

**Lehrstuhl für Mensch-Maschine-Kommunikation**

**Aufmerksamkeitsbasierte  
Lokalisierung und Bewertung  
relevanter Information auf Papierdokumenten**

**Angela Engels  
(geb. Schreyer)**

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. J. Eberspächer

Prüfer der Dissertation:

1. Univ.-Prof. Dr. rer. nat. M. Lang
2. Univ.-Prof. Dr. rer. nat. A. Dengel,  
Universität Kaiserslautern

Die Dissertation wurde am 2.11.1999 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 8.5.2000 angenommen.



## **Danksagung**

Mein besonderer Dank gilt vor allem Herrn Professor Lang, der mein Vorhaben von Anfang an unterstützt und die Arbeit aufmerksam betreut hat. Ebenso danke ich Herrn Professor Dengel für sein stetiges Interesse und die Übernahme des Zweitgutachtens.

Danken möchte ich auch Herrn Kleinschmidt und Herrn Dr. Kämmerer, die die Arbeit im Fachzentrum für Mensch-Maschine-Kooperation (Zentralbereich Technik) der Siemens AG München erst ermöglicht haben.

Dort haben mich in allen Phasen der Arbeit Herr Maderlechner und Herr Suda auf die vielfältigste Art und Weise unterstützt. Gerade bei ihnen möchte ich mich ganz herzlich bedanken.

Besonders erwähnen möchte ich auch alle anderen Kollegen im Fachzentrum, denen ich herzlich für die ständige Bereitschaft zur fachlichen Diskussion und insbesondere für ihre Teilnahme an Befragung und interaktivem Test danke. Auch allen anderen Teilnehmern ein Dankeschön!

In diesem Zusammenhang sind außerdem Frau Professor Kuikka vom Department of Computer Science and Applied Mathematics der University of Kuopio und das finno-ugrische Institut der Ludwig-Maximilian-Universität München hervorzuheben, die bereitwillig finnische Testdokumente für Befragung und Test zur Verfügung gestellt haben.



## Abstract

This work describes a novel, attention-based view of documents in a sender-receiver-model (author-reader-model). According to this attention-based view, the author of a document marks up relevant information on the document using conspicuous layout features. These layout features attract a reader's attention at first glance and enable him to efficiently extract relevant information from the document.

This human mechanism is transferred into a new technical method, that localizes relevant information using exclusively the image of a scanned paper document, and that evaluates the relevance of one piece of information as compared to the others on the document.

In connection with this transfer, three topics are of major interest:

- As the reader's attention is drawn towards conspicuous, i.e. visually attractive document parts, a measure for the visual attractiveness of each object on the document's image is calculated. According to the basic idea, this measure of attractiveness is an implicit measure of relevance for the underlying content.
- The psychological texton theory explains the human visual attention by texture perception, so the technical method determines conspicuous layout features by texture analysis of the document. For use in the technical context, this work describes the texton theory's notion of texture and its implications in a formal mathematical way.
- When defining the measure of attractiveness, the subjective perception of attractiveness for different layout features is accounted for by using different weights. Initial values of these subjective weights are estimated from a survey amongst test persons. Test persons are also needed for the final test of the method, where the perceived attractiveness of all test documents' parts is compared to the calculated measure of attractiveness.

The final test indicates a large correlation between the calculated measure of attractiveness and the human perception of attractiveness. Additionally, a key-sentence extraction algorithm based on the measure of attractiveness produces quite plausible summaries of the analysed document.

# Inhaltsverzeichnis

|  |            |
|--|------------|
| <i>Danksagung</i> .....  | <i>i</i>   |
| <i>Abstract</i> .....  | <i>iii</i> |
| <i>Inhaltsverzeichnis</i> .....  | <i>iv</i>  |
| <i>Abbildungsverzeichnis</i> .....                                     | <i>vi</i>  |
| <i>Tabellenverzeichnis</i> .....                                       | <i>vii</i> |
| <b>1 Einleitung</b> .....  | <b>1</b>   |
| <b>2 Modellierung von Aufmerksamkeit und Attraktivität</b> .....       | <b>5</b>   |
| <b>2.1 Informationsübertragung beim Menschen: Stream-Modelle</b> ..... | <b>5</b>   |
| 2.1.1 Stream.....  | 5          |
| 2.1.2 Stream-Modell I.....   | 5          |
| 2.1.3 Stream-Modell II.....  | 6          |
| 2.1.4 Thematisch verwandte Arbeiten .....                              | 9          |
| <b>2.2 Decodierung</b> .....   | <b>10</b>  |
| 2.2.1 Stream-Modell und automatische Dokumentanalyse.....              | 10         |
| 2.2.2 Decodierung von Dokumenten beim Menschen.....                    | 11         |
| 2.2.3 Decodierung in der aufmerksamkeitsbasierten Dokumentanalyse..... | 13         |
| <b>3 Formalisierung der Begriffe</b> .....                             | <b>21</b>  |
| <b>3.1 Übersicht</b> .....   | <b>21</b>  |
| <b>3.2 Texton, Textur und Texturhierarchie</b> .....                   | <b>21</b>  |
| <b>3.3 Segmentierung und Resegmentierung</b> .....                     | <b>27</b>  |
| <b>3.4 Bewertung</b> .....   | <b>28</b>  |
| <b>3.5 Thematisch verwandte Arbeiten</b> .....                         | <b>30</b>  |
| <b>4 Texturbasierte Attraktivitätsanalyse</b> .....                    | <b>33</b>  |
| <b>4.1 Segmentierung</b> .....   | <b>33</b>  |
| 4.1.1 Segmentierung von Text .....                                     | 33         |
| 4.1.2 Text-Graphik-Trennung.....                                       | 36         |
| 4.1.3 Thematisch verwandte Arbeiten .....                              | 37         |
| <b>4.2 Resegmentierung</b> .....                                       | <b>38</b>  |
| 4.2.1 Übersicht.....   | 38         |
| 4.2.2 Definition von Einrückungen.....                                 | 40         |
| 4.2.3 Definition von Bullets.....                                      | 41         |
| 4.2.4 Definition von Fontstilen .....                                  | 42         |
| 4.2.5 Definition von Freistellung.....                                 | 46         |
| 4.2.6 Definition von Schriftgröße.....                                 | 48         |
| 4.2.7 Definition eines Flächenmaßes .....                              | 48         |
| 4.2.8 Definition von Schwärzungsgrad .....                             | 49         |
| <b>4.3 Attraktivitätsbewertung</b> .....                               | <b>49</b>  |
| 4.3.1 Übersicht.....   | 49         |
| 4.3.2 Normierung.....  | 49         |
| 4.3.3 Gewichte .....   | 50         |
| 4.3.4 Bewertungen der einzelnen Layoutmerkmale.....                    | 51         |
| 4.3.5 Attraktivitätsbewertung als additive Bewertung.....              | 52         |
| 4.3.6 Thematisch verwandte Arbeiten .....                              | 53         |

|            |   |            |
|------------|---|------------|
| <b>5</b>   | <b><i>Befragung zur Wahrnehmung von Layoutmerkmalen</i></b> .....                     | <b>55</b>  |
| <b>5.1</b> | <b>Ziel der Befragung</b> .....   | <b>55</b>  |
| <b>5.2</b> | <b>Durchführung und Teilnehmer</b> .....  | <b>55</b>  |
| <b>5.3</b> | <b>Aufbau des Fragebogens</b> .....   | <b>56</b>  |
| <b>5.4</b> | <b>Auswertung: Ergebnisse und Diskussion</b> .....                                    | <b>58</b>  |
| 5.4.1      | Kopfzeile .....   | 58         |
| 5.4.2      | Frage 1 .....   | 58         |
| 5.4.3      | Frage 2 .....   | 61         |
| <b>6</b>   | <b><i>Experimente und Tests</i></b> .....   | <b>65</b>  |
| <b>6.1</b> | <b>Experimente zur Bestimmung von Entscheidungsschwellen</b> .....                    | <b>65</b>  |
| <b>6.2</b> | <b>Erkennungsraten der Klassifikation bezüglich ausgewählter Layoutmerkmale</b> ..... | <b>67</b>  |
| <b>6.3</b> | <b>Interaktiver Test der Attraktivitätsbewertung</b> .....                            | <b>69</b>  |
| 6.3.1      | Ziel des Tests .....  | 69         |
| 6.3.2      | Durchführung und Teilnehmer.....  | 69         |
| 6.3.3      | Auswertung: Ergebnisse und Diskussion.....  | 73         |
| <b>7</b>   | <b><i>Extraktion „attraktiver“ Information</i></b> .....                              | <b>81</b>  |
| <b>8</b>   | <b><i>Zusammenfassung und Ausblick</i></b> .....                                      | <b>87</b>  |
|            | <b><i>Literatur</i></b> .....   | <b>93</b>  |
|            | <b><i>Verzeichnis der Formelzeichen</i></b> .....                                     | <b>99</b>  |
|            | <b><i>Verzeichnis wichtiger Begriffe</i></b> .....                                    | <b>103</b> |
|            | <b><i>A Fragebogen zur Wahrnehmung von Layoutmerkmalen</i></b> .....                  | <b>109</b> |
|            | <b><i>B Anleitung zum interaktiven Test</i></b> .....                                 | <b>115</b> |
|            | <b><i>C Testdokumente</i></b> .....   | <b>119</b> |
|            | <b><i>D Detaillierte Testergebnisse</i></b> .....                                     | <b>125</b> |
|            | <b><i>E Implementierung</i></b> .....   | <b>127</b> |

## Abbildungsverzeichnis

|   |    |
|---|----|
| Abbildung 2-1: Stream-Modell I – Informationsübertragung durch gesprochene Sprache .....    | 5  |
| Abbildung 2-2: Stream-Modell II – Informationsübertragung durch ein Dokument .....          | 6  |
| Abbildung 2-3: Beispiel zur Kennzeichnung relevanter Information durch Layoutmerkmale ..    | 8  |
| Abbildung 2-4: Beispiele von Julesz zur Untersuchung von Texturwahrnehmung .....            | 12 |
| Abbildung 2-5: Abstraktion vom Betrachter .....   | 13 |
| Abbildung 2-6: Bestimmung von Layoutmerkmalen und Attraktivitätsbewertung .....             | 14 |
| Abbildung 2-7: Ablauf der Decodierung .....   | 15 |
| Abbildung 3-1: Rahmen der Formalisierung .....  | 21 |
| Abbildung 3-2: Texturhierarchie versus Segmentierung und Resegmentierung .....              | 28 |
| Abbildung 4-1: Texturhierarchie für die Segmentierung eines Dokuments .....                 | 34 |
| Abbildung 4-2: Ergebnisse der einzelnen Segmentierungsschritte an einem Beispiel .....      | 35 |
| Abbildung 4-3: Typographische Bezeichnungen für die Definition von Layoutmerkmalen ....     | 40 |
| Abbildung 4-4: Bezeichnungen für die Definition von Einrückungen .....                      | 41 |
| Abbildung 4-5: Bezeichnungen für die Definition von Bullets .....                           | 42 |
| Abbildung 4-6: Zur Definition der Stroke-Schräglage $Slant(s)$ .....                        | 43 |
| Abbildung 4-7: Bezeichnungen für die Definition von Freistellung .....                      | 48 |
| Abbildung 4-8: Zur Herleitung der Gewichte $d_{Graphik}$ und $g_{BlackIntens}$ .....        | 51 |
| Abbildung 5-1: Aufbau des Fragebogens zur Wahrnehmung von Layoutmerkmalen .....             | 56 |
| Abbildung 5-2: Bezeichnungen der zu bewertenden Dokumentbereiche .....                      | 59 |
| Abbildung 5-3: Auswertung der Bewertung von Dokumentbereichen .....                         | 61 |
| Abbildung 5-4: Auswertung der Fragen zur Attraktivität einzelner Layoutmerkmale .....       | 62 |
| Abbildung 5-5: Auswertung der offenen Frage .....   | 62 |
| Abbildung 6-1: Beispiel eines synthetischen Dokuments in Schrift Graphite Light .....       | 65 |
| Abbildung 6-2: Präsentierte Attraktivität (...) der Testdokumente .....                     | 71 |
| Abbildung 6-3: Differenz der präsentierten Attraktivität und des Mittelwertes .....         | 74 |
| Abbildung 6-4: Mittlere Abweichung der korrigierten Attraktivitäten .....                   | 76 |
| Abbildung 7-1: Attraktivitätsbewertung (...) und (...) Extraktion von Schlüsselsätzen ..... | 83 |
| Abbildung 8-1: Flexibilität des Verfahrens zur Attraktivitätsbewertung .....                | 89 |
| Abbildung 8-2: Verwendbarkeit des Verfahrens .....  | 90 |



## Tabellenverzeichnis

|  |            |
|--|------------|
| <i>Tabelle 2-1: Berücksichtigte Layoutmerkmale .....</i>   | <i>17</i>  |
| <i>Tabelle 4-1: Grundlagen von Ähnlichkeits- und Nachbarschaftsrelationen.....</i>                               | <i>34</i>  |
| <i>Tabelle 4-2: Definition des Texton-Merkmals <math>Graphic(t)</math> bei der Text-Graphik-Trennung....</i>     | <i>36</i>  |
| <i>Tabelle 4-3: Definition von Layoutmerkmalen für Textons <math>t_h</math> über verfeinerte Relationen.....</i> | <i>39</i>  |
| <i>Tabelle 4-4: Gewichte zur Bewertung der einzelnen Layoutmerkmale .....</i>                                    | <i>51</i>  |
| <i>Tabelle 4-5: Bewertungen der einzelnen Layoutmerkmale.....</i>  | <i>51</i>  |
| <i>Tabelle 6-1: Mittlere Schräglage der ungefähr senkrechten Strokes .....</i>                                   | <i>66</i>  |
| <i>Tabelle 6-2: Größenverhältnis der Kleinbuchstaben (...) zu den Großbuchstaben .....</i>                       | <i>66</i>  |
| <i>Tabelle 6-3: Erkennungsraten bezüglich Einrückungen, Bullets, Fettdruck (...)</i>                             | <i>68</i>  |
| <i>Tabelle 6-4: Erkennungsraten bezüglich Fettdruck, Kursivdruck (...) aus [Sch98a].....</i>                     | <i>68</i>  |
| <i>Tabelle 6-5: Testbarkeit der Klassifikationsverfahren für Layoutmerkmale.....</i>                             | <i>70</i>  |
| <i>Tabelle 8-1: Testergebnisse bezüglich Einrückungen, Bullets, Fettdruck (...)</i>                              | <i>125</i> |
| <i>Tabelle 8-2: Testergebnisse bezüglich Fettdruck, Kursivdruck und Großschreibung (...)</i>                     | <i>126</i> |
| <i>Tabelle 8-3: Rechenzeit der texturbasierten Attraktivitätsanalyse .....</i>                                   | <i>127</i> |



# 1 Einleitung

---

„Dokument —  
mlat. ‘beweisende Urkunde’, eigentl. ‘das zur Belehrung  
über etwas oder zur Erhellung von etwas Dienliche’, zu lat.  
Docere ‘lehren’ (...)“

„Information —  
(...) Mitteilung, Nachricht, Auskunft (...)“

---

(Brockhaus Enzyklopädie, [Bro88])

„Das zur Belehrung über etwas oder zur Erhellung von etwas Dienliche“ – so definiert die Brockhaus Enzyklopädie ([Bro88]) ein *Dokument*. Demnach dient ein Dokument dazu, *Information* zwischen zwei räumlich und/oder zeitlich getrennten Personen zu übermitteln, wobei Information, ebenfalls nach Brockhaus, eine „Mitteilung, Nachricht“ oder „Auskunft“ ([Bro88]) ist.

Dieser Definition entsprechend ist jedes Medium, das Information zum Zweck der Übertragung speichern kann, ein Dokument. Es gibt also neben schriftlichen Dokumenten wie Briefen, Zeitschriften, E-Mails und WWW-Seiten, auch rein bildhafte Dokumente wie Fotos und Filme und Tondokumente wie Audio-CDs und Tonbänder. Schriftliche Dokumente transportieren dabei trotz Radio, TV und aufkommender Multi-Media-Technik immer noch einen großen Teil der Information, sei es in elektronischer Form oder als Papierdokument.

Da diese schriftlichen Dokumente auf Leser und deren Möglichkeiten der Informationsaufnahme zugeschnitten sind, ist die in einem Dokument enthaltene Information einem Computer oft nicht vollständig und problemlos zugänglich. Auf der anderen Seite werden immer mehr Prozesse der Informationsspeicherung, -übermittlung und -bearbeitung durch Computer unterstützt. Es wäre daher wünschenswert, die in schriftlichen Dokumenten enthaltene Information im Rahmen der Informationstechnik so aufzubereiten, daß sie möglichst effizient elektronisch weiterverarbeitet werden kann. Dieses Problem versucht die *automatische Dokumentanalyse* zu lösen. Ihr Ziel ist es also, sämtliche relevanten Informationen, die in einem Dokument enthalten sind, in eine symbolische Repräsentation zu transformieren, die sich für elektronische Informationsverarbeitung eignet.

In der automatischen Dokumentanalyse ist eine Sichtweise auf Dokumente üblich, wie sie im ODA-Standard ([App90]) festgehalten ist. ODA trennt die Struktur eines Dokuments von seinem Inhalt und unterteilt beide Kategorien in einen geometrischen und einen semantischen Anteil. Da der ODA-Standard zum elektronischen Austausch von Dokumenten entwickelt wurde, beschreibt er in erster Linie die (Daten-)Strukturen, denen ODA-Dokumente genügen. Wie Dokumente erzeugt und verarbeitet werden, ist dabei von geringem Interesse ([App90], S. 18).

Im Gegensatz dazu wird in der vorliegenden Arbeit eine *neue Sichtweise* entwickelt, in der gerade die Erzeugung eines Dokuments durch einen Autor und die Verarbeitung der darin enthaltenen Information durch einen Leser im Mittelpunkt stehen. Grundlage der Arbeit ist

eine psychologische Beschreibung des Zusammenwirkens von Autor und Leser bei der Informationsübertragung. Neben der „Mitteilung, Nachricht“ oder „Auskunft“ ([Bro88]) will der Autor dem Leser nämlich im allgemeinen auch die Bedeutung der Information vermitteln. Dabei geht es nicht um die aus dem ODA-Standard bekannten logischen Strukturen wie Adresse, Unterschrift, Titel oder Datum, sondern um die *Relevanz* der Information. Relevanz ist in diesem Zusammenhang eine Bezeichnung dafür, wie wichtig oder interessant eine Information ist. Sie ist damit eine Kategorie der Einordnung von Information, die einem Leser sehr viel unmittelbarer zur Verfügung steht als die obigen ODA-Kategorien Adresse, Unterschrift, Titel oder Datum.

Wie aber kann Relevanz bei Benutzung eines schriftlichen Dokuments übermittelt werden? Zur Beantwortung dieser Frage berücksichtigt man, daß schriftliche Dokumente Information auf zwei verschiedenen „Kanälen“ übertragen, d.h. sie benutzen *Text* und *Bilder* einerseits und die *Gestaltung* des Dokuments andererseits, um wichtige Inhalte zu vermitteln. Diese zwei Kanäle sind eng miteinander verknüpft, um dem Leser eine optimale Informationsaufnahme aus dem Dokument zu ermöglichen.

Die Arbeit vertritt nun die Auffassung, daß der Gestaltung dabei eine wesentliche Rolle zukommt. Es wird erläutert, daß ein Autor die Relevanz der einzelnen Dokumentbereiche definiert und dementsprechend ihre Gestaltung wählt. Diese Gestaltung lenkt die *Aufmerksamkeit* des Lesers auf diejenigen Teile des Dokuments, in der relevante Information zu finden ist, und sorgt damit dafür, daß der Leser die relevante Information effizient aus dem Dokument extrahiert.

Die Aufmerksamkeit eines Lesers wird dabei auf den ersten Blick auf auffällige gestalterische Merkmale, die sogenannten *Layoutmerkmale*, gelenkt. So springen zum Beispiel **fettgedruckte Wörter** in einem normalen Text geradezu ins Auge. Eine psychologische Theorie ([Jul83]) erklärt dieses Phänomen mit der menschlichen *Texturwahrnehmung*, durch die Änderungen im Blickfeld spontan auffallen. Die Texturwahrnehmung bewirkt also eine unbewußte, visuelle Aufmerksamkeit, die dem Leser eine effiziente Extraktion relevanter Information ermöglicht.

Es erscheint logisch, diese neue Sichtweise auch in der Dokumentanalyse zur Informationsextraktion aus Dokumenten auszunützen. Dadurch entsteht in dieser Arbeit ein völlig neues Verfahren, das ausschließlich auf der Bildinformation des eingescannten Dokuments arbeitet und auf dieser Grundlage den entscheidenden Schritt der Informationsextraktion durchführt. Dieser entscheidende Schritt ist es, die relevante, zu extrahierende Information zu finden, d.h. eine Lokalisierung der relevanten Information durchzuführen und die Relevanz der lokalisierten Information zu bewerten.

Dazu modelliert die vorliegende Arbeit die visuelle Aufmerksamkeit des Menschen, wobei höhere kognitive Prozesse und eine mögliche bewußte Steuerung der Aufmerksamkeit durch den Leser bei der Modellierung nicht berücksichtigt werden, obwohl höhere und niedrigere, bewußte und unbewußte kognitive Prozesse beim Menschen natürlich nie zu trennen sind. Da das komplexe Zusammenwirken der Wahrnehmung mit höheren und niedrigeren kognitiven Prozessen beim Menschen aber noch nicht ausreichend geklärt ist, konzentriert man sich bei der technischen Realisierung auf einen Ausschnitt der menschlichen Informationsverarbeitung. Dabei wurde für diese Arbeit die unbewußte, visuelle Aufmerksamkeit gewählt, die

psychologisch bereits umfangreich untersucht wurde, und die nach der entstandenen psychologischen Theorie den größten Teil der Lokalisierung relevanter Information bewerkstelligt.

