelle bei Piaget und Lewin a. a. O., S. 455.

<sup>32</sup>Hans Ulrich Reck, *Kunst als Medientheorie* (München, 2003).

Brüche aufweist, sondern nur gedankliche Sprünge und Richtungswechsel. In temporaler Sicht, wird Verstehen nicht als stabilen Zustand nach dem Gebrauch der Software erreicht, sondern die andauernde Suche nach Überlagerungen von Aussagen findet allein in und während der Benutzung statt. Dieser prozessuale Aspekt ist vor allem im Hinblick auf wiederholte Anwendung wichtig: denn nur in der Wiederholung kann von der ersten als finalen Interpretation Abstand genommen werden.

*Man kann deshalb das 'Poetische' auch ganz einfach als ein ausdrücklich wahrgenommenes Resultat solcher Abweichungen zwischen zwei Standpunkten oder Blickwinkeln verstehen.*

Der möglichen Reduktion der Software als individuelles Aufzeichnungsgerät singulärer Ereignisse können wir entgegensetzen, dass die Erhebung der Nutzer-Relevanzen **txtkit** Erfahrungswerte aufzeichnet und zu einer Ausdifferenzierung von Bedeutungen führt. Die maschinelle und inhaltliche Analyse bedingen sich wechselseitig und verweigern damit eine automatisierte und standardisierte Benutzung.

Der in **txtkit** angelegte Versuchsanordnung könnte der Vorwurf widerfahren, dass der Gebrauch der Software interesseloses, zufälliges Bedienen und Manipulieren von Textsegmenten begünstigt. Bestimmt man Zufall jedoch als dasjenige für dessen Existenz es keinen Grund gibt, dann spielt der Zufall als Kategorie der Lektüre eine untergeordnete Rolle. Denn nicht eine Begründung, sondern die Bedingung für das Auftauchen einer Aussage in einer bestimmten Konstellation sucht **txtkit** seinen Nutzer an die Hand zu geben. Es geht darum die Kontingenz des jeweiligen Informationsgehaltes einer Aussage mit dem Erfahrungshorizont des Nutzers in Verbindung zu setzen.

Sucht der Nutzer nach einer Begründung für die ein oder andere Aussage, folgt er geradezu dem Anti-Zufall. Nicht pauschales Zweckdenken, sondern die Ideenhypothesen leiten die Lektüre der Nutzer. Indem die Anwender Unterscheidungen vornehmen, Beobachtungen zuordnen, handeln sie und setzen damit Wirklichkeiten, die in ihrer Aktualisierung singulär werden. **txtkit** stellt dem Nutzer unvollständig Bestimmtes zur Verfügung, überlässt ihn damit aber nicht einer Totalität oder Kontinuität. Stattdessen bilden die Widersprüche der erstellten Settings, dass unterschiedliche Kräfte, die eine oder andere Aussage wirksamer erscheinen lassen als diejenigen, die nicht auftauchen.

## 7 Fazit

Dieser Bericht hatte zum Ziel die Entwicklung, den Aufbau und den Gebrauch der Software **txtkit** vorzustellen. Um unser Vorgehen im Hinblick auf die Aufbereitung und Organisation zu erläutern, sind wir zunächst auf die von uns gewonnenen Erkenntnisse bei der Analyse bereits existierender NLP Verfahren im Projekt Megatext eingegangen. In Rückwendungen auf die im Antrag formulierten Absichten und Voraussetzungen konnten wir darlegen, wie die Entmaterialisierung und Entäußerung der bisherigen Form und Struktur von Texten veränderte Methoden zur Aufbereitung und Nutzung von textueller Information möglich macht. Um einen fälschlichen Vergleich und eine anschließende Bewertung von

---

Publikationsformen zu vermeiden, galt es zunächst die unterschiedlichen Nutzungszusammenhänge hervorzuheben und die qualitativen Unterschiede des hier angewandten Verfahrens näherzubringen.

Drei Aspekte möchten wir am Ende herausstellen und ihr Potential für noch nicht realisierte Anwendungen skizzieren.