Das technische Verfahren zur aufmerksamkeitsbasierten Lokalisierung und Bewertung relevanter Information wird aus der psychologischen Modellierung der Aufmerksamkeit abgeleitet, indem der Begriff der Aufmerksamkeit auf das Dokument abgebildet wird. Aus der Aufmerksamkeit wird so die *Attraktivität* von Dokumentbereichen, d.h. sogenannte attraktive Dokumentbereiche ziehen die Aufmerksamkeit an.

Das Verfahren berechnet ausgehend von einer Analyse der gestalterischen Merkmale ein quantitatives Maß der Attraktivität für jeden Dokumentbereich, das man *Attraktivitätsbewertung* nennt. Die Definition der Attraktivitätsbewertung beruht dabei auf zwei wesentlichen Schritten, die Erkenntnisse über die menschliche Wahrnehmung in den technisch-quantitativen Bereich abbilden:

1. Die psychologische Theorie ([Jul83]), die den Zusammenhang zwischen Texturwahrnehmung und visueller Aufmerksamkeit beschreibt, wird für eine technische und damit auch quantitative Behandlung der Thematik formalisiert.
2. Bei der Definition der Attraktivitätsbewertung wird das subjektive Empfinden von Attraktivität für verschiedene Layoutmerkmale durch die Festlegung von sogenannten Gewichten einbezogen. Die Initialwerte dieser Gewichte werden aus einer Befragung von Testpersonen geschätzt. Auch der endgültige Test des Verfahrens und der Definition der Attraktivitätsbewertung kann nur mit Hilfe von Testpersonen durchgeführt werden.

### **Gliederung der Arbeit:**

Im folgenden Kapitel 2 wird zunächst die neue, aufmerksamkeitsbasierte Sichtweise auf Dokumente ausführlich dargestellt. Danach erfolgt in Kapitel 3 die Übertragung dieser psychologischen Beobachtungen und Vorstellungen in den technischen Bereich durch eine Formalisierung der Begriffe, die bis zu einer formalen Darstellung der Attraktivitätsbewertung ausgedehnt wird. In Kapitel 4 wird dann aufbauend auf der Formalisierung der Begriffe und den Ergebnissen der Befragung die technische Realisierung der aufmerksamkeitsbasierten Lokalisierung und Bewertung relevanter Information dargestellt, wobei die Befragung selbst in Kapitel 5 beschrieben wird. Kapitel 6 faßt Experimente und Tests zur Klassifikation bezüglich der Layoutmerkmale zusammen. Insbesondere wird auch der interaktive Test beschrieben, bei dem mit Hilfe von Testpersonen die Attraktivitätsbewertung untersucht wurde. Kapitel 7 demonstriert dann, daß aufbauend auf der entwickelten Attraktivitätsbewertung tatsächlich eine Extraktion relevanter Information realisiert werden kann, indem eine Programmergänzung zur Extraktion von Schlüsselsätzen kurz beleuchtet wird. Eine Zusammenfassung der Arbeit und einen Ausblick auf mögliche Weiterentwicklungen und Anwendungsmöglichkeiten gibt abschließend noch Kapitel 8. In allen Kapiteln finden sich an gegebener Stelle Zusammenfassungen zu den thematisch verwandten Arbeiten.

Die Arbeit wird ergänzt durch eine Zusammenstellung der verwendeten Literatur, durch ein Verzeichnis der Formelzeichen und ein Verzeichnis der wichtigen Begriffe, die in dieser interdisziplinären Arbeit aus verschiedenen Bereichen stammen. Anhang A gibt anschließend den Fragebogen zur Layoutwahrnehmung und Anhang B die Anleitung zum interaktiven Test wieder. Danach sind in Anhang C die Testdokumente abgebildet, die insbesondere zum Test der Klassifikationsverfahren für Layoutmerkmale benutzt wurden. Anhang D listet im Detail

die Ergebnisse dieses Tests auf, und in Anhang E sind Einzelheiten der Implementierung angegeben

# 2 Modellierung von Aufmerksamkeit und Attraktivität

## 2.1 Informationsübertragung beim Menschen: Stream-Modelle

### 2.1.1 Stream

Grundlage für die hier präsentierte Sichtweise auf die automatische Lokalisierung und Bewertung von relevanter Information auf Dokumenten sind zwei Sender-Empfänger-Modelle der Informationsübertragung von Mensch zu Mensch. Sie sollen zunächst vom psychologischen Standpunkt aus erklären, wie Autor und Leser eines Dokuments bei der Informationsübertragung zusammenwirken, und wie die Begriffe Aufmerksamkeit und Attraktivität mit der Relevanz von Dokumentbereichen zusammenhängen.

Wenn Information von Mensch zu Mensch übermittelt werden soll, spielen die sprachlichen Aktivitäten Sprechen und Hören, bzw. Schreiben und Lesen eine wesentliche Rolle. Dabei ist es das Ziel des Informationssenders, seine Gedächtnisstrukturen einem Empfänger als Information mitzuteilen. Diese Gedächtnisstrukturen sind nicht zwangsläufig linear, wohl aber wird die benutzte Sprache durch eine in der Zeit lineare Wortkette übermittelt ([Lin81]). In der vorliegenden Arbeit wird diese lineare Kette als *Stream* bezeichnet. Nach ihm sind die zwei folgenden Sender-Empfänger-Modelle benannt, die die Informationsübertragung über einen idealen Kanal von Mensch zu Mensch beschreiben.

### 2.1.2 Stream-Modell I

Das Stream-Modell I beschreibt die Informationsübertragung von Mensch zu Mensch, wenn als Medium gesprochene Sprache benutzt wird.

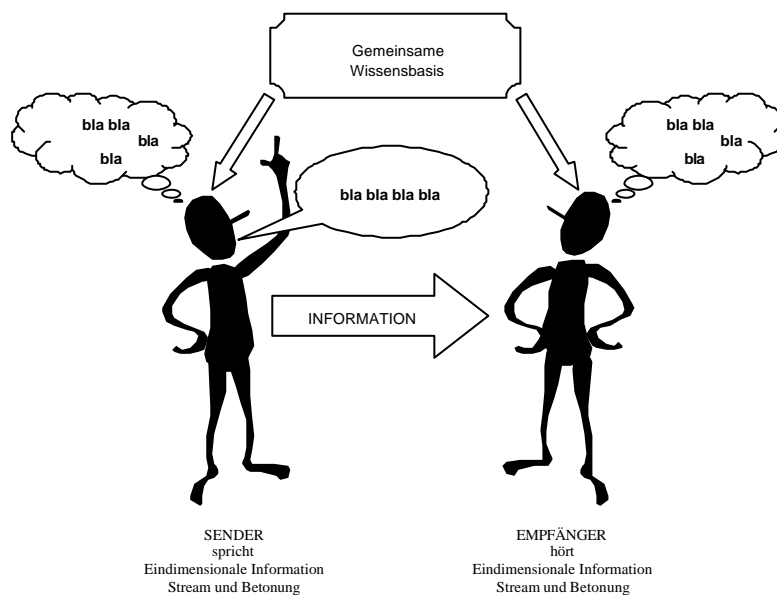


Abbildung 2-1: Stream-Modell I – Informationsübertragung durch gesprochene Sprache

Abbildung 2-1 zeigt, daß der Sender zunächst seine Gedankenstrukturen in einen linearen, d.h. zeitlich eindimensionalen, Stream umwandelt und per gesprochener Sprache zum Empfänger schickt. Der Empfänger nimmt den Stream über seinen auditiven Sinneskanal auf und rekonstruiert im Idealfall aus dem Stream genau die Gedankenstrukturen, die der Empfänger übermitteln wollte.

Dabei hilft dem Empfänger normalerweise die Tatsache, daß der Sender die Teile des Streams, die ihm wichtig erscheinen, besonders hervorhebt, z.B. durch Änderungen in Grundfrequenz oder Amplitude des Sprachsignals, durch Pausen und durch Änderungen in Sprechgeschwindigkeit und Deutlichkeit der Aussprache ([Che92]). Eine Hervorhebung durch Gestik und Mimik wird hier nicht berücksichtigt, da sie bei der Informationsübertragung nur in dem speziellen Fall eine Rolle spielt, in dem sich Sender und Empfänger während der Informationsübertragung sehen können. Der Sender versucht also im vorliegenden Modell durch rein sprachliche Betonung, die Aufmerksamkeit des Empfängers auf die relevanten Anteile im Stream zu lenken, so daß der Empfänger die ursprünglichen Gedanken des Senders möglichst korrekt und effizient aus dem Stream extrahieren kann.

An dieser Stelle wird die in dieser Arbeit verwendete Definition von *Relevanz* deutlich. Relevant sind nämlich diejenigen Anteile des Streams, die der Sender als besonders wichtig erachtet. Der Sender legt also die Relevanz eines Dokumentbereiches fest, der Empfänger dagegen nimmt sie nur wahr.

Diese Art der Informationsübertragung funktioniert selbstverständlich nur dann, wenn beide Beteiligten über eine gemeinsame Wissensbasis verfügen, und zwar nicht nur bezüglich der im Stream verwendeten Sprachanteile, sondern insbesondere auch bezüglich der benutzten Form der Betonung.

### 2.1.3 Stream-Modell II

Wird als Medium der Informationsübertragung nun nicht mehr die gesprochene Sprache, sondern ein (Papier-) Dokument benutzt, ergänzt man das Stream-Modell I um die Verarbeitungsschritte *Codierung* und *Decodierung* (Abbildung 2-2), die jeweils eine deutliche Änderung der Informationsdarstellung beschreiben.

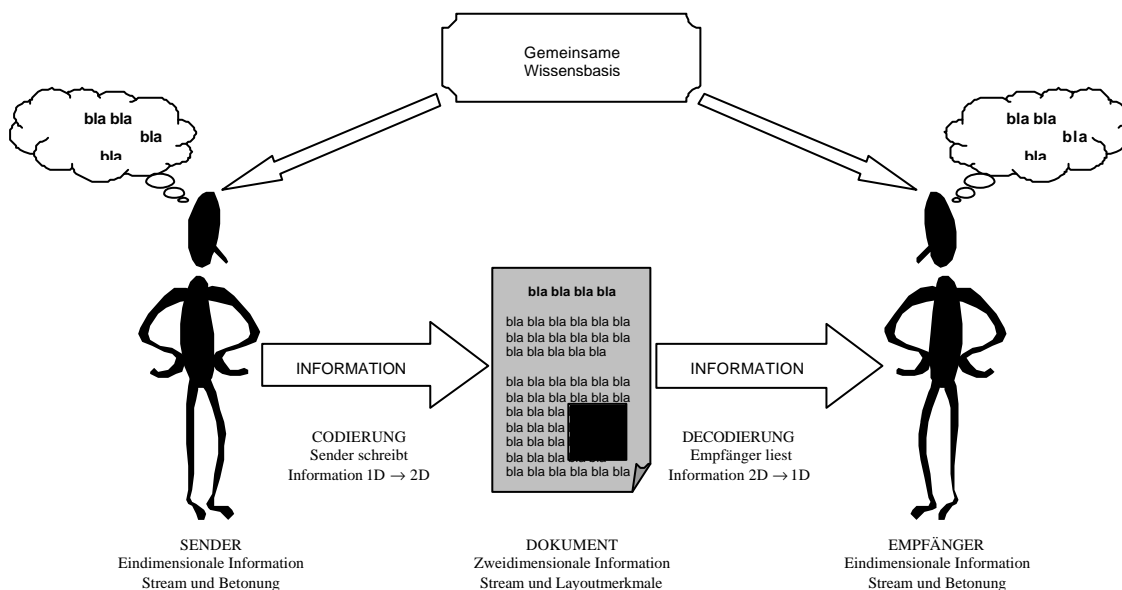


Abbildung 2-2: Stream-Modell II – Informationsübertragung durch ein Dokument



Diese simple Ergänzung des Modells ist insofern erlaubt, als man davon ausgehen kann, daß beim Schreiben die gleichen Grundprozesse ablaufen wie beim Sprechen, und beim Lesen die gleichen wie beim Hören ([And88]). Beide Modelle stellen Sprachverstehen als die genaue Umkehrung von Sprachproduktion dar. Dies ist im Rahmen der Modellbildung möglich, da nach kognitionspsychologischen Erkenntnissen die beiden Prozesse zwar nicht hundertprozentig komplementär sind, aber zumindest starke Ähnlichkeiten aufweisen ([And88]).

Auch beim Stream-Modell II stellt man sich vor, daß der Sender die Gedankenstrukturen, die er dem Empfänger übermitteln will, zunächst in einem eindimensionalen Stream und der dazugehörigen Betonung repräsentiert. Ein Codierungsprozeß ändert dann diese Repräsentation. Beim Schreiben nämlich „gießt“ der Sender seinen eindimensionalen Stream in die zweidimensionale Form des Dokuments.

Der Informationssender kann die Aufmerksamkeit des Empfängers nun nicht mehr durch die entsprechende Akzentuierung der gesprochenen Sprache steuern. Vielmehr markiert er jetzt durch visuelle Signale ([Nor89]), genauer durch Layout und Typographie ([Dut96]), die seiner Meinung nach relevanten Teile des Streams.

Layout bezeichnet dabei die Abmessungen und geometrische Anordnung von Dokumentbereichen, Typographie bezeichnet gestalterische Merkmale wie Schriftgröße und Fontstil. In dieser Arbeit sind Layout und Typographie unter dem Begriff *Layoutmerkmale* zusammengefaßt.

Während der Empfänger das Dokument liest, setzt er die Information bestehend aus Stream und Layoutmerkmalen in einen Stream und dessen Betonung um und rekonstruiert daraus die Gedankenstrukturen des Informationssenders. Auf der Seite des Informationsempfängers findet also eine Decodierung statt.

Dabei wird die Betonung zeitlich vor dem Stream decodiert. Der Stream nämlich wird erst durch das eigentliche Lesen zugänglich, während die Aufmerksamkeit des Empfängers durch bestimmte Layoutmerkmale *schon auf den ersten Blick* auf die aus Sendersicht relevanten Dokumentbereiche gelenkt wird. Aus diesen Bereichen wird folglich auch als erstes der Informationsstream extrahiert. Damit wird die Aufmerksamkeit des Empfängers durch Layoutmerkmale genau so gesteuert, daß relevante Informationen als erstes, d.h. effizient aus dem Dokument extrahiert werden.

Tatsächlich hat Klitz ([Kli92]) in einer Testreihe gezeigt, daß Textbereiche in Fett- und Kursivdruck auf den ersten Blick ins Auge fallen, und zwar ohne daß der Text gelesen werden kann. Für die typographischen Merkmale Fettdruck und Großschreibung (Druck in Großbuchstaben) hat eine Studie ([Fos77]) außerdem ergeben, daß die entsprechend markierten Textbereiche nach dem Lesen besser erinnert werden, die markierte Information also tatsächlich effizient aufgenommen wird.

Abbildung 2-3 verdeutlicht die Bedeutung der Layoutmerkmale noch einmal an einem Beispiel. Man sieht zum einen, daß unsere Aufmerksamkeit durch die Layoutmerkmale auch dann auf die relevanten Dokumentbereiche gelenkt wird, wenn der Text nicht lesbar ist, und zum anderen, wie sehr wir als Leser auf die Kennzeichnung durch Layoutmerkmale angewiesen sind, um die relevante Information zu lokalisieren.

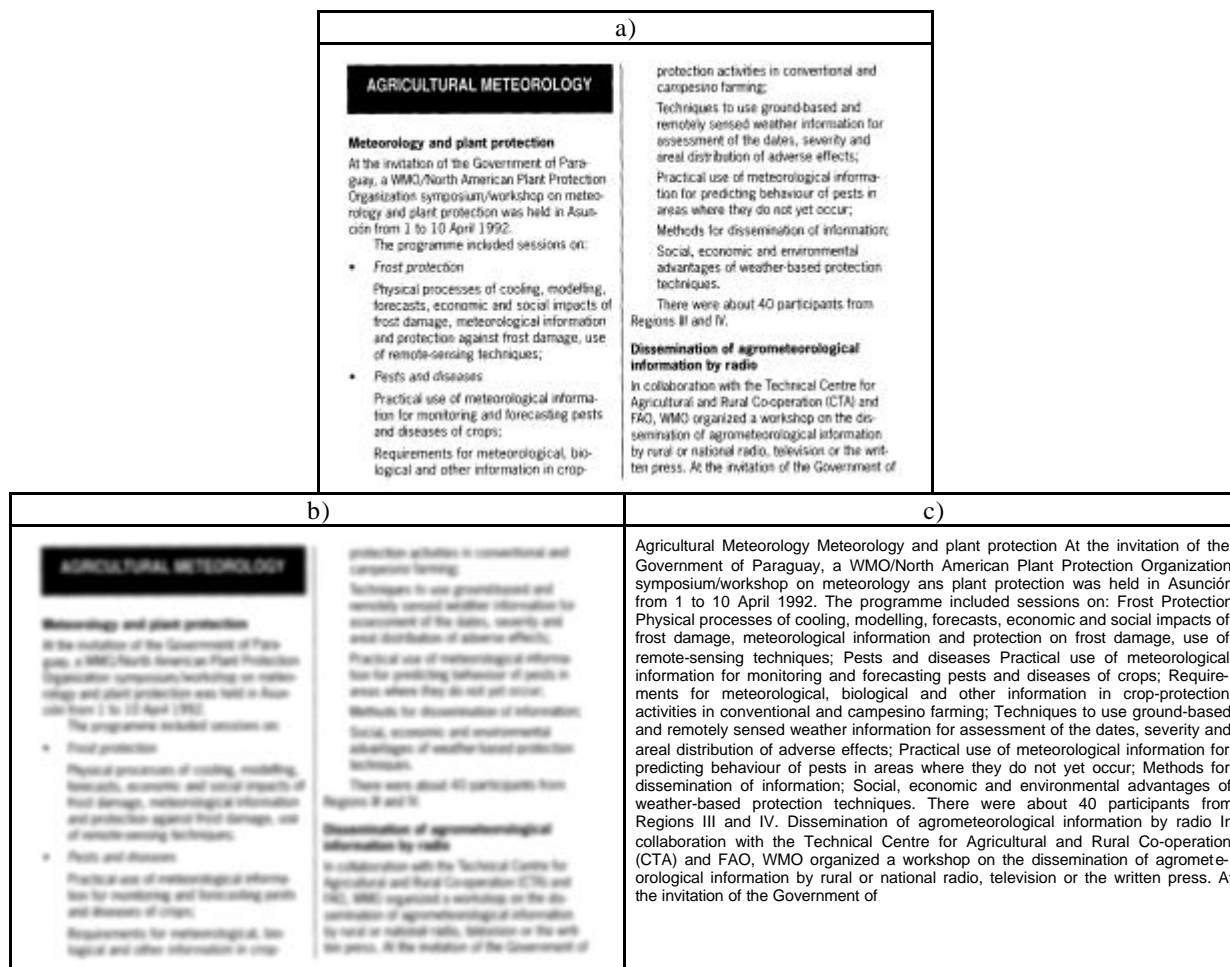


Abbildung 2-3: Beispiel zur Kennzeichnung relevanter Information durch Layoutmerkmale

- a) Stream und Betonung (Original, Ausschnitt aus Dokument S023bin.tif, [UoW96]),  
 b) nur Betonung (Text nicht mehr lesbar),  
 c) nur Stream (Text ohne Betonung durch Layoutmerkmale)

Die mit dem Stream-Modell II beschriebene Informationsübertragung per Dokument funktioniert natürlich nur dann, wenn auch hier Sender und Empfänger eine gemeinsame Wissensbasis haben. In ihr müssen insbesondere „Regeln“ über die Verwendung von Layoutmerkmalen enthalten sein. Tatsächlich wendet der Informationsempfänger natürlich keine Regeln an, sondern seine Aufmerksamkeit wird normalerweise durch seine spontane Wahrnehmung von Layoutmerkmalen gesteuert, ist also eine unbewußte Reaktion auf visuelle Eindrücke. Wie bereits Norrish ([Nor89]) bemerkt, kann man bei der heutigen weiten Verbreitung von computerbasierten Textverarbeitungssystemen auch nicht mehr davon ausgehen, daß auf Senderseite bewußt Regeln angewendet werden, d.h. daß nur ausgebildete Setzer und Designer Dokumente gestalten. Die gemeinsame Wissensbasis ist also eher Ausdruck ähnlicher Erfahrungen bei der Wahrnehmung und Gestaltung von Dokumenten.

Es soll hier noch einmal der *wesentliche Unterschied zur Informationsübertragung durch gesprochene Sprache* betont werden: durch den Übergang von einer eindimensionalen zu einer dauerhaften zweidimensionalen Repräsentation der Information verschwindet die strikte zeitliche Ordnung im Stream. Während dem Empfänger bei gesprochener Sprache zu jedem Zeitpunkt immer nur ein kurzer Ausschnitt des Streams und dessen Betonung präsent sind, liegt dem Empfänger bei Benutzung eines Dokuments die gesamte Information vor. In diesem

Fall ist eine noch effizientere Art der Informationsextraktion möglich, da der Empfänger auf die einzelnen Teile des Streams in beliebiger Reihenfolge zugreifen kann. Der Informationssender kann also durch visuelle Signale, die die Betonung des Streams codieren, die Aufmerksamkeit des Empfängers sprunghaft zwischen einzelnen Teilen des Streams verschieben und so alleine durch die Betonung übergreifendere Zusammenhänge bilden.

#### 2.1.4 Thematisch verwandte Arbeiten

1. Chen und Withgott ([Che92]) identifizieren mit Hilfe eines HMM-Erkenners<sup>1</sup> betonte Wörter in gesprochener Sprache. Zusammenhängende Abschnitte des Sprachsignals, in denen sich betonte Wörter häufen, werden lokalisiert und zu einer Zusammenfassung des gesprochenen Textes zusammengestellt. Diese Arbeit realisiert damit auf dem Gebiet der Sprachverarbeitung das, was hier im Bereich Dokumentverarbeitung demonstriert werden soll: wichtige Teile eines Informationsstreams werden unabhängig vom tatsächlichen Inhalt durch auffällige Merkmale betont; diese Tatsache kann zur Extraktion relevanter Information ausgenutzt werden.
2. In [Sou88], einer Arbeit aus dem Gebiet „Electronic Publishing“, entwickelt Southall einen ähnlichen Blickwinkel auf Dokumente wie die Stream-Modelle. Um Anforderungen an Systeme zum elektronischen Dokumentaustausch zu beschreiben, betrachtet er Dokumente als Medium zur Informationsübertragung und stellt dann die Frage, durch welche Mechanismen Dokumente die Information transportieren. Davon ausgehend definiert er den „Text“ eines Dokuments als die Gesamtheit der Information, die die Mitteilung eines Autors enthält. Ein Teil dieser Information ist der „verbale Inhalt“, der in etwa mit unserem Stream vergleichbar ist. Der verbale Inhalt wird ergänzt durch die „graphische Struktur“ als Ergebnis eines Codierungsprozesses auf Seiten des Autors und durch die „visuelle Struktur“ als die vom Leser wahrgenommene graphische Struktur. Diese Strukturen heißen im Stream-Modell Layoutmerkmale. Obwohl Southall auf die Rolle der Aufmerksamkeit nicht eingeht, erwähnt er als Teil des Codierungsprozesses die Betonung einzelner Wörter durch die graphische Struktur, um diese von ihrer Umgebung hervorzuheben. Southall beschränkt sich in seiner Darstellung auf reine Textdokumente. Im Gegensatz dazu kann der oben beschriebene Stream durchaus auch bildhafte Information beschreiben, die dann in Graphik oder Bildern codiert wird.
3. Im Bereich der automatischen Dokumentanalyse haben Doerman, Rivlin und Rosenfeld ([Doe98]) wie diese Arbeit die Idee entwickelt, daß Dokumente visuelle Übermittler einer Information von Autor zu Leser sind.

Bisher wurden allgemein zwei Strukturen als grundlegende Strukturen eines Dokuments angesehen und mit verschiedenen Verfahren analysiert ([App90], [Har94]): die geometrische Struktur, die dem Layout entspricht, und die semantische Struktur, die den logischen

---

<sup>1</sup> HMM: „Hidden-Markov-Modelle“ in der Sprachverarbeitung sind Netze, deren Knoten Zustände und deren Kanten Übergänge darstellen. Die Zustände werden von einem Anfangszustand über Zwischenzustände bis zu einem Endzustand durchlaufen. Zustände und Übergänge sind durch stochastische artikulatorische bzw. akustische Bestimmungsgrößen gekennzeichnet, d.h. jedem Zustand ist eine Wahrscheinlichkeit für das Emittieren eines bestimmten Symbols (z.B. eines Merkmalsvektors) zugeordnet, die Wahrscheinlichkeit für den Übergang eines Zustandes in einen anderen Zustand wird durch eine Übergangsmatrix festgelegt. Die Ausgabewerte eines HMM stellen sich als Überlagerung von Zustandsübergängen und Emissionen von Symbolen dar. Zu Erkennungszwecken werden in einer Trainingsphase die Wahrscheinlichkeiten festgelegt, danach erfolgt die Klassifikation durch Berechnung der Gesamtwahrscheinlichkeit, mit der eine zu klassifizierende Symbolfolge vom trainierten HMM ausgegeben werden kann.

Inhalt wiedergibt. Laut Doermann gibt es nun zwischen der geometrischen und der logischen Organisationsebene eines Dokuments noch eine sogenannte funktionelle Ebene („functional level“), die die Effizienz steuert, mit der das Dokument Informationen überträgt. Da Doermann dieser Ebene visuelle Markierungen wie Schriftgröße, Großschreibung und Zentriertheit zuordnet, die die Aufmerksamkeit des Lesers auf bestimmte wichtige Wörter oder Sätze lenken, vertritt er offensichtlich eine ähnliche Grundidee wie die vorliegende Arbeit. Das Ziel seiner Arbeit ist es allerdings nicht, die Relevanz von Objekten zu analysieren, sondern er versucht vor allem

- „use classification“ (seitenweise Klassifikation bezüglich der Verwendungsarten Lesen, Überfliegen und Durchsuchen) und
- „type classification“ (seitenweise Klassifikation bezüglich der Typen Titelseite, Dokumentkörper und Literaturverzeichnis).

Die Möglichkeit eines „functional enhancement“, das für den Benutzer relevante Anteile eines Dokuments auswählt, wird angedeutet.

4. Anders als die psychologisch motivierten Stream-Modelle ist das Sender-Kanal-Empfänger-Modell zur Dokumentanalyse von Chou und Kopec ([Cho95]) von der Informationstheorie inspiriert. Es beschreibt, wie eine Nachricht in Form einer logischen Dokumentstruktur in ein ideales zweidimensionales Bild codiert wird, über einen fehlerbehafteten Kanal übertragen und bei der Decodierung wieder in eine logische Struktur umgewandelt wird. Die Quelle der logischen Struktur und der Codierungsprozeß werden explizit mittels einer Grammatik und der Kanal über eine Wahrscheinlichkeitsfunktion modelliert. Die Decodiereinheit benutzt einen der dynamischen Programmierung verwandten Algorithmus, um die Fehlerwahrscheinlichkeit zwischen Original und rekonstruiertem Dokument zu minimieren.

## 2.2 Decodierung

### 2.2.1 Stream-Modell und automatische Dokumentanalyse

Aus der Sicht des Stream-Modells II ersetzt die automatische Dokumentanalyse den menschlichen Informationsempfänger so weit wie möglich durch eine Maschine. Wenn die Maschine Informationen eines Dokuments in eine symbolische Repräsentation transformiert, vollzieht sie also die Decodierung nach.

Das Ziel der vorliegenden Arbeit ist es, die beim Menschen ablaufende Decodierung für die Dokumentanalyse so nachzuvollziehen, daß vor allen Dingen die Betonung der relevanten Information rekonstruiert wird. Dies ermöglicht eine neuartige, effektive Art und Weise, diese relevante Information im Dokument zu lokalisieren und ihre Relevanz zu bewerten.

Das Stream-Modell II hat gezeigt, daß der Sender Informationen, die er für wichtig hält, während der Codierung durch Layoutmerkmale besonders hervorhebt, und daß diese Layoutmerkmale die Aufmerksamkeit des Informationsempfängers bei der Decodierung auf den ersten Blick auf die relevanten Dokumentbereiche lenken. Eine technisch realisierte *aufmerksamkeitsbasierte* Decodierung wird folglich die Layoutmerkmale auf einem Dokument finden und ein Maß dafür angeben, wie stark die Layoutmerkmale eines bestimmten Dokumentbereichs die Aufmerksamkeit eines Betrachters auf den ersten Blick anziehen würden.

Wie oben erwähnt, kann man eine solche Decodierung nicht einfach durch eine Umkehrung der Codierungsregeln bewerkstelligen, weil man nicht von einem auf Senderseite bewußt be-

nutzen Regelwerk ausgehen kann. Deshalb betrachtet diese Arbeit nur die Empfängerseite und versucht hier zuerst zu verstehen, wie die Aufmerksamkeit eines Betrachters modelliert werden kann. Für die Dokumentanalyse wird daraus dann das Konzept der Attraktivität von Dokumentbereichen abgeleitet.

### 2.2.2 Decodierung von Dokumenten beim Menschen

Als Grundlage für eine psychologisch inspirierte Decodierung in der Dokumentanalyse soll im folgenden erläutert werden, wie man sich die Decodierung von Dokumenten, insbesondere die Lokalisierung von Layoutmerkmalen, beim Menschen vorstellen kann.

Anhand experimenteller, psychologischer Studien wurde von Julesz ([Jul83], [Jul91]) die sogenannte *Texton Theorie* entwickelt. Sie ist eines von mehreren bekannten Modellen der menschlichen visuellen Wahrnehmung und erscheint besonders geeignet, die Wahrnehmung von Dokumenten zu beschreiben.

Nach einer Grundidee dieser Theorie arbeitet das visuelle System des Menschen in zwei unterschiedlichen Modi:

- „*Preattentive vision*“: ein Prozeß, der augenblicklich, ohne genaues Hinschauen, ein großes Gesichtsfeld analysiert, und zwar in Parallelverarbeitung unabhängig von der Anzahl der Muster. Er beruht auf Texturwahrnehmung.
- „*Attentive vision*“: ein Prozeß, der beschränkt auf eine kleine Blende eine serielle Suche realisiert, wie bei der Formerkennung. Dieser Prozeß wird von der „preattentive vision“ auf Orte gelenkt, an denen sich die Textur ändert.

Julesz geht davon aus, daß sich ein Großteil des Sehens bereits auf der Ebene der „preattentive vision“ abspielt. Diese untere Ebene dient als Frühwarnsystem. Sie kennzeichnet interessante Punkte, die man genauer anschauen sollte, realisiert also eine *unbewußte, visuelle Aufmerksamkeit*.

Diese visuelle Aufmerksamkeit erleben wir immer dann, wenn wir Änderungen in der Beschaffenheit eines Dokuments, d.h. Änderungen in der Textur eines Dokuments, auf den ersten Blick erkennen. Ein Punkt in einem Dokument, der durch ein Layoutmerkmal gekennzeichnet ist, ist eine solche Änderung in der Beschaffenheit und damit nach Julesz visuell interessant. Wie das Stream-Modell II zeigt, fallen aber im speziellen Fall von Dokumenten visuell interessante Punkte mit inhaltlich relevanten Dokumentbereichen zusammen, weil der Autor diesen Zusammenhang bei der Codierung seiner Gedanken verwendet. Leser können relevante Information so effizient aus einem Dokument extrahieren.

Die „preattentive vision“ beruht nach Julesz auf Texturwahrnehmung. In der von Julesz aufgestellten Texton Theorie setzten sich Texturen aus diskreten Wahrnehmungseinheiten, den *Textons*, zusammen. Auch in dieser Hinsicht erscheint es logisch, Text als Textur zu betrachten, da sich Text auf einem Dokument offensichtlich auch aus diskret wahrnehmbaren Einheiten zusammensetzt, nämlich aus Strichen, Buchstaben, Wörtern, Textzeilen und Blöcken. In die Vorstellung einer Block-Textur lassen sich sogar Bildblöcke integrieren.

Textons zeichnen sich dadurch aus, daß sie einfache, direkt beobachtbare Merkmale haben, z.B. Farbe, Orientierung, Breite und Höhe. Die visuelle Aufmerksamkeit wird an die Orte im Gesichtsfeld gelenkt, an denen sich die Merkmale der Textons ändern, oder an denen sich ihre

Dichte ändert. Dies sind Orte der Texturänderung, die dazwischen liegenden *Texturen* sind also Anhäufungen von ähnlichen Wahrnehmungseinheiten, die räumlich zusammenhängen.

In Kapitel 3 wird für den Gebrauch in der Dokumentanalyse von dieser Vorstellung ausgehend eine Formalisierung der Begriffe Textur und Texton vorgenommen. Es wird der zusätzliche Begriff der Texturhierarchie eingeführt, da Dokumente komplexere visuelle Strukturen sind als die von Julesz untersuchten einfach strukturierten Beispiele (siehe Abbildung 2-4).



Abbildung 2-4: Beispiele von Julesz zur Untersuchung von Texturwahrnehmung (Fig. 2 aus [Jul83]), Spontane Wahrnehmung von Texturunterschieden (visuelle Aufmerksamkeit), links bedingt durch Textons unterschiedlicher Größe, rechts bedingt durch Textons unterschiedlicher Orientierung

Die Decodierung der kompletten Information eines Dokuments ist mit dem Lokalisieren relevanter Dokumentbereiche über Texturwahrnehmung selbstverständlich noch nicht abgeschlossen. Vielmehr laufen danach beim Lesen und Interpretieren der relevanten Teile höhere kognitive Prozesse ab, durch die der Mensch den Inhalt des Dokuments erst erschließt. Eine Trennung zwischen der reinen Wahrnehmung und den höheren kognitiven Prozessen ist einem Leser im allgemeinen nicht möglich, kann unter künstlichen Bedingungen wie in der Befragung zur Wahrnehmung von Layoutmerkmalen (Kapitel 5) aber versucht werden.

### Thematisch verwandte Arbeiten

1. Watt ([Wat93]) betrachtet, wie diese Arbeit auch, Text als Textur. Er untersucht unter diesem Gesichtspunkt die visuelle Auswirkung von verschiedenen Wort- und Zeilenabständen.
2. Die „feature integration theory“ von Treisman und Gelade ([Tre80]) ist der Texton Theorie verwandt. Auch Treisman geht davon aus, daß das visuelle System in einem parallelen, automatischen und großflächigen Prozeß zuerst einzelne Merkmale wie Farbe, Orientierung, räumliche Frequenz, Helligkeit und Bewegungsrichtung registriert. Objekte, die nur durch eine dieser Basisdimensionen charakterisiert sind, fallen auf den ersten Blick ins Auge. Die sogenannte „focussed attention“ ist erst dann nötig, wenn mehrere Basisdimensionen miteinander in Beziehung gesetzt werden müssen, um so Objekte zu bilden.
3. Die bisher besprochenen Arbeiten bewegen sich im Rahmen der Psychologie. Ein Zusammenhang der menschlichen Texturwahrnehmung mit den in der Neurophysiologie entdeckten rezeptiven Feldern ist allerdings ebenfalls allgemein anerkannt ([And88], [Bac97], [Jul83]). Rezeptive Felder sind spezifische neuronale Filter, die für die Registrierung der visuellen Basisdimensionen Helligkeit, Orientierung, Ortsfrequenz, Bewegung, Stereo und Farbe zuständig sind. Laut Bach ([Bac97]) überlagern sich bei der Textursegmentierung die neuronalen Signale additiv, ohne sich gegenseitig zu behindern. Dies wird

bei der Definition einer Attraktivitätsbewertung als additive Bewertung noch eine Rolle spielen (Kapitel 3.4 und 4.3).

4. Koch und Ullman ([Koc85]) simulieren mit Hilfe eines künstlichen neuronalen Netzes die Verschiebung der visuellen Aufmerksamkeit auf einer Szene. Ihrer Vorstellung nach wird eine Reihe von elementaren Merkmalen (z.B. Farbe, Orientierung, Bewegungsrichtung) parallel in verschiedenen topographischen neuronalen Karten repräsentiert. Danach werden diese Karten entsprechend von Auswahlkriterien selektiv auf eine nicht-topographische Karte abgebildet. Dieser zweite Prozeß ist die Grundlage für die visuelle Aufmerksamkeit.

Itti, Koch und Niebur zeigen in [Itt98] dann die Weiterentwicklung des Verfahrens und die Anwendung auf natürliche Szenen. Als elementare Merkmale werden in diesem Fall Kontrast, Farbe und Orientierung durch lineare Filterung des Bildes bestimmt.

5. Das für diese Arbeit übernommene Modell von Julesz ([Jul83], [Jul91]) und Treisman ([Tre80]), aber auch von Koch, Ullmann ([Koc85]) und Itti et al. ([Itt98]) beschreibt Aufmerksamkeit als unbewußten Prozeß, der nur von der Beschaffenheit der visuellen Szene auf relevante Teile gelenkt wird (visuelle Aufmerksamkeit). Wie bei Anderson ([And88]) und Lindsay und Norman ([Lin81]) dargestellt, kann aber jeder seine Aufmerksamkeit auch bewußt lenken. In der vorliegenden Arbeit wird dieser zweite Aspekt der Aufmerksamkeit nicht berücksichtigt. Er stellt eine höhere kognitive Leistung dar, kommt also für eine Modellierung der Wahrnehmung „auf den ersten Blick“ nicht in Frage.

### 2.2.3 Decodierung in der aufmerksamkeitsbasierten Dokumentanalyse

Es wurde bisher dargestellt, wie die Aufmerksamkeit eines Betrachters von visuell interessanten Strukturen auf dem Dokument zu inhaltlich relevanter Information gelenkt wird. Nur dieser beim Menschen wesentliche Teil der Decodierung, nicht die ganze Decodierung, soll im folgenden technisch nachvollzogen werden. Um solchermaßen ein automatisches Verfahren zur Informationsextraktion zu entwickeln, wird im nächsten Schritt vom Betrachter abstrahiert.

Das Konzept der *Aufmerksamkeit* wird dazu übertragen in ein Konzept der *Attraktivität von Dokumentbereichen*. Unter diesem neuen Blickwinkel wird nicht mehr die Aufmerksamkeit eines Betrachters von einem Dokumentbereich angezogen, sondern ein Dokumentbereich hat für einen Betrachter eine gewisse Attraktivität (Abbildung 2-5). Dabei geht die Modellierung über vom dynamischen Prozeß der Aufmerksamkeit - man schaut einzelne Dokumentbereiche in einer gewissen Reihenfolge an - zur statischen Attraktivitätsverteilung auf einem Dokument.

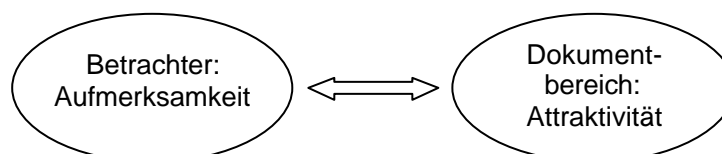


Abbildung 2-5: Abstraktion vom Betrachter:  
Konzept der Aufmerksamkeit wird übertragen in  
Konzept der Attraktivität von Dokumentbereichen

Diese *Attraktivitätsverteilung* ergibt sich, indem jedem einzelnen Dokumentbereich ein Maß für seine Attraktivität zugeordnet wird. Diese sogenannte *Attraktivitätsbewertung* läßt einen Rückschluß darauf zu, wie stark der Dokumentbereich die Aufmerksamkeit eines Betrachters auf den ersten Blick anziehen würde.

Die Aufmerksamkeit wird entsprechend dem Stream-Modell II (Kapitel 2.1.3) durch Layoutmerkmale gesteuert. Das heißt übertragen, daß auch die Attraktivität eines Dokumentbereiches durch Layoutmerkmale gegeben ist.

Eine automatische Decodierung, die sich am Ablauf der menschlichen Decodierung orientiert, wird also als erstes die Layoutmerkmale der einzelnen Dokumentbereiche bestimmen müssen. Abbildung 2-6 illustriert, daß die Bestimmung von Layoutmerkmalen mit ganz unterschiedlichen technischen Methoden vorgenommen werden muß, je nachdem, wie das Dokument repräsentiert ist. Ist das Dokument ein eingescanntes Papierdokument, bieten sich Methoden der Bildverarbeitung an, ist es dagegen ein elektronisches Dokument, z.B. ein HTML-Dokument, sind die Layoutmerkmale in der jeweiligen Beschreibungssprache codiert und können direkt abgelesen werden. Die anschließende Berechnung einer Attraktivitätsbewertung aus den Layoutmerkmalen ist dagegen von der ursprünglichen Repräsentation des Dokuments unabhängig.

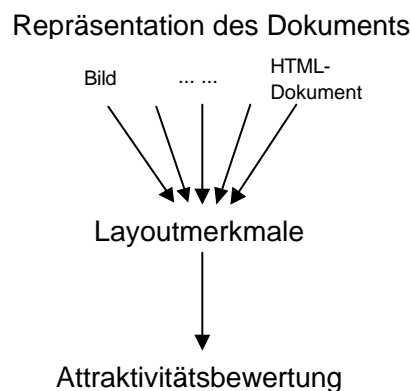


Abbildung 2-6: Bestimmung von Layoutmerkmalen und Attraktivitätsbewertung aus unterschiedlichen Repräsentationen des Dokuments

Diese Arbeit betrachtet Papierdokumente, und zwar aus zweierlei Gründen. Zum einen sind die bisher existierenden Methoden zur Lokalisierung relevanter Information auf Papierdokumenten noch nicht ausgereift ([Doe97]). Zum anderen ist es das Ziel der Arbeit, sich zur möglichst optimalen Implementierung einer Attraktivitätsbewertung möglichst nah an der menschlichen Wahrnehmung zu orientieren, der schriftliche Dokumente üblicherweise als Bild und nicht in elektronisch codierter Form vorliegen.

Man bestimmt die Layoutmerkmale also ausgehend vom Bild des eingescannten Papierdokuments mittels Bildverarbeitung. Die dazu benutzten Methoden werden in Anlehnung an die menschliche Wahrnehmung gewählt.

Kapitel 2.2.2 hat erläutert, daß ein Betrachter ein Dokument auf den ersten Blick mit seinem Mechanismus zur Texturwahrnehmung analysiert. Man geht also davon aus, daß ein Betrachter einerseits die Gruppierung von Dokumentbereichen, andererseits auch die Lokalisierung von Layoutmerkmalen durch Texturwahrnehmung bewerkstelligt. Folglich untersucht



auch ein technisches Verfahren ein Dokument durch Texturanalyse, um eine Segmentierung des Dokuments durchzuführen und die Layoutmerkmale zu lokalisieren. Dabei ist der verwendete Texturbegriff dem Juleszschen Begriff verwandt.

Die bisher gewonnenen psychologischen Erkenntnisse über die Informationsverarbeitung auf Seiten des Lesers werden also umgesetzt in eine automatische Decodierung von Dokumenten. Abbildung 2-7 zeigt den Ablauf der gesamten automatischen Decodierung. Nach der *Datenerfassung und Vorverarbeitung*, erfolgen die texturbasierten Schritte *Segmentierung* und *Resegmentierung bezüglich der Layoutmerkmale*. Diese Schritte entsprechen im Modell der menschlichen Decodierung eindeutig der „preattentive vision“. Auch die Zusammenfassung zu einer *Attraktivitätsbewertung* wird im Vergleich noch eher dieser niedrigen kognitiven Ebene zugeordnet, während die *Extraktion* von Inhalt bereits einem höheren kognitiven Prozeß entspricht.

Obwohl die hier beschriebene Decodierung von der Decodierung beim Menschen inspiriert ist, kann der Ablauf selbstverständlich nicht direkt verglichen werden. Die Decodierung beim Menschen ist kein sequentieller Vorgang, sondern höhere und niedrigere kognitive Prozesse überschneiden sich teilweise und sind folglich nur schwer voneinander zu trennen.