**1. Der Status der textuellen Information wird durch die Verfahren des Parsing und der Datenbankstruktur als zentrale, vernetzte Zugriffsquelle festgelegt.**

Will man die mediale Differenz im Hinblick auf andere Textquellen in den Blick bekommen, müssen die technischen Verfahren mit denen das Material aufbereitet und bearbeitet wird, berücksichtigt werden. Durch die variabel und wiederholt durchführbare Parsing-Methode werden Texte in ihre Einzelteile zerlegt und indexikalisiert. Die Entmaterialisierung macht hier die numerische Erfassung des Textes und statistischer Angaben möglich, was die Grundlage sowohl für die maschinelle als auch die inhaltliche Analyse und deren Visualisierung schafft. Zwar gleicht die zentral und via Internet zugängliche Textspeicherung anderen Textquellen im Internet, da sie unabhängig lokaler oder institutioneller Bedingungen von mehreren Anwendern parallel gelesen und bearbeitet werden können. Ein wesentliche Unterschied ist aber, dass die Datenbank die Interventionen der Akteure aufzeichnet, um sie in die Visualisierung einfließen zu lassen und damit auch alle nachfolgenden Lektüren beeinflussen. Somit erweitert die hier angelegte Versuchsanordnung statische Strukturabbildungen oder Archivfunktionen, vernetzte Publikations- oder Distributionsmedien um die Dimension der dynamischen und explorativen Wissensentwicklung.

Eine mögliche Weiterentwicklung sollte die Verfeinerung von Parsing-Verfahren und deren regulären Ausdrücke verfolgen, ebenso wie auch die Bezugnahme von Mustererkennung größerem Masstabes. Alternativ zu dem von uns realisierten Datenbankmodell würde sich die Verwendung dezentraler Organisationsstrukturen anbieten, mit denen sich Server-unabhängig informelle Textnetze installieren liessen.

**2. Die kollektive Wissensentwicklung basiert auf der Rückkopplung von maschineller und inhaltlicher Analyse und deren Visualisierung.**

Das in **txtkit** installierte Datenbanksystem sichert die Textdaten in Verbindung mit den Gebrauchsspuren der Anwender. Durch die Erhebung von Relevanzen wird die kontinuierliche und gemeinschaftliche Bearbeitung des Materials numerisch erfasst und durch den **txtvbot** visualisiert. Die Frage nach dem in der Software **txtkit** realisierten Modell gemeinschaftlicher Wissensentwicklung mündet somit einerseits in der nicht einsehbaren Struktur der Datenbank und andererseits in der visuellen Ausgabe der Daten in Form des **txtvbot**. Beide Bereiche formatieren singuläre Ereignisse und heterogene Einzelbeobachtungen, übertragen sie in das von uns festgelegte Ordnungssystem. Immer wenn der **txtvbot** die Interventionen des einzelnen Anwenders protokolliert, werden die Bewegungen anderer Nutzer sichtbar. Die Relevanzen und Related Informations bedeuten also keine verzichtbare Zusatzinformation, sie organisieren den

weiteren Verlauf der Lektüre. Weil die irreversible Aufzeichnung der Daten Beobachtung und Schlussweisen beeinflussen, handelt es sich bei dem interpersonalen Notationssystem jedoch nicht um eine Vereinheitlichung heterogener Standpunkte und Perspektiven, sondern um eine Vergemeinschaftlichung der Transformationsspuren.

**3. Die Versuchsanordnung vereint die Aufbereitung und die Erkundung textueller Informationen. Aufschluss über die externe Geltung dieser Objekte des Wissens, gibt die Software allerdings nicht.**

Die in **txtkit** angewandten Verfahren zur Protokollierung von Valenzen basieren auf den dem Nutzer zur Verfügung gestellten Operationen. Wie oben beschrieben, enthalten die aufgezeichneten Operationen immer schon die Bedingung der Möglichkeit ihrer Auswertung und Bedeutung in Form von Relevanzen. Dies hat zur Folge, dass sich die Aufzeichnungen nicht von ihrer Auswertung trennen lassen, und wie anhand der logischen Figur der Abduktion nachgewiesen werden konnte, sich wechselseitig beeinflussen. Bestimmt man nun die partielle Hypothesenbildung als ein wesentliches Verfahren explorativer Wissensentwicklung, stellt sich die Frage nach dem Stellenwert oder der Tragweite der Hypothesen im Hinblick auf die Erkenntnisgewinnung. Nicht ein übergeordnetes Deutungsschema oder Interpretationsmuster bestimmt die Geltung der hier gesammelten Wissensobjekte, sondern ihre Pragmatik bzw. Funktionalität im Rahmen einer sich kontinuierlich ändernden Textsammlung. **txtkit** ist ein System, das den Vollzug der Unterscheidung von Wahrnehmung und Verstehen Kontur verleiht und in einen neuen Darstellungsraum überführt, dies jedoch nicht um die Einheit von Beobachtung, Bewertung und Bedeutung zu behaupten, sondern um die medialen Differenzen dieser Bereiche in den Blick zu bekommen.