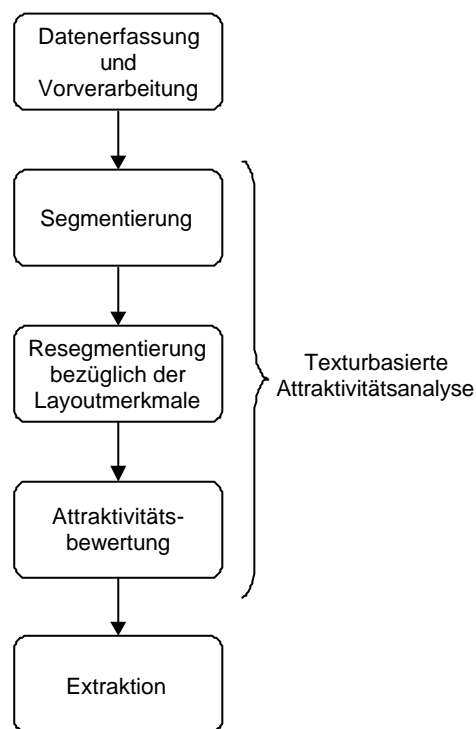


Abbildung 2-7: Ablauf der Decodierung in der aufmerksamkeits-, bzw. attraktivitätsbasierten Dokumentanalyse

### Datenerfassung und Vorverarbeitung

Grundlage für die automatische Dokumentanalyse ist eine Erfassung und elektronische Repräsentation der zu bearbeitenden Papierdokumente. Die Papierdokumente werden mit einem optischen Scanner erfaßt, wodurch das Dokument in ein Raster von Bildpunkten („pixels“) überführt wird. Je nachdem, ob ein Bildpunkt die Werte 0 oder 1, 0 bis 255 oder in drei Kanälen 0 bis 255 annehmen kann, spricht man von Binär-, Grauwert- oder Farbbildern. In der

vorliegenden Arbeit werden Binärbilder benutzt. Sie werden in der Arbeit als Bilder der Dokumente, kurz als *Dokumente*, bezeichnet.

Die Dokumente werden nach der Erfassung einer Vorverarbeitung unterzogen, die vor allen Dingen deshalb nötig ist, weil die Dokumente nicht wie im Modell (Abbildung 2-2) über einen idealen Kanal von Sender zu Empfänger übertragen werden. Vielmehr entstehen durch etwaiges Kopieren oder Faxen und durch das Einscannen Artefakte auf dem Dokument, d.h. entweder entstehen zusätzliche Objekte, die das Original nicht enthält, oder Objekte, die das Original enthält, verschwinden. Diese Verfälschung der ursprünglichen Information stört die spätere Analyse und muß deshalb soweit wie möglich korrigiert werden. Auch die Korrektur einer beim Scannen entstandenen Schräglage des Dokuments gehört zur Vorverarbeitung, da eine Schräglagenkorrektur nachfolgende Algorithmen normalerweise deutlich vereinfacht. Für diese Arbeit ist die Schräglagenkorrektur mit der nachfolgend beschriebenen Segmentierung verknüpft.

### **Segmentierung**

Die hier realisierte Segmentierung findet auf dem Dokument Texturen, d.h. sie gruppiert Strukturen ähnlicher, benachbarter Objekte zu größeren Einheiten. Diese Einheiten können dann in einer Texturhierarchie wieder Objekte einer erneuten Segmentierung werden. Auf diese Art und Weise entstehen aus den Strokes, d.h. aus grundlegenden strichartigen Objekten, die Objekte Buchstabe, Wort, Zeile und Block (Kapitel 4.1).

Die Segmentierung ist außerdem für die sogenannte Text-Graphik-Trennung zuständig, d.h. die Segmentierung unterscheidet zwischen Objekten, die nur Text enthalten, und solchen, die keinen Text, sondern bildhafte Darstellungen, enthalten. *Graphikobjekte* müssen später anders behandelt werden als *Textobjekte*.

### **Resegmentierung bezüglich der Layoutmerkmale**

Von den genannten Objekten werden im Codierungsprozeß im allgemeinen Wörter und Text- oder Graphikblöcke zur Betonung markiert. Deshalb untersucht die Decodierung diese Objekte auf Layoutmerkmale. Tabelle 2-1 gibt die in der Arbeit berücksichtigten Layoutmerkmale an. Ihre Auswahl ist durch eine Befragung zur Wahrnehmung von Layoutmerkmalen (Kapitel 5) bestätigt worden.

Man beachte, daß einige dieser Layoutmerkmale (Einrückung, Bullets, Fontstile) ein Objekt durch ihr Vorhandensein attraktiv machen, während bei anderen (Freistellung, Schriftgröße, Flächenmaß, Schwärzungsgrad) die Stärke ihrer Ausprägung für ihre Attraktivität verantwortlich ist.

Betrachtet man nun die mit Layoutmerkmalen markierten Wörter, Text- und Graphikblöcke als Textons im Sinne der vorne beschriebenen Texton Theorie, dann sind die Layoutmerkmale Merkmale dieser Textons, die die umliegenden Textons nicht haben. Die solchermaßen markierten Wörter und Blöcke bilden also jeweils eigene Texturen, die die visuelle Aufmerksamkeit anziehen.

In der automatischen Decodierung findet man nun diese Texturen durch eine Resegmentierung der bereits bestehenden Segmentierung, wobei die in Tabelle 2-1 vorgegebenen Layoutmerkmale als Merkmale der Textons benutzt werden. Dies wird in Kapitel 4.2 im Detail erläutert werden.

| Text    | Wort                    |   |
|---------|-------------------------|---|
|         | Einrückung              | Weißer Einzug in der ersten Zeile eines Absatzes. Das Wort nach der Einrückung ist besonders attraktiv.                         |
|         | Bullet                  | Kleines Symbol, um die einzelnen Punkte einer Aufzählung zu markieren. Das Wort direkt nach dem Bullet ist besonders attraktiv. |
|         | Fontstil Fettdruck      | z.B. <b>xxxxxxx</b>   |
|         | Fontstil Kursivdruck    | z.B. <i>xxxxxxx</i>   |
|         | Fontstil Großschreibung | z.B. <b>XXXXXXXX</b>  |
|         | Textblock               |   |
|         | Freistellung            | Die Größe des Weißbereichs, der den Block umgibt.   |
|         | Schriftgröße            | Die mittlere Zeilenhöhe im Block.   |
| Graphik | Graphikblock            |   |
|         | Freistellung            | Die Größe des Weißbereichs, der den Block umgibt.   |
|         | Flächenmaß              | Ein Maß für die Ausdehnung des Graphikblocks.   |
|         | Schwärzungsgrad         | Ein Maß für die visuelle Dichte des Blocks, das zur groben Unterscheidung von Rasterbildern und Liniengraphiken benutzt wird.   |

Tabelle 2-1: Berücksichtigte Layoutmerkmale

Die Resegmentierung liefert außerdem für jedes Layoutmerkmal einen Zahlenwert, der die Klassifikation des Textons bezüglich des Layoutmerkmals repräsentiert. Dieser Zahlenwert kennzeichnet für Einrückung, Bullets und die Fontstile das Vorhandensein des jeweiligen Merkmals mit dem Wert 1, für alle anderen Layoutmerkmale die Stärke der Ausprägung.

### Attraktivitätsbewertung

Anschließend müssen für jedes Objekt die Ergebnisse der Resegmentierung so zusammengefaßt werden, daß sich ein Maß für die Attraktivität des Objekts ergibt. Neuere elektrophysiologische Untersuchungen ([Bac97]) deuten an, daß sich bei der Texturwahrnehmung die verschiedenen visuellen Basisdimensionen additiv überlagern, ohne sich gegenseitig zu behindern. Da die einzelnen Layoutmerkmale durch verschiedene Basisdimensionen gekennzeichnet sind, liegt es nahe, die einzelnen Zahlenwerte additiv zu überlagern.

Aus unserer Erfahrung bei der Wahrnehmung von Layoutmerkmalen wissen wir aber, daß nicht alle Layoutmerkmale gleich attraktiv sind. Eine gewichtete additive Überlagerung der Werte erscheint also angebracht. Da die Attraktivität der einzelnen Layoutmerkmale für jeden Betrachter subjektiv ist, wurden die Gewichte durch eine Befragung ermittelt (siehe Kapitel 5).

Es entsteht so eine Attraktivitätsbewertung aller Objekte eines Dokuments. Eine formale Beschreibung der Attraktivitätsbewertung erfolgt in Kapitel 4.3.

### Extraktion

Letzter Schritt der Decodierung ist die Extraktion des Inhalts. Dabei wird die im Dokument enthaltene relevante Information in eine symbolische Repräsentation transformiert, die sich für eine elektronische Informationsverarbeitung eignet.

Bei der hier präsentierten aufmerksamkeitsbasierten Extraktion werden dazu die Objekte entsprechend ihrer Attraktivitätsbewertung sortiert, um relevante Informationen zu lokalisieren.

Eine Extraktion des Inhalts kann dann mit Hilfe einer OCR<sup>2</sup> durchgeführt werden, und zwar entweder so, daß der OCR nur besonders attraktive Schlüsselwörter zugeführt werden, oder auch so, daß ganze Schlüsselsätze extrahiert werden. Die Extraktion von Schlüsselsätzen wird in Kapitel 7 demonstriert. Dort werden auch die in der Literatur genannten bisherigen Ansätze zur Informationsextraktion besprochen.

Als entscheidende Verarbeitungsschritte einer aufmerksamkeitsbasierten Informationsextraktion stehen in dieser Arbeit die Segmentierung, und besonders die Resegmentierung und die Attraktivitätsbewertung im Mittelpunkt. Sie werden im folgenden unter der Bezeichnung *texturbasierte Attraktivitätsanalyse* zusammengefaßt und näher beleuchtet.

### Thematisch verwandte Arbeiten

1. Text wurde in der Dokumentanalyse schon verschiedentlich als Textur betrachtet, allerdings immer nur in direktem Zusammenhang mit grundlegenden Segmentierungsaufgaben.

Um Lage und Abmessungen von Zeilen und Blöcken auf reinen Text-Dokumenten zu finden, wurde von O’Gorman ([OG93]) die „docstrum“-Methode vorgeschlagen. Sie kann als texturbasierte Methode gelten, auch wenn der Autor dies nicht explizit erwähnt. Grundelemente der Textur sind Connected Components, die entsprechend ihren wechselseitigen Abständen und Neigungen gruppiert werden.

Eine Aufgabe der Segmentierung ist es außerdem, die lokalisierten Objekte als Grundlage für eine weitere Bearbeitung in grobe Strukturklassen einzuteilen. Wang und Srihari beschreiben in [Wan89] ein Verfahren, das durch „run-length smoothing“ gruppierte Blöcke in die Klassen Halbtonbild, Text mit großer Schriftgröße, Text mit kleiner Schriftgröße und Liniengraphik einteilt. Dazu wird die Textur der Blöcke mit einer Art Frequenzanalyse von Schwarz-Weiß-Übergängen untersucht.

Jain und Bhattacharjee ([Jai92]) segmentieren ein Dokument durch Texturanalyse in die drei Klassen Text, gleichförmiger Bereich und Übergangsbereich. Dazu bilden sie zunächst mit acht zweidimensionalen Gabor-Filtern und anschließend mit einem sigmoiden Filter Merkmale, die dann über einen Abstandsklassifikator den drei Klassen zugeordnet werden. In [Jai96] präsentieren Jain und Zhong einen weiterentwickelten Algorithmus zur Texturanalyse auf Dokumenten mit dem Ziel, die vier Klassen Text, Halbtonbild, Liniengraphik und Hintergrund zu unterscheiden. Ein neuronales Netz lernt nun für das spezifische Segmentierungsproblem charakteristische Filter, die von Gabor-Filtern abgeleitet sind und diese ersetzen.

Etemad, Doermann und Chellappa ([Ete97]) ersetzen die Gabor-Filterung durch eine orthogonale Wavelet-Packet-Transformation, um auf einem Dokument Text-, Halbtonbild- und Graphikregionen zu klassifizieren. Die von der Wavelet-Transformation gelieferten Merkmalsvektoren werden lokal integriert und mit Hilfe eines neuronalen Netzes mit „weichen“ Entscheidungsschwellen klassifiziert.

---

<sup>2</sup> OCR: „Optical Character Recognition“ ist ein automatisches Verfahren, das durch Mustererkennung gedruckten (oder handgeschriebenen) Buchstaben eine entsprechende symbolische Repräsentation, z.B. das entsprechende ASCII-Zeichen, zuordnet. Im allgemeinen durchläuft das Verfahren ausgehend von einem Papierdokument die Verarbeitungsschritte Scannen des gesamten Dokuments, Segmentierung, Vorverarbeitung, Merkmalsextraktion, Erkennung und evtl. Nachbearbeitung.

Weitere Anwendungen der Texturanalyse in der Dokumentanalyse sind die Trennung von Text und Hintergrund ([Che97b]) und die automatische Erkennung von Skript und Sprache eines Dokuments ([Tan98]).

2. Die Theorie von Julesz wurde bereits von mehreren Autoren in einen Algorithmus zur Texturanalyse umgesetzt. Rearick entwirft in [Rea85] eine Methode zur Textursegmentierung in Bildern aus künstlichen und realen Texturen, die stark von der Juleszschen Vorstellung der „preattentive vision“ inspiriert ist. Grundlage der Methode ist die Berechnung der drei Texton-Merkmale Dichte der Endpunkte, integrierte Texton-Länge und Dichte von Ecken nach der Binarisierung des eingegebenen Grauwertbildes.

Clark und Bovik segmentieren in [Cla86] Bilder, die aus realen Texturen zusammengesetzt sind, und in [Cla89] Muster, die aus künstlichen Texton-Mustern nach Vorbildern bei Julesz gebildet wurden. Clark und Bovik nehmen an, daß das visuelle System des Menschen Gabor-Filter zur Wahrnehmung von Textons realisiert hat, und benutzen diese Filter daher zum Finden von Textons und zur Merkmalsextraktion. Ganz ähnliche Filter, nämlich Differenzen von gegeneinander verschobenen Gauß-Funktionen, verwenden auch Malik und Perona in [Mal90], um künstliche Texton-Muster nach Julesz zu segmentieren.

Voorhees und Poggio beschreiben ausgehend vom psychologischen Begriff des Textons in [Voo88] eine Methode, um nicht nur künstliche Bilder aus realen Texturen, sondern tatsächlich Grauwertbilder realer Szenen in Bereiche verschiedener Textur zu segmentieren. Die verwendeten Textons sind dunkle Flecken, die hier durch Filterung des Originals mit einem Laplace-Filter erzeugt werden. Für diese Textons werden die Merkmale Kontrast, Orientierung, Breite, Länge, Fläche und Flächendichte berechnet und mit Hilfe einer lokalen Statistik so ausgewertet, daß Texturgrenzen lokalisiert werden können. Voorhees und Poggio bezweifeln, daß die von Julesz propagierten Textons namens Überschneidung und Endpunkt neben den diskreten Wahrnehmungseinheiten („Flecken“) überhaupt eine Rolle spielen. Interessanterweise ist auch in der Sichtweise der vorliegenden Arbeit nur der Begriff der diskreten Wahrnehmungseinheit maßgebend.

Die zitierten Veröffentlichungen zeigen, daß bei der Analyse von realen Texturen das Problem auftaucht, die Textons als diskrete Einheiten im Bild zu definieren, und man sich im allgemeinen mit einer passenden Filterung des Grauwertbildes behilft. Da in der vorliegenden Arbeit Binärbilder benutzt werden, taucht dieses Problem hier nicht auf, denn durch die Binarisierung werden automatisch diskrete Einheiten auf dem Dokument gebildet. Sie werden dann die Grundlage der Textons.



# 3 Formalisierung der Begriffe

## 3.1 Übersicht

Die texturbasierte Attraktivitätsanalyse ist inspiriert von der psychologischen Texton Theorie von Julesz ([Jul83]), die qualitativ die menschliche Wahrnehmung von Texturen beschreibt. Für eine technisch eindeutige Beschreibung und eine quantitative Umsetzung der dort dargestellten Grundsätze werden im folgenden die zentralen Punkte der Theorie, Textur und Texton, formal gefaßt. In einer Weiterentwicklung wird außerdem eine Texturhierarchie eingeführt, die sich speziell dafür eignet, die Verwobenheit von Texturen auf Dokumenten zu beschreiben. Außerdem wird bereits hier der später bei der Berechnung eines Attraktivitätsmaßes benötigte Begriff der Bewertung definiert.

Abbildung 3-1 zeigt, daß man im Rahmen einer formalen Darstellung prinzipiell trennt zwischen der Welt der Objekte, dargestellt durch den Objektraum, und der Welt der Beobachtungen, dargestellt durch den Merkmalsraum. Der Objektraum enthält die realen Objekte, während der Merkmalsraum Repräsentationen der Objekte enthält, die durch das Vermessen der Objekte zustande gekommen sind. Die entsprechende Messung wird durch eine Registrierungsfunktion beschrieben.

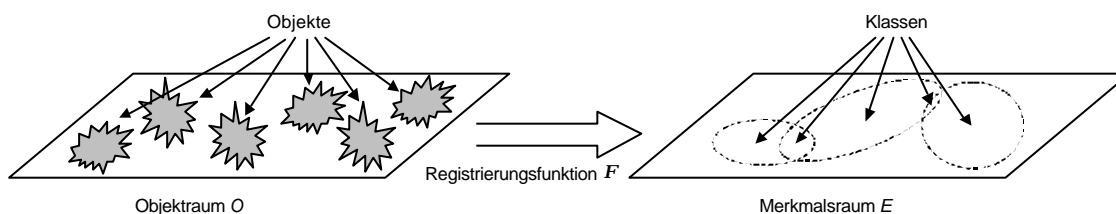


Abbildung 3-1: Rahmen der Formalisierung – Objektraum (Welt der realen Objekte), Registrierungsfunktion und Merkmalsraum (Welt der Beobachtungen)

## 3.2 Texton, Textur und Texturhierarchie

In der psychologischen Texton Theorie wird eine Textur als eine Anhäufung von ähnlichen Wahrnehmungseinheiten definiert, wobei diese Wahrnehmungseinheiten räumlich zusammenhängen müssen. Um diese Vorstellung zu formalisieren, wird im folgenden zuerst der Begriff der Wahrnehmungseinheiten, d.h. der Textons, formalisiert.

Zunächst wird ein Objektraum als endlicher Ausschnitt aus einem metrischen Raum definiert. Dieser Objektraum entspricht aus psychologischer Sicht unserem begrenzten Blickfeld, in dem man Objekte wahrnehmen kann, wobei im formalen Rahmen in Zukunft nicht von Wahrnehmung sondern allgemeiner von Messung die Rede sein wird.

### Definition:

Ein Paar

$$M = (X, d) \tag{3-1}$$

heißt *metrischer Raum (Meßraum)*, wenn

1.  $X$  ist eine Menge. Ihre Elemente werden als Punkte des Raums bezeichnet.
2.  $d: X \times X \rightarrow \mathbf{R}_0^+$  ist eine Abstandsfunktion, d.h. für alle  $x, y, z \in X$  gilt
  - a)  $d(x, y) = 0 \Leftrightarrow x = y$
  - b)  $d(x, y) = d(y, x)$
  - c)  $d(x, y) + d(y, z) \geq d(x, z)$

**Definition:**

Ein Paar

$$O = (M, \mathbf{G}) \tag{3-2}$$

heißt *Objektraum* über dem Meßraum  $M$ , wenn

1.  $M = (X, d)$  ist ein metrischer Raum mit  $X \neq \emptyset$ .
2.  $\emptyset \neq \mathbf{G} \subseteq \text{Pow}(X) - \emptyset$  und für alle  $g, h \in \mathbf{G}$  gilt:  $g \cap h = \emptyset$ .
3.  $\mathbf{G}$  ist endlich.
4. Es gibt ein  $\mathbf{a} > 0$  und ein  $x \in X$  so, daß für alle  $g \in \mathbf{G}$  gilt:  
 $g \subset U(x, \mathbf{a}) = \{y: y \in X \wedge d(x, y) \leq \mathbf{a}\}$ .  
 Man nennt alle  $g \in \mathbf{G}$  *Objekte* und  $\mathbf{G}$  die *Menge aller Objekte*.

Mit diesen beiden Definitionen ist ein Objektraum also ein Raum, in dem sich mindestens ein Objekt befindet. Die Objekte sind nicht-leere Mengen von Punkten im Meßraum und außerdem paarweise disjunkt. Diese Modellierung spiegelt die Darstellung der Texton Theorie wieder, nach der die Grundeinheiten wahrgenommen, d.h. vermessen werden können und nach der sie voneinander unterscheidbar sind. Daß diese Objekte nur in einem begrenzten Blickfeld wahrgenommen werden können, wird durch Bedingung 4. modelliert.

Dem Objektraum steht der Merkmalsraum gegenüber.

**Definition:**

Ein *Merkmalsraum* ist das kartesische Produkt

$$E = E_1 \times \dots \times E_n, \tag{3-3}$$

mit den Mengen  $E_i \subseteq \mathbf{R}$  der möglichen Ausprägungen des Merkmals  $i, i = 1 \dots n$ .

Dabei muß jede Menge  $E_i$

1. beschränkt sein und
2. mindestens zwei Elemente enthalten, da eine Messung des Merkmals sonst sinnlos wäre.

Auf psychologischer Seite fordert die Texton Theorie, daß die benutzten Merkmale relativ einfach und direkt meßbar sein sollen. Diese Forderung kann nicht formal beschrieben werden, sie sollte jedoch bei einer späteren technischen Umsetzung so weit wie möglich berücksichtigt werden. Dabei ist zu beachten, daß für Menschen andere Merkmale direkt meßbar sind als für ein technisches Verfahren. Der Merkmalsraum wird also üblicherweise technisch direkt meßbare Merkmale („unmittelbare Merkmale“) und aus den unmittelbaren Merkmalen berechnete Merkmale („mittelbare Merkmale“) enthalten.

Bei einer Messung wird einem Objekt aus dem Objektraum ein *Merkmalsvektor*  $\langle e_1, \dots, e_n \rangle$  aus dem Merkmalsraum zugewiesen. Dieser Vorgang wird durch die Registrierungsfunktion beschrieben.



**Definition:**

Eine *Registrierungsfunktion* ist eine Abbildung

$$F: \mathbf{G} \rightarrow E, \quad (3-4)$$

die jedem Objekt  $\mathbf{g} \in \mathbf{G}$  aus einem Objektraum  $O$  einen Merkmalsvektor aus dem Merkmalsraum  $E$  zuordnet.

Im Merkmalsraum sind die realen Objekte nun also durch einen Vektor ausgewählter Merkmale repräsentiert, wobei die Auswahl der Merkmale und die Festlegung der Registrierungsfunktion eine Frage der konkreten Implementierung, nicht der formalen Darstellung, ist. An dieser Stelle werden bei der Implementierung die Erkenntnisse von Julesz einfließen können.

Die Verbindung eines Objektraumes mit einem Merkmalsraum über eine Registrierungsfunktion wird durch die sogenannte Textonstruktur dargestellt.

**Definition:**

Eine *Textonstruktur*  $TS$  ist ein Tripel

$$TS = (O, E, F) \quad (3-5)$$

mit den Eigenschaften

1.  $O$  ist ein Objektraum.
2.  $E$  ist ein Merkmalsraum.
3.  $F$  ist eine Registrierungsfunktion.

In einer solchen Textonstruktur  $TS$  heißt ein Paar  $t = (\mathbf{g}, F(\mathbf{g}))$  *Texton*. Ein Texton ist also die Kombination eines Objekts mit seinen über eine Registrierungsfunktion gemessenen Merkmalen, ganz ähnlich der Juleszschen Definition.

Im folgenden bezeichnet  $Dom(t)$  den *Träger*  $\mathbf{g}$  und  $Prop(t)$  den *Merkmalsvektor*  $F(\mathbf{g})$  des Textons  $t$ , wenn  $t = (\mathbf{g}, F(\mathbf{g}))$ . Da Textons über paarweise disjunkten Objekten aufgebaut sind, gilt selbstverständlich für zwei Textons  $s$  und  $t$ , daß  $Dom(s) \cap Dom(t) \neq \emptyset \Leftrightarrow s = t$ .

$TEX_{O,E,F}$  soll die Menge aller Textons einer gegebenen Textonstruktur  $TS = (O, E, F)$  bezeichnen. In eindeutigen Fällen schreibt man  $TEX$  anstelle von  $TEX_{O,E,F}$ .

Wie bereits dargestellt, werden die Texturen, die hier modelliert werden sollen, als Anhäufungen von ähnlichen, räumlich benachbarten Textons aufgefaßt. Das nächste Ziel der Formalisierung wird es also sein, die Begriffe der Ähnlichkeit und des räumlichen Zusammenhangs zu definieren. Dazu werden zwei Relationen eingeführt.

**Definition:**

Sei  $E$  ein Merkmalsraum, dann ist

$$S \subseteq E \times E \quad (3-6)$$

eine *Ähnlichkeitsrelation*, wenn  $S$  reflexiv, symmetrisch und transitiv, also eine Äquivalenzrelation, ist.

Da  $\mathcal{S}$  eine Äquivalenzrelation ist, zerlegt sie den Merkmalsraum, in dem sich die den Objekten zugeordneten Merkmale befinden, in disjunkte Äquivalenzklassen. Dabei ist eine Äquivalenzklasse wie folgt definiert:

Sei  $A$  eine beliebige Menge und  $Z \subseteq A \times A$  eine Äquivalenzrelation. Dann ist eine Äquivalenzklasse  $Z[x]$  gegeben durch  $Z[x] := \{y: (x, y) \in Z\}$  und es gilt

- a)  $y \in Z[x] \Rightarrow Z[y] = Z[x]$   
 b)  $Z[y] \cap Z[x] = \emptyset$ , falls  $y \notin Z[x]$ . (3-7)

Zwei Textons  $s$  und  $t$  sind bezüglich einer Ähnlichkeitsrelation  $\mathcal{S}$  genau dann *ähnlich*, wenn  $(Prop(s), Prop(t)) \in \mathcal{S}$ , ihre beiden Merkmalsvektoren also in einer Äquivalenzklasse des Merkmalsraumes liegen.

Die Ähnlichkeitsrelation  $\mathcal{S}$  wird in der Realität im allgemeinen aus zwei Anteilen bestehen, die unterschiedlich definiert sind:

1. Zum einen kann man eine Ähnlichkeitsrelation über ein Klassifikationssystem definieren. Dieses Klassifikationssystem wird aus einer Reihe von Erzeugendenmengen („Grundklassen“) gebildet, die explizit vorgegeben werden und im Merkmalsraum Merkmalsvektoren zu sogenannten *expliziten Klassen* zusammenfassen. Aus den Durchschnitten und Vereinigungen dieser Erzeugenden ergeben sich weitere explizite Klassen.

Gibt man beispielsweise für das grundlegende Klassifikationssystem die Erzeugenden „Fettdruck“ und „Kursivdruck“ vor, so ergibt sich durch Überschneidung dieser beiden Klassen die weitere Klasse „Fett-Kursivdruck“. Es existieren letztendlich drei disjunkte Klassen, innerhalb derer ein Ähnlichkeitsbegriff über die Zugehörigkeit zu diesen Klassen gegeben ist.

Diese so entstehenden disjunkten Klassen kann man selbstverständlich mit Hilfe einer Äquivalenzrelation beschreiben, die hier  $\mathcal{S}_K$  heißen soll.

2. Auf der anderen Seite gibt es aber auch einen Ähnlichkeitsbegriff, der durch explizite Klassen nicht beschrieben werden kann. So ist es üblich, zwei Textons dann als ähnlich zu bezeichnen, wenn für beide der Meßwert eines Merkmals innerhalb einer gewissen Toleranzgrenze übereinstimmt. Man bezeichnet zwei Textons z.B. als ähnlich, wenn sie in etwa die gleiche Höhe haben, ohne daß man explizit angibt welchen Wert die Messung der Höhe ergeben muß.

Um solche *implizite Klassen* zu beschreiben, benutzt man eine weitere Äquivalenzrelation, die hier  $\mathcal{S}_R$  heißen soll. Die entstehenden Äquivalenzklassen können sich mit den durch  $\mathcal{S}_K$  gebildeten Klassen durchaus überschneiden.

$\mathcal{S}$  setzt sich also intern aus den zwei Anteilen  $\mathcal{S}_K$  und  $\mathcal{S}_R$  zusammen, und zwar als

$$\mathcal{S} = \mathcal{S}_K \cap \mathcal{S}_R. \quad (3-8)$$

Die Äquivalenzrelation zerlegt also den Merkmalsraum in Äquivalenzklassen. Dies bewirkt, daß auch jedem Texton, zu dem ja immer ein Merkmalsvektor des Merkmalsraumes gehört,

eine Klasse zugeordnet wird. Diesen Vorgang nennt man Klassifikationsprozeß und sein Ergebnis eine Klassifikation.

**Definition:**

Sei  $TS = (O, E, F)$  eine Textonstruktur. Dann heißt für jedes  $t \in TEX$

$$C(t) := \mathcal{S}[Prop(t)] \quad (3-9)$$

die *Klassifikation* von  $t$  in  $TS$ .

Eigentlich müßte man hier von *endgültiger Klassifikation* sprechen, da man oft auch die Zuordnung eines Textons zu den expliziten Klassen als Klassifikation bezeichnet. Man würde beispielsweise davon sprechen, ein Wort zunächst als fettgedruckt und danach als kursiv gedruckt zu klassifizieren. Dies entspricht der Zuordnung zu den beiden expliziten Klassen „Fettdruck“ und „Kursivdruck“. Die endgültige Klassifikation des Wortes wäre aber die Klasse „kursiver Fettdruck“.

Ohne Klassifikation ist das Auffinden von Texturen nicht möglich, da durch Klassifikation ein Ähnlichkeitsbegriff nicht nur im Merkmalsraum, sondern auch im Objektraum definiert werden kann.

Die räumliche Beziehung zwischen Textons soll durch eine Nachbarschaftsrelation beschrieben werden.

**Definition:**

Sei  $TS = (O, E, F)$  eine Textonstruktur und  $TEX$  die Menge aller Textons. Dann heißt

$$N \subseteq TEX \times TEX \quad (3-10)$$

*Nachbarschaftsrelation* in  $TEX$ , wenn

1.  $\forall t \in TEX: (t, t) \in N$
2.  $\forall s, t \in TEX: (s, t) \in N \rightarrow (t, s) \in N$ .

Man beachte, daß die Nachbarschaftsrelation räumliche Zusammenhänge in der Regel unter Einbeziehung von Merkmalen der Textons beschreibt. Beispielsweise werden Buchstaben aufgrund ihres gegenseitigen Abstandes zu Texturen, den Wörtern, zusammengefaßt. Dieser Abstand hängt aber von der Buchstabenhöhe ab.

Da nun Ähnlichkeit und Nachbarschaft definiert sind, kann man im nächsten Schritt als Grundlage des Texturbegriffes definieren, wann zwei Textons zusammenhängend sind.

**Definition:**

Sei  $TS = (O, E, F)$  eine Textonstruktur,  $TEX$  die Menge aller Textons,  $\mathcal{S} \subseteq E \times E$  eine Ähnlichkeitsrelation und  $N \subseteq TEX \times TEX$  eine Nachbarschaftsrelation. Dann sind zwei Textons  $s, t \in TEX$   *$\mathcal{S}$ - $N$ -zusammenhängend*, wenn es eine Kette  $t_1, \dots, t_n \in TEX$  gibt, so daß:

1.  $s = t_1$  und  $t = t_n$
2.  $\forall i = 1, \dots, n-1: (t_i, t_{i+1}) \in N$  und  $(Prop(t_i), Prop(t_{i+1})) \in \mathcal{S}$ . (3-11)

Es ist offensichtlich, daß  $t$  auch  *$\mathcal{S}$ - $N$ -zusammenhängend* mit  $s$  ist, wenn  $s$   *$\mathcal{S}$ - $N$ -zusammenhängend* mit  $t$  ist.

**S-N-Zusammenhang** ist der Schlüssel zur Definition von Texturen, weil Texturen nach der Grundidee aus zusammenhängenden Textons bestehen.

**Definition:**

Sei  $TS = (O, E, F)$  eine Textonstruktur,  $TEX$  die Menge aller Textons mit  $s, t \in TEX$ ,  $S \subseteq E \times E$  eine Ähnlichkeitsrelation und  $N \subseteq TEX \times TEX$  eine Nachbarschaftsrelation. Dann ist  $T$  eine **S-N-Textur**, wenn

1.  $\forall s, t \in T: s$  und  $t$  sind **S-N-zusammenhängend** und
2.  $\forall s \notin T: s$  gibt kein  $t \in T$ , so daß  $s$  und  $t$  **S-N-zusammenhängend** sind. (3-12)

Nach dieser Definition sind Texturen maximale Mengen von zusammenhängenden Textons. Diese Mengen sind disjunkt, denn seien  $S$  und  $T$  **S-N-Texturen**, dann gilt:  $T \cap S \neq \emptyset \Leftrightarrow T = S$ . Das heißt, daß sich Texturen nie Textons teilen. Außerdem gilt für Texturen  $S$  und  $T$  mit  $S \neq T$ , daß sich die Mengen  $\bigcup_{s \in S} Dom(s)$  und  $\bigcup_{t \in T} Dom(t)$  nicht überschneiden.

Als Spezialfall gibt es selbstverständlich auch Texturen, die nur ein Texton enthalten. Die Elemente solcher Texturen hängen nur mit sich selbst zusammen, d.h. sie sind nur sich selbst ähnlich und/oder nur zu sich selbst benachbart.

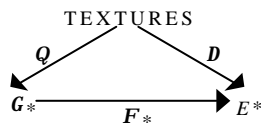
Wie bereits in Kapitel 2.2.2 erläutert, wurde die Texton Theorie anhand relativ einfach strukturierter Beispiele aufgestellt. Betrachtet man komplexere Texturen, z.B. wie in dieser Arbeit Dokumente, fällt auf, daß Texturen Hierarchien bilden können. Eine *Texturhierarchie* ergibt sich wie folgt:

**Lemma:**

Sei  $TS = (O, E, F)$  eine Textonstruktur,  $TEX$  die Menge aller Textons und  $TEXTURES$  die Menge aller **S-N-Texturen**, die durch eine Ähnlichkeitsrelation  $S \subseteq E \times E$  und eine Nachbarschaftsrelation  $N \subseteq TEX \times TEX$  erzeugt werden.

Sei außerdem  $E^* = E^*_1 \times \dots \times E^*_m$  ein Merkmalsraum,  $D:TEXTURES \rightarrow E^*$  und  $Q:TEXTURES \rightarrow G^*$ , wobei  $G^* = \{ \bigcup_{t \in T} Dom(t) : T \in TEXTURES \}$  und  $Q(T) = \bigcup_{t \in T} Dom(t)$ .

Dann existiert eine Funktion  $F^*:G^* \rightarrow E^*$ , so daß das Diagramm



kommutiert, d.h.  $F^*(Q(T)) = D(T)$ . (3-13)

Da  $G^*$  die Bedingungen 2. – 4. der Definition (3-2) erfüllt, und  $F^*$  eine Registrierungsfunktion nach Definition (3-4) ist, ist  $TS^* = (O^*, E^*, F^*)$  eine Textonstruktur.

Die Abbildung  $D$  reflektiert, daß die Merkmale eines auf höherer Hierarchiestufe konstruierten Textons aus Merkmalen der Textons berechnet werden, die zur Konstruktion benutzt wurden. Der theoretische Rahmen erlaubt es auch, den Textons einer höheren Stufe statistische Merkmale zuzuweisen, wie es später bei der Klassifikation einiger Layoutmerkmale geschieht.

Nachdem Textons einer höheren Hierarchiestufe aus Texturen (Mengen von Textons einer niedrigeren Stufe) aufgebaut werden, schließen sich die Textons in der Hierarchie natürlicherweise ein. Um anzugeben, daß ein Texton  $t$  einer niedrigeren Stufe an der Konstruktion eines Textons  $s$  auf höherer Stufe beteiligt war, schreibt man  $t \succ s$ , insbesondere auch  $Dom(t) \succ Dom(s)$ .

Im allgemeinen benutzt die Dokumentanalyse eher einfache Merkmale, um Textons zu charakterisieren. Dazu gehören Abmessungen und Position eines Textons  $t$  wie folgt:

$$\begin{aligned} Top(t) &= \max\{y : (x, y) \in Dom(t)\}, & Bot(t) &= \min\{y : (x, y) \in Dom(t)\} \\ Left(t) &= \min\{x : (x, y) \in Dom(t)\}, & Right(t) &= \max\{x : (x, y) \in Dom(t)\} \\ Height(t) &= Top(t) - Bot(t) + 1, & Width(t) &= Left(t) - Right(t) + 1 \end{aligned} \quad (3-14)$$

Damit kann man eine Abbildung  $Prop(t) = (Top(t), Bot(t), Left(t), Right(t), Height(t), Width(t))$  definieren, die einen grundlegenden Merkmalsvektor liefert. Von diesen grundlegenden geometrischen Merkmalen abgesehen, kann ein Texton viele zusätzliche Merkmale haben, z.B. Fläche, Orientierung und, für Textons höherer Hierarchiestufe, die Anzahl der Textons niedrigerer Stufe, die zur Konstruktion benutzt wurden, usw..

### 3.3 Segmentierung und Resegmentierung

In der Bildverarbeitung wird häufig von Segmentierung gesprochen, wenn das Blickfeld in Teilbereiche zerlegt wird. Betrachtet man Segmentierung auf der Basis des Texturbegriffs, dann sind diese Teilbereiche gerade die Texturen. Segmentierung ist also die Zerlegung des Blickfeldes in Texturen.

#### Definition:

Sei  $TS = (O, E, F)$  eine Textonstruktur,  $TEX$  die Menge aller Textons und  $TEXTURES$  die Menge aller  $S-N$ -Texturen, die durch eine Ähnlichkeitsrelation  $S \subseteq E \times E$  und eine Nachbarschaftsrelation  $N \subseteq TEX \times TEX$  erzeugt werden. Dann heißt ein Prozeß zur Bestimmung aller  $S-N$ -Texturen von  $TS$  ein *Segmentierungsprozeß* und die Menge  $TEXTURES$  eine *S-N-Segmentierung* von  $TEX$ , kurz eine *Segmentierung* von  $TEX$ . (3-15)

Im folgenden wird der Segmentierungsprozeß dem üblichen Sprachgebrauch entsprechend auch als Segmentierung bezeichnet, da im Kontext normalerweise keine Verwechslungsmöglichkeit besteht.

Eine Segmentierung nach Definition (3-15) ist immer vollständig, da jedes Objekt einer Textur zugeordnet wird, und sei es einer Textur, die nur ein Element enthält.

Man stellt zudem fest, daß der hier eingeführte Segmentierungsbegriff nicht absolut, sondern abhängig von den ausgewählten Merkmalen, d.h. insbesondere von Ähnlichkeits- und Nachbarschaftsrelation ist. Dies spiegelt zum einen deutlich die Realität wieder, bei der je nach wahrgenommenen Merkmalen und Interpretation wahrgenommener Merkmale sowohl Menschen als auch Computer zu verschiedenen Segmentierungen einer visuellen Szene, z.B. eines Dokuments, kommen können.

Zum anderen ermöglicht es, den Begriff der *Resegmentierung* einzuführen. Die Resegmentierung erlaubt es, bei einer erneuten Segmentierung eines bereits segmentierten Bildes mehr Details herauszuarbeiten, ohne bereits erzielte Ergebnisse zu verändern, d.h. verfeinerte Tex-

turen zu erzeugen. Dazu verfeinert man Ähnlichkeitsrelation und/oder Nachbarschaftsrelation, wobei die Verfeinerungen wie folgt definiert sind:

**Definition:**

Sei  $A$  eine beliebige Menge, und seien  $Z \subseteq A \times A$  und  $Z^* \subseteq A \times A$  Äquivalenzrelationen, dann heißt  $Z^*$  *feiner* als  $Z$ , wenn es für jede Äquivalenzklasse  $Z[x]$  in  $A$  eine Äquivalenzklasse  $Z^*[x]$  in  $A$  gibt, so daß  $Z^*[x] \subseteq Z[x]$ . (3-16)

$Z^*$  ist also genau dann feiner als  $Z$ , wenn  $Z^*$   $A$  in mehr Klassen zerlegt, aber die durch  $Z$  vorgegebenen Grenzen respektiert.

Diese Definition der Verfeinerung gilt zunächst nur für die Ähnlichkeitsrelation, da diese eine Äquivalenzrelation ist. Nachbarschaftsrelationen sind dagegen nicht transitiv und deshalb auch keine Äquivalenzrelationen. Von Verfeinerung einer Nachbarschaftsrelation kann man aber trotzdem sprechen, sobald man nicht mehr die Nachbarschaftsrelation selbst, sondern ihren transitiven Abschluß betrachtet. Da die Verfeinerung bereits bestehende Texturen behandelt, ist dies auch erlaubt, denn zwei Elemente einer Textur liegen immer auch im transitiven Abschluß der Nachbarschaftsrelation.

An dieser Stelle sei ausdrücklich darauf hingewiesen, daß die Texturhierarchie nicht mit der Resegmentierung gekoppelt ist. Innerhalb jeder Hierarchiestufe finden Segmentierung und Resegmentierung statt. Dann werden die sich ergebenden Texturen zu neuen Textons auf einer höheren Hierarchiestufe. In dieser höheren Hierarchiestufe gibt es neben den neuen Textons normalerweise auch eine völlig neue Ähnlichkeits- und Nachbarschaftsrelation. Bei einer Resegmentierung dagegen werden die gleichen Textons wie bei der Segmentierung behandelt, nur die bestehenden Relationen werden verfeinert (Abbildung 3-2).

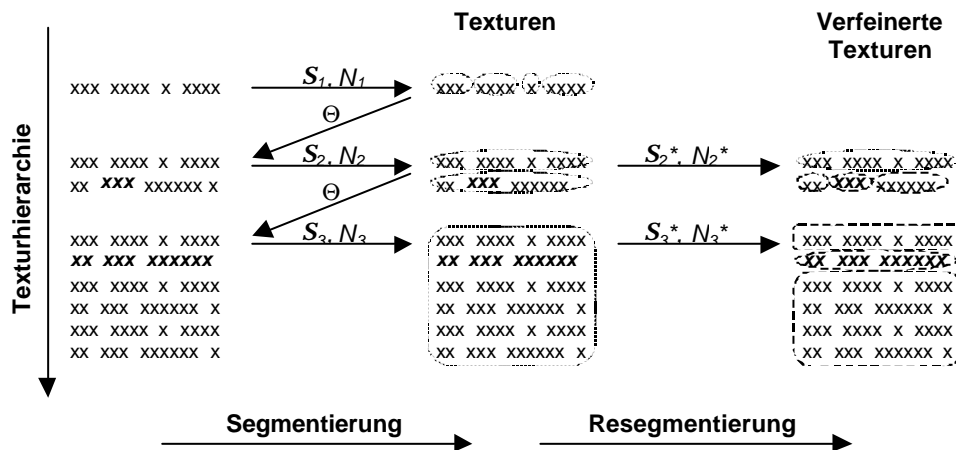


Abbildung 3-2: Texturhierarchie versus Segmentierung und Resegmentierung

**3.4 Bewertung**

Das Ziel der Arbeit ist es, die Attraktivität von Objekten, bzw. Textons, zu bewerten. Daher soll nun in Weiterführung des formalen Rahmens der Begriff der Bewertung definiert werden.