Der Versuch eine abschliessende Bewertung der Software **txtkit** vorzunehmen erscheint in unseren Augen aus zwei Gründen problematisch: zum einen liegt es für uns in der Natur der Sache, dass nicht allein der Entwurf und die Programmierung Aufschluss über Geltung und Wirksamkeit der Software geben, sondern vor allem die Zeit danach, ihre Verbreitung, Verwendung oder gar ihre Umfunktionierung.

Der zweite Grund liegt in der Anlage dieses Versuchs begründet: fragt man danach, welche Einsichten über das Verhältnis von Wissenschaft und Kunst auf Grundlage dieses Bausteins vorgenommen werden können, liegt es nahe, die Software sowohl aus den Augen eines Informatikers, als auch durch Experten der Kunst zu beurteilen. Dies möchten wir als Designer weder versuchen, noch erscheint es uns lohnenswert, denn beide Disziplinen weisen andere Bewertungsgrundlagen auf.

Dennoch ist die Entwicklung eines solchen Prototypens appellativ: Es ist klar, dass sich durch die Anwendung von Computern und der Verbreitung mathematischer Darstellung von Strukturen die Wissenschaften selbst verändern. Unser Blick wechselt vom Objektwissen zu Symbolsystemen: welche Maschine, welches Programm wird entwickelt, um Information herzustellen bzw. sich ihrer bedienen zu können? In diesem Gerangel darum, welche Repräsentation, Simu-

---

lation oder welches Denkmodell das Geeignete zur Erschliessung von Realitäten ist, darf sich jedoch auf keine Konkurrenzsituationen eingelassen werden, bei der die Künste womöglich fremde Spielregeln übernehmen und ökonomischen Prozessen unterliegen.

Stattdessen haben wir während der einjährigen Entwicklungszeit von **txtkit** gesehen, dass der Entwurf, die Setzung und Besetzung von Symbolen, Codes und ihrer Bearbeitung trotz der Technologie inhärenten Grenzen und Zwängen offene und eben-nicht totalisierende Aussagen und Handlungszusammenhänge generiert. Diese Qualität kann sich wiederum mit anderen wissenschaftlichen Anwendungen messen lassen, denn ihre Pragmatik dient keinem geschlossenen Wissensgebiet, sondern kehrt die Hierarchien um: nicht die Tatsachen bestimmen das Denken, sondern die Erwartungen der Anwender lenken das Potential des Werkzeugs.

## 8 Credits Software

**txtkit** 1.0  
Visual text mining tool

[www.txtkit.sw.ofcd.com](http://www.txtkit.sw.ofcd.com)

Copyright ©2003 Schoenerwissen/OfCD

Konzept: Schoenerwissen/OfCD, Hans Ulrich Reck  
Entwicklung, Design und Programmierung: Schoenerwissen/OfCD  
Mitarbeit Programmierung: Thomas Chille  
Beratung: Hans Ulrich Reck, Georg Trogemann, Thomas Chille, Jochen Viehoff

Includes and Libraries:

SMySQL Framework, Serge Cohen  
<http://mysql-cocoa.sourceforge.net>

CHARVA, Rob Pitman  
<http://www.pitman.co.za/projects/charva>

Cocoa Open Source, Karelia Software, LLC  
<http://cocoa.karelia.com>

NSFont Category for OpenGL, Bryan L Blackburn  
<http://www.withay.com/macosx/opengl/nsfont.html>

Connector/J, MySQL AB  
<http://www.mysql.com/products/connector-j>

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA.