**Definition:**

Sei  $TS = (O, E, F)$  eine Textonstruktur. Dann heißt eine Abbildung

$$\mathbf{m}^*: E = E_1 \times \dots \times E_n \rightarrow \mathbf{R} \quad (3-17)$$

eine *Bewertung* in  $TS$ , wenn gilt:

1.  $\forall \langle e_1, \dots, e_n \rangle \in E: \mathbf{m}^*(\langle e_1, \dots, e_n \rangle) \geq 0$
2.  $\exists \mathbf{b} > 0$  so, daß  $\forall t \in TEX: \mathbf{m}^*(\langle e_1, \dots, e_n \rangle) = \mathbf{m}^*(Prop(t)) \leq \mathbf{b}$

Nach dieser Definition ist eine Bewertung also eine nicht-negative, beschränkte Abbildung aus dem Merkmalsraum nach  $\mathbf{R}$ . Sie induziert auf natürliche Weise eine Abbildung

$$\mathbf{m} \text{ } TEX \rightarrow \mathbf{R}, \text{ mit } \mathbf{m}(t) := \mathbf{m}^*(Prop(t)). \quad (3-18)$$

Aus dieser Festlegung des Bewertungsbegriffs wird sofort klar, daß man die Bewertung eines Textons auch als mittelbares, d.h. berechnetes, Merkmal des Textons auffassen kann.

In den meisten Fällen basiert die Bewertung auf komponentenweisen Bewertungen, d.h. für jedes Merkmal  $i$  gibt es eine nicht-negative und beschränkte Funktion

$$\mathbf{m}_i^*: E_i \rightarrow \mathbf{R} \quad (3-19)$$

die jedes Merkmal einzeln bewertet. Die einzelnen Bewertungen werden dann zu einer Gesamtbewertung aufaddiert. Man definiert also eine *additive Bewertung*

$$\mathbf{m}^*(\langle e_1, \dots, e_n \rangle) = \sum_{i=1, \dots, n} \mathbf{m}_i^*(e_i) \quad (3-20)$$

Die später berechnete Attraktivitätsbewertung ist gerade eine solche additive Bewertung.

In einer Texturhierarchie wird man die Bewertung im allgemeinen als Abbildung

$$\mathbf{m}: \bigcup_{j \text{ ist Stufe}} TEX_j \rightarrow \mathbf{R} \quad (3-21)$$

einführen und folgendes zusätzlich verlangen:

1. Für jede Stufe  $j$  ist  $\mathbf{m}|_{TEX_j}$  (Beschränkung von  $\mathbf{m}$  auf  $TEX_j$ ) eine Bewertung.
2. Ist  $t_j$  ein Texton der Stufe  $j > 0$  und gilt für ein Texton  $t_{j-1}$  der Stufe  $j-1$ , daß  $t_{j-1} \succ t_j$  ist, so ist  $\mathbf{m}(t_{j-1}) \geq \mathbf{m}(t_j)$ .

Die zweite Bedingung besagt also, daß Textons einer niedrigeren Hierarchiestufe eine höhere Bewertung erhalten, als Textons einer höheren Hierarchiestufe. Das ist insofern einleuchtend, als die Textons niedrigerer Hierarchiestufe aufgrund einer (oder mehrerer) ihrer vielen Merkmale zu Texturen und damit Textons höherer Hierarchiestufe zusammengefaßt werden. Das Texton höherer Hierarchiestufe ist also eine Generalisierung der Textons niedrigerer Hierarchiestufe, umgekehrt sind die Textons niedrigerer Hierarchiestufe Spezialisierungen. Diese Spezialisierungen haben neben der zur Texturbildung benutzten Merkmale noch weitere Merkmale, die bewertet werden müssen, so daß ihre Gesamtbewertung zwangsläufig höher ist als die der Textons höherer Hierarchiestufe.

### 3.5 Thematisch verwandte Arbeiten

1. Clark und Bovik ([Cla89]) segmentieren Texturen, die sich aus Textons Julesz'scher Definition zusammensetzen, durch Filterung mit Gabor-Funktionen. Um diesen Vorgang signaltheoretisch zu beschreiben, formalisieren sie das von ihnen benutzte Texturmodell. Eine Textur ist definiert als

$$f(x, y) = \sum_m \sum_n t(x - mX, y - nY) = t(x, y) * \sum_m \sum_n \mathbf{d}(x - mX, y - nY) \quad ((10) \text{ in } [Cla89]),$$

wobei ein Texel  $t(x, y)$  nur als „Anordnung von Textons“, nicht aber formal beschrieben ist.

Die Texel sind regulär auf dem Gitter

$$\sum_m \sum_n \mathbf{d}(x - mX, y - nY)$$

angeordnet. Dieses Gitter beschreibt also die Nachbarschaftsbeziehung zwischen Textur-elementen. Im oben eingeführten formalen Rahmen könnte man das durch eine entsprechend gewählte Nachbarschaftsrelation  $N$  ausdrücken, wobei  $N$  wohlgemerkt nicht nur starre Gitteranordnungen sondern beliebige räumliche Anordnungen zuläßt.

Die Ähnlichkeit der Textons wird in [Cla89] durch die Gabor-Filter gemessen. Nach Clark und Bovik realisiert die Filterung mit Gabor-Funktionen gleichzeitig eine Bestimmung der Texton-Merkmale (entsprechend dem obigen  $Prop(t)$ ) und einen Vergleich der Merkmale (entsprechend der obigen Ähnlichkeitsrelation  $\mathcal{S}$ ), letzteres angezeigt durch die Amplitude des gefilterten Bildes.

2. Weitere Möglichkeiten, Textur zu formalisieren, werden bei Ballard und Brown (Kapitel 6 in [Bal82]) genannt. Ballard und Brown gehen wie diese Arbeit davon aus, daß sich Texturen aus visuellen Grundeinheiten zusammensetzen, die einige unveränderliche Merkmale haben und sich im Bereich der Textur an verschiedenen Positionen, in verschiedener Orientierung und lokaler Ausprägung wiederholen. Diese Grundeinheiten werden nicht Textons, sondern „Texel“ genannt.

Bei räumlich stark strukturierten Texturen steht eine Beschreibung der Nachbarschaftsbeziehung zwischen den Texeln im Vordergrund, wobei dies nicht wie in der oben stehenden Formalisierung durch eine Nachbarschaftsrelation  $N$  geschieht, sondern durch Grammatiken, z.B. durch

- „Shape“-Grammatiken: relativ komplexe Texel einer bestimmten Form werden durch sogenannte Formregeln aneinandergelinkt.
- „Tree“-Grammatiken: in einer zweistufigen Grammatik wird durch eine erste Grammatik ein Baum definiert, der die reguläre Anordnung von Texeln beschreibt, anschließend wird durch eine zweite Grammatik die Zusammensetzung jedes Texels aus Pixeln definiert.
- „Array“-Grammatiken: ähnlich wie bei Tree-Grammatiken erfolgt die Definition zweistufig, allerdings wird die Lage der Texel nicht durch Baumstrukturen sondern durch den Gebrauch von Platzhaltern in der Grammatik definiert.

Bei räumlich wenig strukturierten Texturen müssen nicht unbedingt Texel definiert werden. Vielmehr eignen sich zur Modellierung solcher Texturen statistische Modelle. Sie be-



schreiben Texturen so, daß sie mit Methoden der statistischen Mustererkennung segmentiert werden können, also z.B. mit Abstandsklassifikatoren oder mittels Wahrscheinlichkeitsdichtefunktionen. Grundlage dafür sind Merkmalsvektoren, die unterschiedlich berechnet werden können, z.B.

- aus dem Leistungsspektrum: wenn eine Textur räumlich periodisch ist, hat das per Fourier-Transformation ermittelte Spektrum für die korrespondierenden räumlichen Frequenzen Maxima, die die Textur charakterisieren. Dieser Zusammenhang läßt sich auch im Ortsbereich nachweisen.
- durch die räumliche Abhängigkeit von Graustufen: in einem Zwischenschritt wird eine Matrix berechnet, in der jedes Element angibt, wie oft zwei Grauwerte in einer bestimmten Lage zueinander auftreten. Aus dieser Matrix werden dann die eigentlichen Merkmale gewonnen. Diese Merkmale sind nicht mehr psychologisch fundiert.
- aus bildbasierten Texteln: die zu analysierende Textur wird als erstes in Regionen, z.B. Ellipsen, segmentiert, die dann als Textel aufgefaßt werden. Die Form der Regionen wird durch Parameter beschrieben, die dann die Merkmale zur Charakterisierung der Textur liefern.

Während die ersten beiden Beschreibungen von Textur von der Juleszschen Vorstellung diskreter Wahrnehmungseinheiten weit entfernt sind, und daher auch nicht mehr zu der obigen von Julesz abgeleiteten Formalisierung in Bezug gesetzt werden können, wird beim dritten Verfahren die Entstehung von Texteln erzwungen. Hier könnte Textur dann wieder mit den oben eingeführten Begriffen der Ähnlichkeit und der Nachbarschaft beschrieben werden.

3. Der Begriff des Textons wird in der Dokumentanalyse auch von Dori et al. ([Dor96]) benutzt, allerdings völlig anders definiert als bei Julesz ([Jul83]). Während ein Texton bei Julesz eine Einheit der Wahrnehmung ist – und in dieser Arbeit entsprechend eine Einheit im Bild des Dokuments –, ist ein Texton bei Dori eine Einheit in der logischen Struktur eines Dokuments. Bezogen auf diese logische Struktur bilden die Textons bei Dori eine Hierarchie, weil ein Texton selbst wieder Textons enthalten kann.



# 4 Texturbasierte Attraktivitätsanalyse

## 4.1 Segmentierung

### 4.1.1 Segmentierung von Text

Kapitel 3.3 hat bisher allgemein beschrieben, daß ein Segmentierungsprozeß die Menge aller Textons einer Textonstruktur in Teilmengen, d.h. Texturen, zerlegt. Nun soll gezeigt werden, wie dies im konkreten Fall auf Dokumenten geschieht. Hier kann man die bisher als Objekte bezeichneten Einheiten in Kombination mit ihren Merkmalen auch als Textons auffassen.

Auf dem Binärbild des eingescannten Dokuments werden zunächst die Zusammenhangskomponenten, englisch „Connected Components“, berechnet, wobei Connected Components Gebiete von 8-zusammenhängenden<sup>3</sup> schwarzen Runs<sup>4</sup> sind. Für die hier vorliegende Arbeit werden Zwischenergebnisse dieser Connected-Components-Berechnung weiterverwendet. Die Zwischenergebnisse sind strichartige Objekte, die jeweils Träger eines Strokes  $Dom(s)$  sind. Der Träger eines Strokes  $s$  ist ein Gebiet von 8-zusammenhängenden schwarzen Runs, das keine Split- oder Merge-Runs enthält. Dabei zeichnet sich ein Split bzw. Merge-Run dadurch aus, daß es in der nächsten bzw. vorhergehenden Bildzeile zwei Runs gibt, die in horizontaler Richtung mit dem Split- bzw. Merge-Run überlappen.

Ausgehend von diesen Strokes als Textons der niedrigsten Hierarchiestufe werden dann Textons aufgrund ihrer Ähnlichkeit und ihrer Nachbarschaftsbeziehung im Rahmen einer Texturhierarchie (Gleichung 3-13) zu Texturen zusammengefaßt. Abbildung 4-1 verdeutlicht, wie mehrere aus einzelnen Strokes gebildete Texturen entstehen, und jede dieser Texturen dann auf der nächsthöheren Hierarchiestufe ein Texton „Buchstabe“ darstellt. Entsprechend werden aus Buchstaben-Texturen Wörter, aus Wort-Texturen Zeilen und aus Zeilen-Texturen Blöcke. Dabei kommt man in der Texturhierarchie jeweils zur nächsthöheren Stufe, so daß letztlich gilt:  $s \succ c \succ w \succ l \succ b$ .

Auch für Blöcke kann man sich vorstellen, daß sie noch einmal zu einer Textur zusammengefaßt werden, und diese Textur als Texton aufgefaßt ein „Superblock“ wäre. Diese Vorstellung wird für die in Kapitel 4.2 dargestellte Resegmentierung nützlich sein.

Ausschlaggebend für die korrekte Funktionsweise dieses Segmentierungsprozesses ist die Wahl der verwendeten Ähnlichkeits- und Nachbarschaftsrelationen  $S$  und  $N$ . Tabelle 4-1 führt für die einzelnen Segmentierungsschritte die Grundlagen von Ähnlichkeits- und Nachbarschaftsrelationen auf, wie sie in der aktuellen Implementierung des Verfahrens gewählt sind.

---

<sup>3</sup> 8-Zusammenhang: Ein Pixel (rechteckig!) hängt mit einem anderen zusammen, wenn die beiden Pixel eine gemeinsame Kante oder Ecke haben. Im Gegensatz dazu steht der 4-Zusammenhang: Ein Pixel hängt mit einem anderen zusammen, wenn die beiden Pixel eine Kante gemeinsam haben.

<sup>4</sup> Run: Man bezeichnet als Run eine zusammenhängende Folge von schwarzen oder weißen Pixeln innerhalb einer Bildzeile.

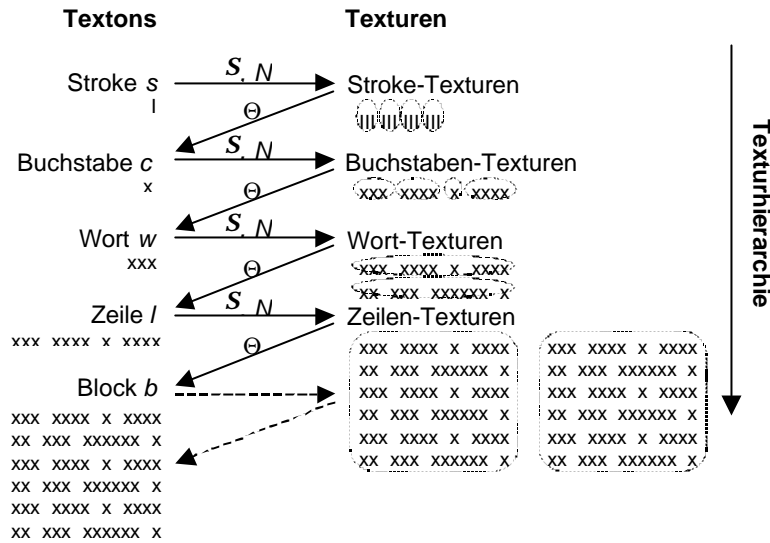


Abbildung 4-1: Texturhierarchie für die Segmentierung eines Dokuments

| Segmentierungsschritt | $Prop(t_n)$  | Grundlage von $N_n$  |
|-----------------------|--|--|
| $t_n \rightarrow t_h$ |  |  |
| $s \rightarrow c$     | Strokebreite,<br>Strokehöhe,<br>Strokeorientierung | 8-Zusammenhang zwischen Strokes  |
| $c \rightarrow w$     | Buchstabenhöhe                                     | Horizontaler Abstand,<br>Vertikale Überlappung nach rechts                               |
| $w \rightarrow l$     | Worthöhe,<br>Wortorientierung                      | Horizontaler Abstand,<br>Vertikale Überlappung nach rechts                               |
| $l \rightarrow b$     | Mittlere Buchstabenhöhe der Zeile                  | Vertikaler Abstand,<br>Horizontale Überlappung,<br>Ausrichtung der Zeilen am linken Rand |

Tabelle 4-1: Grundlagen von Ähnlichkeits- und Nachbarschaftsrelationen für die Segmentierung eines Dokuments

Der komplizierteste Segmentierungsschritt ist dabei die Bildung von Zeilen-Texturen, bzw. Block-Textons. Für ein Dokument gibt es oft eine Vielzahl von möglichen Blockanordnungen, da die Zeilen eines Dokuments durch mehrere Textonmerkmale gruppiert werden. Je nachdem, welches Merkmal man in den Vordergrund der Analyse stellt, ergibt sich eine andere Segmentierung. Das grundlegende Merkmal ist aber sicher die Zeilenhöhe, die Grundlage der Nachbarschaftsrelation der vertikale Abstand zweier benachbarter Zeilen, ihre Überlappung und gegenseitige Ausrichtung.

Abbildung 4-2 gibt für ein Beispiel die Ergebnisse der einzelnen Segmentierungsschritte wieder. Die gute Segmentierung zeigt, daß die Ähnlichkeits- und die Nachbarschaftsrelationen gut an das zu bearbeitende Dokument angepaßt sind, obwohl das verwendete Verfahren ein Dokument aus Zeitgründen in nur einem Durchgang pro Hierarchiestufe segmentiert, und eine Anpassung der Relationen deshalb nur in sehr begrenztem Umfang möglich ist. Aus Erfahrung weiß man allerdings, daß diese relativ starren Relationen nicht in allen Fällen zu einer optimalen Segmentierung führen können.

| Segmentierungsschritt | Ergebnis   |
|-----------------------|--|
| s → c                 | <p><i>Pieni kierros eri alojen ihmisten parissa paljastaa, että journalismissa vallitsevat lait ja säännöt ovat monimuutoksen kehityksen tulosta; alan toimintatavat eivät selity pelkästään naismies-akselilla.</i></p> <p>Tiedottaja <b>Outi Pärnänen</b> Suomen Punaisesta Rististä sanoo aiheen olleen mielessä erityisesti Jugoslavian sodan takia. "Kului vuosi puolitoista ennen kuin kirjoittelu alkoi olla muutenkin kuin sotatapahtumien raportointia. Vasta väkivallan avauduttua alkoivat paljastua siviilisiin kohdistuneet kauhuudet."</p> <p>Kuitenkin esimerkiksi musliminaisten raiskaukset unohtuvat otsikoista varsin pian. "Eletään uutisen armoilla ja uutiseen tarvitaan aina jotain uutta. Tavallisten ihmisten elämä ei ole uutisarvoista."</p> <p>Viime aikana ovat Ruandan tapahtumat olleet kauhisteltavana. Pärnäsen arvion mukaan melu ei olisi ollut niin valtaisa ellei siellä olisi ollut eurooppalaisia. "Meidän tiedonvälityskulttuurissamme aiheen pitää jollain lailla liittyä Eurooppaan ja eurooppalaisiin ennen kuin se on raportointisarvoista."</p> <p>Toistuminen vähentää nimikään uutisarvoa. Bangladeshin kaksi kertaa vuodessa toistuvat tulvat ovat nykyisin rutimajuttu. Luonnonkatastrofin mielenkiinto punnitetaan sen mukaan kuinka montaa kymmentä tuhatta ihmistä se vahingoittaa.</p> <p>Pärnänen ei usko, että yksi nainen tekisi kesän, tarvitaan suurempi joukko, porukka, muuttamaan valtavirtaa. Nainen päällikkönäkään ei yhdessä yössä muuta mitään, tämä sukupolvi on sisäistänyt nykyiset uutiskriteerit. "Tämä on kasvatus- ja kulttuurisidonnan juttu".</p> |
| c → w                 | <p><i>Pieni kierros eri alojen ihmisten parissa paljastaa, että journalismissa vallitsevat lait ja säännöt ovat monimuutoksen kehityksen tulosta; alan toimintatavat eivät selity pelkästään naismies-akselilla.</i></p> <p>Tiedottaja <b>Outi Pärnänen</b> Suomen Punaisesta Rististä sanoo aiheen olleen mielessä erityisesti Jugoslavian sodan takia. "Kului vuosi puolitoista ennen kuin kirjoittelu alkoi olla muutenkin kuin sotatapahtumien raportointia. Vasta väkivallan avauduttua alkoivat paljastua siviilisiin kohdistuneet kauhuudet."</p> <p>Kuitenkin esimerkiksi musliminaisten raiskaukset unohtuvat otsikoista varsin pian. "Eletään uutisen armoilla ja uutiseen tarvitaan aina jotain uutta. Tavallisten ihmisten elämä ei ole uutisarvoista."</p> <p>Viime aikana ovat Ruandan tapahtumat olleet kauhisteltavana. Pärnäsen arvion mukaan melu ei olisi ollut niin valtaisa ellei siellä olisi ollut eurooppalaisia. "Meidän tiedonvälityskulttuurissamme aiheen pitää jollain lailla liittyä Eurooppaan ja eurooppalaisiin ennen kuin se on raportointisarvoista."</p> <p>Toistuminen vähentää nimikään uutisarvoa. Bangladeshin kaksi kertaa vuodessa toistuvat tulvat ovat nykyisin rutimajuttu. Luonnonkatastrofin mielenkiinto punnitetaan sen mukaan kuinka montaa kymmentä tuhatta ihmistä se vahingoittaa.</p> <p>Pärnänen ei usko, että yksi nainen tekisi kesän, tarvitaan suurempi joukko, porukka, muuttamaan valtavirtaa. Nainen päällikkönäkään ei yhdessä yössä muuta mitään, tämä sukupolvi on sisäistänyt nykyiset uutiskriteerit. "Tämä on kasvatus- ja kulttuurisidonnan juttu".</p> |
| w → l                 | <p><i>Pieni kierros eri alojen ihmisten parissa paljastaa, että journalismissa vallitsevat lait ja säännöt ovat monimuutoksen kehityksen tulosta; alan toimintatavat eivät selity pelkästään naismies-akselilla.</i></p> <p>Tiedottaja <b>Outi Pärnänen</b> Suomen Punaisesta Rististä sanoo aiheen olleen mielessä erityisesti Jugoslavian sodan takia. "Kului vuosi puolitoista ennen kuin kirjoittelu alkoi olla muutenkin kuin sotatapahtumien raportointia. Vasta väkivallan avauduttua alkoivat paljastua siviilisiin kohdistuneet kauhuudet."</p> <p>Kuitenkin esimerkiksi musliminaisten raiskaukset unohtuvat otsikoista varsin pian. "Eletään uutisen armoilla ja uutiseen tarvitaan aina jotain uutta. Tavallisten ihmisten elämä ei ole uutisarvoista."</p> <p>Viime aikana ovat Ruandan tapahtumat olleet kauhisteltavana. Pärnäsen arvion mukaan melu ei olisi ollut niin valtaisa ellei siellä olisi ollut eurooppalaisia. "Meidän tiedonvälityskulttuurissamme aiheen pitää jollain lailla liittyä Eurooppaan ja eurooppalaisiin ennen kuin se on raportointisarvoista."</p> <p>Toistuminen vähentää nimikään uutisarvoa. Bangladeshin kaksi kertaa vuodessa toistuvat tulvat ovat nykyisin rutimajuttu. Luonnonkatastrofin mielenkiinto punnitetaan sen mukaan kuinka montaa kymmentä tuhatta ihmistä se vahingoittaa.</p> <p>Pärnänen ei usko, että yksi nainen tekisi kesän, tarvitaan suurempi joukko, porukka, muuttamaan valtavirtaa. Nainen päällikkönäkään ei yhdessä yössä muuta mitään, tämä sukupolvi on sisäistänyt nykyiset uutiskriteerit. "Tämä on kasvatus- ja kulttuurisidonnan juttu".</p> |

Abbildung 4-2: Ergebnisse der einzelnen Segmentierungsschritte an einem Beispiel (Ausschnitt aus Dokument I3), siehe Kapitel 6.3)

|                   |   |   |   |
|-------------------|---|---|---|
| $l \rightarrow b$ | <p><i>Pieni kierros eri alojen ihmisten parissa paljastaa, että journalismissa vallitsevat lait ja säännöt ovat monimutkaisen kehityksen tulosta; alan toimintatavat eivät selity pelkästään nais-mies-akselilla.</i></p> | <p>Tiedottaja <b>Outi Pärnänen</b> Suomen Punaisesta Rististä sanoo aiheen olleen mielessä erityisesti Jugoslavian sodan takia. "Kului vuosi puolitoista ennen kuin kirjoittelu alkoi olla muutakin kuin sotatapahtumien raportointia. Vasta vankileirien avauduttua alkoivat paljastua siviileihin kohdistuneet kaibeudet."</p> <p>Kuitenkin esimerkiksi musliminaisten raiskaukset unohtuivat otsikoista varsin pian. "Eletään uutisen armoilla ja uutiseen tarvitaan aina jotain uutta. Tavallisten ihmisten elämä ei ole uutisarvoista."</p> <p>Vuime aikana ovat Ruandan tapahtumat olleet kauhisteltavina. Pärnäsen arvion mukaan melu ei olisi ollut niin valtaisa ellei siellä olisi ollut eurooppalaisia. "Meidän tiedonvälityskulttuurissamme aiheen pitää jollain lailla liittyä Eurooppaan ja eurooppa-</p> | <p>laisin ennen kuin se on raportoit-<br/>sen arvoista."</p> <p>Toistuminen vähentää nimikään uutisarvoa. Bangladeshin kaksi kertaa vuodessa toistuvat tulvat ovat nykyisin rutiinjuttu. Luonnonkatastrofin mielenkiinto punnitaan sen mukaan kuinka montaa kymmentä tuhatta ihmistä se vahingoittaa.</p> <p>Pärnänen ei usko, että yksi nainen tekisi kesän, tarvitaan suurempi joukko, porukka, muuttamaan valtavirtaa. Nainen päällikkönäkään ei yhdessä yössä muuta mitään, tämä sukupolvi on sisäistänyt nykyiset uutiskriteerit. "Tämä on kasvatus- ja kulttuurisidonnainen juttu".</p> |
|-------------------|---|---|---|

Fortsetzung zu Abbildung 4-2: Ergebnisse der einzelnen Segmentierungsschritte an einem Beispiel (Ausschnitt aus Dokument I3), siehe Kapitel 6.3)

Im folgenden bezeichnet *STROKES* die Menge aller Stroke-Textons, *CHARS* die Menge aller Buchstaben-Textons, *WORDS* die Menge aller Wort-Textons, *LINES* die Menge aller Zeilen-Textons und *BLOCKS* die Menge aller Block-Textons in einem Dokument.

#### 4.1.2 Text-Graphik-Trennung

Im Verlauf des oben dargestellten Segmentierungsprozesses werden Textons zum einen aufgrund ihrer Nachbarschaftsbeziehung, zum anderen aufgrund ähnlicher Merkmale zu Texturen zusammengefaßt. Dabei sind ausschlaggebende Merkmale die Abmessungen der Textons, insbesondere ihre Höhe.

Diejenigen Textons  $t$ , die sich in ihren Abmessungen von den anderen Textons deutlich unterscheiden, können auf der jeweiligen Hierarchiestufe nicht mit Stroke-, Buchstaben-, Wort- oder Zeilen-Texturen zusammenhängen. Ihre Träger  $Dom(t)$  werden als Graphikobjekte betrachtet, und sie bekommen das zusätzliche Merkmal, ein Graphik-Texton zu sein,  $Graphic(t)$  (siehe Tabelle 4-2).

Laut Tabelle 4-2 wird dabei ein Block zu einem Graphikblock, wenn er besonders niedrig und gleichzeitig besonders breit ist. Aber auch auf andere Art und Weise können Graphikblöcke entstehen. Die Graphik-Textons auf einem Dokument können nämlich unabhängig von der Hierarchiestufe ihrer Entstehung aufgrund ihres ähnlichen Merkmals  $Graphic(t)$  und einer benachbarten Position, also aufgrund zweier Relationen  $S$  und  $N$ , zu Graphik-Texturen zusammengefaßt werden. Diese Graphik-Texturen sind wiederum Graphikblöcke.

|            |                                       |
|------------|---------------------------------------|
| Texton $t$ | $Graphic(t) :\Leftrightarrow$         |
| $s$        | $Height(s) > 140 \vee Width(t) > 250$ |
| $c$        | $Height(c) < 3$                       |
| $l$        | $Height(l) > 140$                     |
| $b$        | $Height(b) < 5 \wedge Width(b) > 40$  |

Tabelle 4-2: Definition des Texton-Merkmals  $Graphic(t)$  bei der Text-Graphik-Trennung

Außerdem wird für Buchstaben ein sehr schwaches Formkriterium als Merkmal verwendet, denn Buchstaben, die aus mehr als einer gewissen Anzahl von Strokes bestehen, werden als Graphik-Textons betrachtet.

Im Gegensatz zur Segmentierung von reinem Text, bei der fast ausschließlich implizite Klassen und entsprechende Ähnlichkeitsrelationen  $S_R$  verwendet werden, benutzt man zur Text-Graphik-Trennung explizite Klassen und die entsprechenden Ähnlichkeitsrelationen  $S_K$ . Tabelle 4-2 enthält also genau die Definitionen der expliziten Klassen für jede Hierarchiestufe. Die angegebenen Schwellen zur Trennung der Klassen sind dabei Extremwerte, bei deren Erreichen eine Ähnlichkeit mit Text sicher nicht mehr gegeben ist.

#### 4.1.3 Thematisch verwandte Arbeiten

1. Eine Übersicht über die Bestimmung von „geometrischem“ und „logischem“ Layout gibt Haralick in [Har94]. Die Bestimmung des geometrischen Layouts entspricht dem hier Segmentierungsprozeß genannten Vorgang.

- Haralick beschreibt dabei zunächst eine Reihe von Bottom-Up-Verfahren<sup>5</sup> zur Bestimmung von Blöcken und Zeilen, die sich im wesentlichen aus den Schritten
  1. Unterabtastung, englisch „Subsampling“, d.h. Tiefpaß-Filterung,
  2. Anwendung morphologischer Operationen und
  3. Connected-Components-Analyse
 zusammensetzen. Dabei gibt es sowohl Verfahren, bei denen das einleitende Subsampling fehlt, als auch Verfahren, bei denen die morphologischen Operationen entfallen.

Sieht man davon ab, daß diese Verfahren nie alle in der obigen Texturhierarchie auftretenden Objekte erzeugen, kann man durchaus Ähnlichkeiten im Vorgehen erkennen. So entspricht das „Verschmieren“ des Dokumentbildes durch Subsampling dem Auswählen der gewünschten Hierarchiestufe. In den morphologischen Operationen ist eine Ähnlichkeitsrelation  $S$  implizit durch die Form und Verknüpfungsart des Strukturelements gegeben, die Ausdehnung des Strukturelements und die Connected-Components-Analyse beinhaltet implizit eine Nachbarschaftsrelation  $N$ .

- Ein weiteres der beschriebenen Bottom-Up-Verfahren faßt Buchstabenketten zu Textgruppen zusammen, indem die Regularität einer Höhen-Abstandsbeziehung untersucht wird. Dies entspricht implizit der Anwendung einer Nachbarschaftsrelation  $N$ , die auf einem von der Buchstabenhöhe abgeleiteten Abstand beruht.
- Haralick stellt außerdem noch zwei Top-Down-Verfahren dar, die Blöcke durch Analyse der weißen Bereiche auf dem Dokument finden. Es werden also nicht wie in dieser Arbeit Texturen auf den schwarzen Pixeln des Dokuments gebildet, sondern Texturen auf den weißen Pixeln des Bildes. Die durchgeführten morphologischen Operationen realisieren in diesem Fall eine Ähnlichkeits- und eine Nachbarschaftsrelation  $S$  und  $N$  auf den Weißbereichen.

Eine überarbeitete Version des Überblicks von Haralick ist Kapitel 2 in [Dor96].

---

<sup>5</sup> Bottom-Up-Verfahren, Top-Down-Verfahren: In einem Bottom-Up-Verfahren werden von den Pixeln ausgehend immer größerer Bereiche des Dokuments zusammengefaßt. Im Gegensatz dazu wird ein Dokument in einem Top-Down-Verfahren zuerst in große Teilbereiche zerlegt, die dann in immer kleinere Bereiche unterteilt werden.

2. Auf die Segmentierung durch Texturanalyse wurde in dieser Arbeit bereits in Kapitel 2.2.3 im Detail eingegangen.

## 4.2 Resegmentierung

### 4.2.1 Übersicht

Die im vorhergehenden Kapitel 4.1 dargestellte Segmentierung erzeugt in der Texturhierarchie unter anderem Wort-, Textblock- und Graphikblock-Textons. Gerade diese Textons werden im Codierungsprozeß (Stream-Modell II) durch Layoutmerkmale betont. Die Layoutmerkmale sind Merkmale dieser Textons, die die umliegenden Textons nicht haben. Die markierten Textons bilden deshalb jeweils eigene Texturen, die die Aufmerksamkeit anziehen.

Im folgenden wird beschrieben, wie diejenigen Textons, die eines oder mehrere Layoutmerkmale tragen, durch eine Resegmentierung der bestehenden Segmentierung identifiziert werden, d.h. wie implizit durch Resegmentierung die Lokalisierung relevanter Information realisiert wird.

Eine derartige Resegmentierung ist leicht vorstellbar für Zeilen und Textblöcke, aus denen durch die Resegmentierung Wörter herausgetrennt werden. Aber auch das Finden von markierten Blöcken ist durch Resegmentierung darstellbar. Geht man nämlich auf die oben entwickelte Vorstellung des „Superblocks“ zurück, so kann man aus diesem „Superblock“, der alle Blöcke eines Dokuments enthält, die markierten Blöcke heraustrennen.

Wie in Kapitel 3.3 erläutert, ist die Grundlage einer Resegmentierung die Verfeinerung der bereits zur Segmentierung verwendeten Ähnlichkeits- und Nachbarschaftsrelationen  $S$  und  $N$ . Verfeinert man die bestehenden Ähnlichkeitsrelationen  $S$ , indem man als neue Merkmale der Textons die Layoutmerkmale hinzunimmt, so werden bei der Resegmentierung genau die Textons mit Layoutmerkmalen als eigene Texturen herausgetrennt.

Exakt die gleichen Textons werden herausgetrennt, wenn man als Grundlage der Resegmentierung die Ähnlichkeits- und die Nachbarschaftsrelation  $S$  und  $N$  verfeinert wie in Tabelle 4-3 angegeben, denn Tabelle 4-3 enthält eine übersichtliche Definition der Layoutmerkmale, die in den folgenden Kapiteln detailliert ausgeführt wird.

Hervorzuheben ist, daß jedes Layoutmerkmal im Merkmalsraum Klassen bildet. Dabei entstehen durch die Layoutmerkmale Einrückung, Bullets und durch die Fontstile Fettdruck, Kursivdruck und Großschreibung jeweils zwei Klassen, z.B. „Fettdruck“ und „Nicht-Fettdruck“. Alle anderen Layoutmerkmale bilden jeweils eine ganze Reihe von Klassen. Man betrachtet bei diesen Layoutmerkmalen im Rahmen der Meßgenauigkeit und des Meßwertebereichs nämlich jeden Meßwert der Merkmalsausprägung als eigene Klasse. Dadurch entstehen in der Regel mehr als zwei, aber trotzdem endlich viele explizite Klassen.

Letzteres ist auch dafür verantwortlich, daß der „Superblock“ (Kapitel 4.1.1) bei der Resegmentierung normalerweise in genau so viele Texturen zerfällt, wie er Blöcke enthält. So hat z.B. jeder Block im Rahmen der Meßgenauigkeit einen anderen Meßwert für die Freistellung, d.h. jeder Block wird einer anderen Freistellungs-klasse zugeordnet.



Bei der Resegmentierung werden also die einzelnen Wort-, Textblock- oder Graphikblock-Textons einer der Klassen zugeordnet, die durch die Layoutmerkmale gebildet werden, d.h. die Textons werden bezüglich der Layoutmerkmale klassifiziert.

Für die Klassifikationen der Wort-Textons in zwei Klassen braucht man jeweils Entscheidungsschwellen, die den großflächigen Vergleich der Textons bei der Texturwahrnehmung dadurch nachbilden, daß sie die globale Ausprägung von  $Prop(t_h)$  und der Grundlage von  $N_h$  (siehe Tabelle 4-3) in der Umgebung des zu klassifizierenden Objekts mit berücksichtigen. Dies wird entweder

- durch adaptive Schwellen realisiert, die sich an die Ausprägung des Merkmals auf dem Dokument anpassen, oder
- durch empirisch festgelegte Schwellen, die durch eine vorhergehende Analyse mehrerer Dokumente ermittelt wurden.

Zur robusten Bestimmung von adaptiven Entscheidungsschwellen oder Merkmalen, werden dabei zum Teil Informationen benutzt, die sich bei der Segmentierung erst auf höherer Hierarchiestufe ergeben.

| Layoutmerkmal         | $Dom(t_h)$ | $Prop(t_h)$                       | Grundlage von $N_h$                                       | $Dom(t_n \succ t_h)$ | $Prop(t_n)$    |
|-----------------------|------------|-----------------------------------|---|----------------------|----------------|
| ... für Wörter        |            |                                   |   |                      |                |
| Einrückung            | Wort $w$   | -                                 | Versatz von übereinanderstehenden Wörtern am Zeilenanfang | -                    | -              |
| Bullets               | Wort $w$   | Schwärzungsgrad, Aspektverhältnis | Abstände zu den umliegenden Wörtern                       | Pixel $p$            | Farbe          |
| Fontstil Fettdruck    | Wort $w$   | Schwärzungsgrad                   | -   | Pixel $p$            | Farbe          |
| Fontstil Kursivdruck  | Wort $w$   | Mittlere Schräglage               | -   | Stroke $s$           | Schräglage     |
| FontstilGroßschreibg. | Wort $w$   | Höhe aller Zeichen                | -   | Buchstabe $c$        | Buchstabenhöhe |
| ... für Textblöcke    |            |                                   |   |                      |                |
| Freistellung          | Block $b$  | -                                 | Abstände zu den umliegenden Blöcken                       | -                    | -              |
| Schriftgröße          | Block $b$  | Mittlere Zeilenhöhe               | -   | Zeile $l$            | Zeilenhöhe     |
| ... für Graphikblöcke |            |                                   |   |                      |                |
| Freistellung          | Block $b$  | -                                 | Abstände zu den umliegenden Blöcken                       | -                    | -              |
| Flächenmaß            | Block $b$  | Fläche                            | -   | -                    | -              |
| Schwärzungsgrad       | Block $b$  | Schwärzungsgrad                   | -   | Pixel $p$            | Farbe          |

Tabelle 4-3: Definition von Layoutmerkmalen für Textons  $t_h$  über verfeinerte Relationen: Verfeinerung von  $S$  mit Hilfe von  $Prop(t_h)$  und von  $N$  mit Hilfe der angegebenen Grundlage von  $N_h$ ; Berechnung von  $Prop(t_h)$  aus den Merkmalen der Textons  $t_n$ , die zur Konstruktion von  $t_h$  beigetragen haben ( $t_n \succ t_h$ ).

In den folgenden Kapiteln wird im Detail beschrieben, wie die einzelnen Layoutmerkmale definiert sind, und damit die jeweiligen Textons klassifiziert werden. Die Darstellung der Definitionen benutzt dabei einige typographische Bezeichnungen, die an dieser Stelle eingeführt werden sollen (Abbildung 4-3).

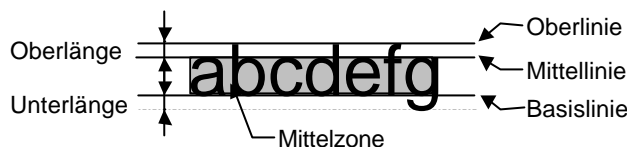


Abbildung 4-3: Typographische Bezeichnungen für die Definition von Layoutmerkmalen

Die *Basislinie* eines Wortes  $w$  ist die Ausgleichsgerade<sup>6</sup> durch die untersten Punkte der Buchstaben ohne Unterlänge im Wort. Ebenso ist die Basislinie einer Zeile  $l$  definiert. Die *Oberlinie* ist die Ausgleichsgerade durch die obersten Punkte der Großbuchstaben und Kleinbuchstaben mit Oberlänge. Da wir annehmen, daß das ganze Dokument schräglagenkorrigiert ist, sind die Ausgleichsgeraden gegeben durch die Werte  $Base(w, l)$  und  $Top(w, l)$ .

Um die *Mittellinie* eines Wortes  $w$  zu erhalten, bestimmt man zuerst in  $y$ -Richtung die lokalen Maxima der Komponenten. In der horizontalen Projektion dieser Punkte definiert das Maximum den  $y$ -Wert der Mittellinie  $Middle(w)$ .

Das zugrunde gelegte Koordinatensystem hat seinen Ursprung in der linken unteren Ecke des Dokuments.

#### 4.2.2 Definition von Einrückungen

Die Wörter sind im allgemeinen am linken Rand eines Textblocks ausgerichtet, d.h. ihre linken Ränder stehen exakt untereinander. Eine Einrückung am Blockanfang unterbricht diese regelmäßige Nachbarschaftsbeziehung und damit die Textur.

Ein Wort  $w_1$  am Anfang der ersten Zeile  $l_1$  eines Blockes  $b$  ( $w_1 \succ l_1 \succ b$ ,  $w_2 \succ l_2 \succ b$ ) (Abbildung 4-4), ist durch eine *Einrückung* markiert, wenn der folgende Ausdruck  $Indent(w_1) = 1$ :

$$Indent(w_1) := \begin{cases} 1, & \text{falls } Th_1 < Left(w_1) - Left(w_2) < Th_2 \\ 0, & \text{sonst} \end{cases} \quad (4-1)$$

Die Entscheidungsschwellen sind

$$\begin{aligned} Th_1 &= h \cdot (Top(l_1) - Base(l_1) + 1) \text{ und} \\ Th_2 &= k \cdot (Top(l_1) - Base(l_1) + 1), \end{aligned} \quad (4-2)$$

mit den empirischen Faktoren  $h = 0.5$  und  $k = 3.0$ .

Diese Definition muß erweitert werden, um zu verhindern, daß Wörter in zentrierten Zeilen als eingerückt klassifiziert werden. Es ergibt sich

$$Indent(w_1) := \begin{cases} 1, & \text{falls } Th_1 < Left(w_1) - Left(w_2) < Th_2 \wedge \\ & |(Right(w_n) - Right(w_m)) - (Left(w_1) - Left(w_2))| > Th_1 \\ 0, & \text{sonst} \end{cases} \quad (4-3)$$

<sup>6</sup> Ausgleichsgerade: Die Ausgleichsgerade von  $N$  Meßpunkten  $(x_i, y_i)$  ist gegeben durch  $y = a \cdot x + b$ , mit

$$a = \frac{N \cdot [xy] - [y][x]}{N \cdot [x^2] - ([x])^2} \text{ und } b = \frac{[y][x^2] - [xy][x]}{N \cdot [x^2] - ([x])^2} \text{ (siehe [Bro91], S. 788).}$$

Verwendet wird dabei die Gaußsche Summenschreibweise  $\sum_{i=1}^N x_i \stackrel{def}{=} [x]$ .

Der in der Texton Theorie geforderte großflächige Vergleich eines Textons mit den Textons seiner Umgebung ist hier durch die empirischen Faktoren realisiert, die sich aus der Betrachtung einer ganzen Reihe von Dokumenten ergaben.

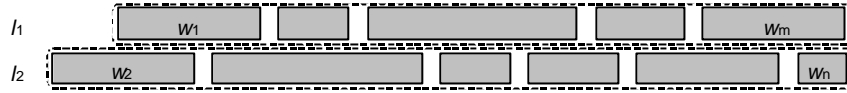


Abbildung 4-4: Bezeichnungen für die Definition von Einrückungen

### 4.2.3 Definition von Bullets

Ein Bullet fällt ins Auge, weil es sich durch seine Merkmale und seinen Abstand von den umgebenden Wörtern unterscheidet, also nicht zur Textur der umgebenden Wörter gehört.

Da ein Bullet dazu dient, das nachfolgende Wort hervorzuheben, formuliert man im folgenden die Definition bezüglich des markierten Wortes.