---

## 9 Anhang

### Literatur

- Balke, Dr. Friedrich, 'Programm des Kulturwissenschaftlichen Forschungskollegs Medien und kulturelle Kommunikation an der Universität zu Köln' (URL: [http://www.uni-koeln.de/inter-fak/fk-427/prog/prog\\_fk\\_intro.html](http://www.uni-koeln.de/inter-fak/fk-427/prog/prog_fk_intro.html)).
- Cahn, Michael, 'Die Rhetorik der Wissenschaft im Medium der Typographie. Zum Beispiel die Fussnote', in: u.a., Hans-Jörg Rheinberger (Hrsg.), *Räume des Wissens. Repräsentation, Codierung, Spur* (Berlin, 1997), S. 91 – 111.
- Cramer, Florian, 'Exe.cut[up]able Statements: The Insistence of Code', in: *Code. The Language of our Time, Ars Electronica 2003* (Osterfildern-Ruit, 2003).
- Fischer-Lichte, Prof. Dr. Erika, 'Gesamtkonzept Kulturen des Performativen, Sonderforschungsbereich FU Berlin' (URL: [http://www.sfb-performativ.de/seiten/frame\\_impresum.html](http://www.sfb-performativ.de/seiten/frame_impresum.html)).
- Fuchs, Peter, 'Die Signatur des Bewusstseins', in: Huber, Jörg (Hrsg.), *Interventionen 12. Person/Schauplatz*, 123 - 137 (Wien/New York, 2003).
- Gramelsberger, Gabriele, 'Semiotik und Simulation: Fortführung der Schrift ins Dynamische. Entwurf einer Symboltheorie der numerischen Simulation und ihrer Visualisierung', Dissertation, FU Berlin (Berlin, 2002), (URL: <http://darwin.inf.fu-berlin.de/2002/118/>).
- Håkansson, M. und Lindqvist, M., 'Megatext. Visualising Semantic Relations', Diplomarbeit, Computational Linguistics, University of Göteborg (2002), (URL: [http://www.cling.gu.se/~c18mhaka/MegaText\\_compressed.pdf](http://www.cling.gu.se/~c18mhaka/MegaText_compressed.pdf)).
- Huber, Joachim, *Urbane Topologie. Architektur der randlosen Stadt* (Weimar, 2002).
- Kanellos, Ioannis, 'Texte et intertextes, concepts de médiation entre les nouvelles technologies et le requisit d'une assistance à la compréhension', in: Sadin, Eric (Hrsg.), *éc/artS 3* (Paris, 2002).
- Kittler, Wolf, 'Thermodynamik und Guerilla. Zur Methode von Michel Foucaults Archäologie des Wissens', *Trajekte. Newsletter des Zentrums für Literaturforschung Berlin* 4 (2002).
- Krämer, Sybille, 'Kalküle als Repräsentation. Zur Genese des operativen Symbolismus', in: u.a., Hans-Jörg Rheinberger (Hrsg.), *Räume des Wissens. Repräsentation, Codierung, Spur* (Berlin, 1997), S. 111 – 123.
- Krämer, Sybille, 'Textualität, Visualität und Episteme. über ihren Zusammenhang in der frühen Neuzeit', in: Renate Lachmann, Stefan Rieger (Hrsg.), *Text und Wissen, Technologische und anthropologische Aspekte* (Tübingen, 2003), S. 17 – 28.
- Manovich, Lev, *The Language of New Media* (Cambridge, Mass., 2001).
- Peirce, Charles Sanders, *Schriften zum Pragmatismus und Pragmatizismus* (Frankfurt a.M., 1991).
- Porombka, Stephan, *Hypertext. Zur Kritik eines digitalen Mythos* (München, 2001).
- Reck, H. U. und Trogemann, G., 'Exposé zum Antrag im Modellversuchsförderprogramm Kulturelle Bildung im Medienzeitalter' (URL: <http://www.khm.de/kmw/kit/pdf/expose.pdf>).
- Reck, Hans Ulrich, *Kunst als Medientheorie* (München, 2003).
- Reichle, Ingeborg, 'Kunst im Zeitalter der technischen Reproduzierbarkeit des Menschen. Zur Rezeption der Gentechnik in der zeitgenössischen Kunst', (URL: <http://www.kunsttexte.de/download/bwt/reichle.pdf>) – Zugriff am 2003.

- Rheinberger, H.-J., Wahrig-Schmidt, B. und Hagner, M., 'Räume des Wissens. Repräsentation, Codierung, Spur.' (1997), S. 7–23.
- Rheinberger, Hans-Jörg, *Experimentalsysteme und epistemische Dinge. Eine Geschichte der Proteinsynthese im Reagenzglas* (Göttingen, 2001).
- Serres, M. und Farouki, N., *Thesaurus der exakten Wissenschaften* (Frankfurt a. M., 2001).
- Simanowski, Roberto, *Interfictions. Vom Schreiben im Netz* (Frankfurt a. M.: Suhrkamp, 2002).
- Small, David, 'The Talmud Project: Rethinking the book', Dissertation, School of Architecture and Planning (MIT Media Lab, 1999), (URL: <http://acg.media.mit.edu/projects/theses.html>) – Zugriff am 2001.
- Wardrip-Fruin, N. und Montfort, N., *The New Media Reader* (Cambridge, Mass., 2003).