Mit den Bezeichnungen

1.  $WordDist(w_1) = Left(w_2) - Right(w_1)$
2.  $AspectRatio(w_1) = Width(w_1) / Height(w_1)$
3.  $BlackIntens(w_1) = |Area(w_1) \cap Dom(w_1)| / |Area(w_1)|$ ,  
mit der Fläche eines Textons  $t$ :

$$Area(t) = \{(x, y) : Bot(t) \leq y \leq Top(t) \wedge Left(t) \leq x \leq Right(t)\} \quad (4-4)$$

wird die Definition wie folgt beschrieben:

Seien zwei Wörter  $w_1$  und  $w_2$  am Anfang der Zeile  $l$  ( $w_1, w_2 \succ l$ ) (Abbildung 4-5), dann ist  $w_1$  ein *Bullet* und  $w_2$  das dadurch markierte Wort, wenn der folgende Ausdruck  $Bullet(w_2) = 1$ :

$$Bullet(w_2) := \begin{cases} 1, \text{ falls } \left\{ \begin{array}{l} c:c \succ w_1 \wedge c \in CHARS \\ Th_1 < AspectRatio(w_1) < Th_2 \wedge \\ BlackIntens(w_1) > Th_3 \wedge \\ Th_4 < WordDist(w_1) < Th_5 \end{array} \right\} = 1 \wedge \\ 0, \text{ sonst} \end{cases} \quad (4-5)$$

Die empirischen Entscheidungsschwellen sind

$$\begin{aligned} Th_1 &= 0.8, \\ Th_2 &= 1.1, \\ Th_3 &= 0.65, \end{aligned} \quad (4-6)$$

sowie

$$\begin{aligned} Th_4 &= h \cdot (Top(l) - Base(l) + 1) \\ Th_5 &= k \cdot (Top(l) - Base(l) + 1), \end{aligned} \quad (4-7)$$

mit den Faktoren  $h = 0.3$  und  $k = 1.4$ .

Die ersten drei Bedingungen der Definition beziehen sich auf Merkmale  $Prop(t_h)$ , die letzte Bedingung bezieht sich auf die Nachbarschaftsbeziehung  $N_h$ .

Auch diese Definition realisiert den großflächigen Vergleich eines Textons mit den Textons seiner Umgebung durch die empirischen Schwellen und Faktoren, die sich aus der Betrachtung einer ganzen Reihe von Dokumenten ergaben.



Abbildung 4-5: Bezeichnungen für die Definition von Bullets

## 4.2.4 Definition von Fontstilen

### 4.2.4.1 Definition von Fettdruck

Das Fettdruck charakterisierende Merkmal des Wort-Textons  $w$  ist sein Schwärzungsgrad, der durch die Anzahl der schwarzen Pixel  $p$  pro Fläche in der Mittelzone des Wortes gegeben ist.

Wie in Abbildung 4-3 gezeigt, ist die Mittelzone derjenige Bereich eines Wortes  $w$ , der von allen Buchstaben besetzt wird. Wenn man den Schwärzungsgrad nur in diesem Bereich mißt, erreicht man eine gewisse Unabhängigkeit von der Struktur der Buchstaben, d.h. von vorhandenen Unter- oder Oberlängen. Die *Mittelzone* von  $w$  ist also gegeben durch

$$MiddleZone(w) = \{(x,y): Left(w) \leq x \leq Right(w) \wedge Base(w) \leq y \leq Middle(w)\}. \quad (4-8)$$

Der *Schwärzungsgrad* eines Wortes  $w$  ist dann gegeben als

$$BlackInt(w) = |MiddleZone(w) \cap Dom(w)| / |MiddleZone(w)|. \quad (4-9)$$

Um eine robuste Klassifikation zu gewährleisten, definiert man auch den *Schwärzungsgrad für jeden Block  $b$* :

$$BlackInt(b) = \frac{\sum_{w \in \{w:w \succ b \wedge w \in WORDS\}} BlackInt(w)}{|\{w:w \succ b \wedge w \in WORDS\}|} \quad (4-10)$$

Die Entscheidungsschwelle wird abgeleitet aus dem Histogramm der Schwärzungsgrad-Werte auf dem ganzen Dokument. Dadurch adaptiert sich die Schwelle automatisch an die globale Ausprägung des Schwärzungsgrades auf dem Dokument.

Sei  $1 \leq i, j \leq N_{bin}$ , dann ist das *Histogramm des Schwärzungsgrades* gegeben durch

$$H(i) = \left| \left\{ w: w \in WORDS \wedge i = \lfloor N_{bin} \cdot BlackInt(w) \rfloor \right\} \right| \quad (4-11)$$

und die abgeleitete *Entscheidungsschwelle* durch

$$Th = (\max\{i : \forall j : H(i) \geq H(j)\} + d) / N_{bin}, \quad (4-12)$$

mit den empirisch hergeleiteten Werten  $N_{bin} = 100$  und  $d = 8$ .

Damit ist das Layoutmerkmal  $Bold(w)$  für ein Wort-Texton  $w$  in einem Block-Texton  $b$  mit  $w \succ b$

$$\mathit{Bold}(w) := \begin{cases} 1, & \text{falls } \mathit{BlackInt}(w) > Th \vee \mathit{BlackInt}(b) > Th \\ 0, & \text{sonst} \end{cases} \quad (4-13)$$

Eine Klassifikation entsprechend dieser Definition funktioniert gut für eine ausreichend große Mittelzone, d.h. für Wörter mit mehr als 3 Buchstaben. Für kürzere Wörter erweitert man die Definition, indem man den direkten Vorgänger  $\mathit{Pred}(w)$  und den direkten Nachfolger  $\mathit{Suc}(w)$  definiert:

$$\mathit{Bold}(w) := \begin{cases} 1, & \text{falls } \mathit{BlackInt}(w) > Th \wedge \\ & (\mathit{BlackInt}(\mathit{Pred}(w)) > Th \vee \mathit{BlackInt}(\mathit{Suc}(w)) > Th) \\ 0, & \text{sonst} \end{cases} \quad (4-14)$$

#### 4.2.4.2 Definition von Kursivdruck

Das Merkmal für ein kursiv gedrucktes Wort-Texton  $w$  ist die mittlere Schräglage der Stroke-Textons  $s \succ w$ . Dabei sind besonders die ungefähr senkrechten Strokes charakteristisch.

Die *Schräglage eines Strokes* ist gegeben durch den Winkel  $\mathbf{a}$  zwischen der Senkrechten und der Verbindungslinie der Mitte des obersten und der Mitte des untersten schwarzen Runs in einem Stroke (Abbildung 4-6):

$$\mathit{Slant}(s) = \tan(\mathbf{a}) = \frac{\mathit{Cent}(\mathit{TopMostRun}(s)) - \mathit{Cent}(\mathit{BottomMostRun}(s)) + 1}{\mathit{Height}(s)} \quad (4-15)$$

wobei  $\mathit{Cent}(r)$  die horizontale Mitte eines Runs  $r$  bezeichnet.

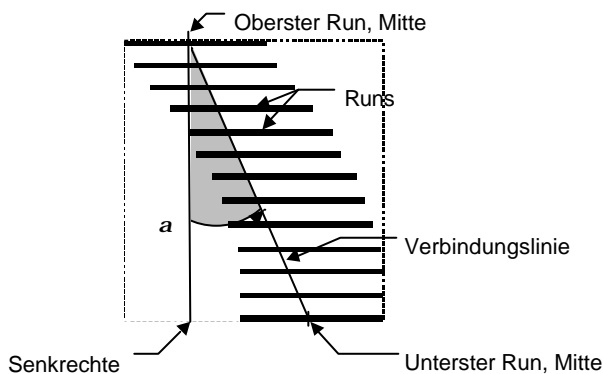


Abbildung 4-6: Zur Definition der Stroke-Schräglage  $\mathit{Slant}(s)$

Die mittlere Schräglage der Strokes innerhalb eines Wort-Textons  $w$  berechnet sich aus den Schräglagen der *ungefähr senkrechten Strokes*, die zur Konstruktion des Wortes beigetragen haben:

$$\mathit{Verticals}(w) = \{s: s \succ w \wedge s \in \mathit{STROKES} \wedge \mathit{Height}(s) > \mathit{Width}(s)\} \quad (4-16)$$

Die *mittlere Schräglage innerhalb eines Wortes*  $w$  ist gegeben durch

$$\mathit{MeanSlant}(w) = \frac{\sum_{s \in \mathit{Verticals}(w)} \mathit{Slant}(s)}{|\mathit{Verticals}(w)|} \quad (4-17)$$

und die *mittlere Schräglage innerhalb eines Blocks*  $b$  durch

$$MeanSlant(b) = \frac{\sum_{w \in \{w: w \succ b \wedge w \in WORDS\}} MeanSlant(w)}{|\{w: w \succ b \wedge w \in WORDS\}|} \quad (4-18)$$

Mit diesen Merkmalen definiert man das Layoutmerkmal  $Italics(w)$  für ein Wort-Texton  $w$  in einem Block-Texton  $b$  mit  $w \succ b$ :

$$Italics(w) := \begin{cases} 1, & \text{falls } MeanSlant(w) > Th \vee MeanSlant(b) > Th \\ 0, & \text{sonst} \end{cases} \quad (4-19)$$

Entsprechend Tabelle 6-1 (Kapitel 6.1) ist die Entscheidungsschwelle  $Th = 0.12 \approx \tan(7^\circ)$ . Die Versuche, die zur Festlegung dieser Schwelle geführt haben, sind der vorweggenommene großflächige Vergleich eines Textons mit den Textons seiner Umgebung.

Wegen der statistischen Natur des Merkmals ist die Schräglage eines Wortes nur für Wörter mit einer ausreichend großen Anzahl von ungefähr senkrechten Strokes zuverlässig. Experimente haben eine gute Klassifikationsleistung für Wörter mit mehr als drei Buchstaben ergeben.

#### 4.2.4.3 Definition von Großschreibung

Das Merkmal für ein Wort  $w$  in Großschreibung ist die Höhe der enthaltenen Buchstaben  $c \succ w$ . Zur Klassifikation eines Wortes bezüglich der Großschreibung vergleicht man die Höhe jedes einzelnen Buchstabens mit einer adaptiven Entscheidungsschwelle, die aus der mittleren Zeilenhöhe des Blocks  $b$  ( $w \succ b$ ) berechnet wird. Diese Entscheidungsschwelle realisiert damit wieder den Vergleich der lokalen Ausprägung eines Merkmals mit der Ausprägung in der weiteren Umgebung des zu klassifizierenden Wort-Textons.

Die *mittlere Zeilenhöhe eines Blocks  $b$*  (siehe Kapitel 4.2.6, Gleichung 4-32) ist gegeben durch

$$MeanHeight(b) = \frac{\sum_{l \in \{l: l \succ b \wedge l \in LINES\}} (Top(l) - Base(l) + 1)}{|\{l: l \succ b \wedge l \in LINES\}|} \quad (4-20)$$

Die davon abgeleitete Entscheidungsschwelle ist

$$Th = h \cdot MeanHeight(b) \quad (4-21)$$

mit  $h = 0.85$ . Experimente auf synthetischen Dokumenten (Kapitel 6.1) haben das Verhältnis 0.74 zwischen der Höhe der Kleinbuchstaben ohne Unter- und Oberlänge und der Höhe der Großbuchstaben ergeben. Für eine robuste Klassifikation bezüglich Großschreibung wird aber ein etwas höherer Wert gewählt.

Ein Wort  $w$  in der Zeile  $l$  eines Blocks  $b$  ( $w \succ l \succ b$ ) wird als großgeschrieben definiert, wenn  $Capitals(w) = 1$ :

$$Capitals(w) := \begin{cases} 1, & \text{falls } \forall c \succ w, c \in CHARS : TOP(c) - BASE(l) + 1 > Th \\ 0, & \text{sonst} \end{cases} \quad (4-22)$$

Um die Klassifikation robuster zu machen, definiert man zusätzlich für einen Block  $b$  mit den Buchstaben  $c$  in der Zeile  $l$  ( $w \succ l \succ b$ ), daß

$$Capitals(b) := \begin{cases} 1, & \text{falls } \frac{|\{c: c \succ b \wedge TOP(c) - BASE(l) + 1 > Th \wedge c \in CHARS\}|}{|\{c: c \succ b \wedge c \in CHARS\}|} > 0.67 \\ 0, & \text{sonst} \end{cases} \quad (4-23)$$

Die *Capitals*-Definition für Blöcke zählt im wesentlichen die großgeschriebenen Buchstaben. Wenn mehr als zwei Drittel der Buchstaben eines Blocks großgeschrieben sind, wird der ganze Block als großgeschrieben klassifiziert.

Daraus ergibt sich letztendlich eine Definition für das Layoutmerkmal *AllCapitals(w)* eines Wortes  $w \succ b$ :

$$AllCapitals(w) := \begin{cases} 1, & \text{falls } Capitals(w) \vee Capitals(b) \\ 0, & \text{sonst} \end{cases} \quad (4-24)$$

Der Klassifikator entsprechend dieser Definition behandelt erst Wörter mit mehr als fünf Buchstaben, weil für diese Wörter

- zum einen keine Zweideutigkeit besteht zwischen den ganz großgeschriebenen Wörtern und denen, die sowohl Großbuchstaben als auch Kleinbuchstaben mit Oberlänge enthalten, und
- zum anderen die Verwechslungsgefahr mit Abkürzungen gering ist.

#### 4.2.4.4 Thematisch verwandte Arbeiten

Klassifikationsverfahren für Fontstile wurden bereits mit unterschiedlichen Methoden und Zielsetzungen entworfen:

1. Bloomberg ([Blo91]) entwirft Klassifikationsverfahren für Kursivdruck und Fettdruck, weil diese Merkmale seiner Auffassung nach die wichtigen Wörter eines Dokuments kennzeichnen. Er realisiert die Verfahren wie in dieser Arbeit durch Texturanalyse, allerdings modelliert er die Texturbestandteile nicht explizit, sondern verwendet morphologische Operationen über das ganze Dokument.
2. Chaudhuri und Garain beschreiben in [Cha98] Klassifikationsverfahren für Kursivdruck, Fettdruck und Großschreibung. Diese Klassifikationsverfahren arbeiten nicht mit morphologischen Operationen, sondern ähnlich den hier präsentierten Verfahren mit expliziten Merkmalen von Texteinheiten wie Strichen oder Buchstaben. Ein Zusammenhang mit Texturanalyse wird bei Chaudhuri und Garain aber nicht hergestellt. Besonders die Merkmale zur Klassifikation von Fettdruck sind dort dazu aber auch zu wenig an der menschlichen Wahrnehmung orientiert.

Obwohl erwähnt wird, daß Fontinformation bei der Extraktion von Information aus Datenbanken hilfreich sein kann, ist das Hauptziel der Autoren, die Klassifikationsleistung von OCR-Algorithmen zu verbessern.

3. Bei Doermann ([Doe98]) ist die funktionelle Ebene eines Dokuments durch gestalterische Merkmale gekennzeichnet (siehe Kapitel 2.1.4). Deshalb werden auch in [Doe98] Klassifikationsverfahren für Kursivdruck und Fettdruck beschrieben. Ohne Bezug zur Texturanalyse berechnet das Verfahren für Kursivdruck aus Buchstaben ein explizites Merkmal für Kursivdruck, Fettdruck wird dagegen morphologisch klassifiziert.

Ein Arbeitsgebiet, das der Klassifikation bezüglich Fontstilen verwandt ist, ist die Klassifikation bezüglich Font. Während es bei der Klassifikation bezüglich der Fontstile gerade darauf ankommt, von der aktuellen Schrift (engl. „typeface“) und der aktuellen Schriftgröße zu abstrahieren, geht es bei der Erkennung von Font darum, diese Informationen zusätzlich zum Fontstil zu gewinnen. Fontklassifikation wird im wesentlichen als Vorverarbeitungsschritt für OCR verwendet, um deren Erkennungsleistung zu erhöhen.

1. Khoubyari und Hull ([Kho96]) bestimmen den vorherrschenden Font eines Textes, indem sie die Bilder der Wörter auf einem Dokument ausschneiden und zu ähnlichen Gruppen zusammenfassen. Häufige, kurze Wörter werden als sogenannte „function words“ identifiziert, die in jedem Dokument vorkommen. Für das zu untersuchende Dokument wird für jedes „function word“ durch Überlagerung ein Prototyp gebildet, der dann zur Fontklassifikation mit den in einer Datenbank abgespeicherten „function words“ in verschiedenen Fonts verglichen wird.
2. In [Zra98] beschreiben Zramdini und Ingold ein System zur Optical Font Recognition (OFR). Der Font eines Textes wird bei diesem Ansatz zeilenweise bestimmt. Dazu werden aus jeder Zeile acht Merkmale extrahiert, die auf Connected Components und horizontalen und vertikalen Projektionen beruhen. Diese Merkmale werden dann mittels Bayes-Klassifikator und einer Datenbank von Font-Modellen klassifiziert.

Ein Font ist laut [Zra98] unter anderem gekennzeichnet durch Gewicht (fünf Abstufungen zwischen Dünn- und Fettdruck) und Neigung (normal oder kursiv). Die Klassifizierung bezüglich dieser Eigenschaften hat große Ähnlichkeit mit der Klassifikation bezüglich der Fontstile Fett- und Kursivdruck durch explizite Merkmale.

3. Interessant ist auch der Ansatz von Morris ([Mor92]), Fonts mit einem spektralen Ansatz zu klassifizieren, der sich auf das visuelle System des Menschen beruft. Allerdings kann dieser Ansatz nur Fonts einer einzigen Schriftgröße erkennen, und auch nur auf synthetischen Dokumenten ohne Störungen wie sie beim Scannen oder Kopieren entstehen.
4. Shi und Pavlidis ([Shi97]) zeigen die Verknüpfung von klassifizierter Fontinformation mit einem OCR-Algorithmus, und die daraus resultierende Verbesserung der OCR-Erkennungsergebnisse. Es genügt in diesem Beispiel bereits die Fontfamilie des Textes zu bestimmen, d.h. Schrift mit Serifen von Schrift ohne Serifen und normale von kursiv gedruckter Schrift zu unterscheiden.

#### 4.2.5 Definition von Freistellung

Die einzelnen Block-Textons auf einem Dokument bilden nur noch schwer eine gemeinsame Textur. Neben Unterschieden in Layoutmerkmalen wie der Schriftgröße bei Textblöcken (siehe Kapitel 4.2.6) und der Fläche bei Graphikblöcken (siehe Kapitel 4.2.7), ist oft auch keine regelmäßige Nachbarschaftsbeziehung mehr gegeben. So haben die einzelnen Blöcke eines Dokuments oft sehr unterschiedliche Abstände zueinander, so daß der Eindruck von



verschieden großen Weißbereichen entsteht, die die Blöcke umgeben. Für diese unterschiedliche Freistellung wird nun ein Maß definiert.

Dazu definiert man zunächst einen *Rahmen* (Abbildung 4-7), der alle Blöcke  $b \in BLOCKS$  auf einem Dokument umgibt, aber den weißen Seitenrand abschneidet:

$$\begin{aligned} FrameTop &= \max\{Top(b) : b \in BLOCKS\} \\ FrameBot &= \min\{Bot(b) : b \in BLOCKS\} \\ FrameLeft &= \min\{Left(b) : b \in BLOCKS\} \\ FrameRight &= \max\{Right(b) : b \in BLOCKS\} \end{aligned} \quad (4-25)$$

Außerdem definiert man zwei Regeln für *horizontale und vertikale Überlappung* von zwei Blöcken  $b$  und  $b_1$ .

$$OverHor(b, b_1) := \begin{cases} 1, & \text{falls } \max\{Left(b), Left(b_1)\} - \min\{Right(b), Right(b_1)\} < 0 \\ 0, & \text{sonst} \end{cases} \quad (4-26)$$

$$OverVert(b, b_1) := \begin{cases} 1, & \text{falls } \max\{Bot(b), Bot(b_1)\} - \min\{Top(b), Top(b_1)\} < 0 \\ 0, & \text{sonst} \end{cases}$$

Damit sind für einen Block  $b_1 \in BLOCKS$  die Abstände zu den nach oben, unten, links und rechts nächstgelegenen Blöcken folgendermaßen gegeben:

$$\begin{aligned} \text{Wenn } \exists b: Bot(b) - Top(b_1) > 0 \wedge OverHor(b, b_1) \wedge b \in BLOCKS, \\ \text{dann } BlockDistTop(b_1) &= \\ & \min\{Bot(b) - Top(b_1) : Bot(b) - Top(b_1) > 0 \wedge \\ & \quad OverHor(b, b_1) \wedge b \in BLOCKS\} \\ \text{sonst } BlockDistTop(b_1) &= FrameTop - Top(b_1) \end{aligned} \quad (4-27)$$

$$\begin{aligned} \text{Wenn } \exists b: Top(b) - Bot(b_1) < 0 \wedge OverHor(b, b_1) \wedge b \in BLOCKS, \\ \text{dann } BlockDistBot(b_1) &= \\ & \min\{Top(b) - Bot(b_1) : Top(b) - Bot(b_1) < 0 \wedge \\ & \quad OverHor(b, b_1) \wedge b \in BLOCKS\} \\ \text{sonst } BlockDistBot(b_1) &= Bot(b_1) - FrameBot \end{aligned} \quad (4-28)$$

$$\begin{aligned} \text{Wenn } \exists b: Right(b) - Left(b_1) < 0 \wedge OverVert(b, b_1) \wedge b \in BLOCKS, \\ \text{dann } BlockDistLeft(b_1) &= \\ & \min\{Right(b) - Left(b_1) : Right(b) - Left(b_1) < 0 \wedge \\ & \quad OverVert(b, b_1) \wedge b \in BLOCKS\} \\ \text{sonst } BlockDistLeft(b_1) &= Left(b_1) - FrameLeft \end{aligned} \quad (4-29)$$

$$\begin{aligned} \text{Wenn } \exists b: Left(b) - Right(b_1) > 0 \wedge OverVert(b, b_1) \wedge b \in BLOCKS, \\ \text{dann } BlockDistRight(b_1) &= \\ & \min\{Left(b) - Right(b_1) : Left(b) - Right(b_1) > 0 \wedge \\ & \quad OverVert(b, b_1) \wedge b \in BLOCKS\} \\ \text{sonst } BlockDistRight(b_1) &= FrameRight - Right(b_1) \end{aligned} \quad (4-30)$$

Das Maß für die *Freistellung* eines Blockes  $b_1 \in BLOCKS$  berechnet sich als Summe der einzelnen Abstände

$$Free(b_1) := BlockDistTop(b_1) + BlockDistBot(b_1) + BlockDistLeft(b_1) + BlockDistRight(b_1) \quad (4-31)$$

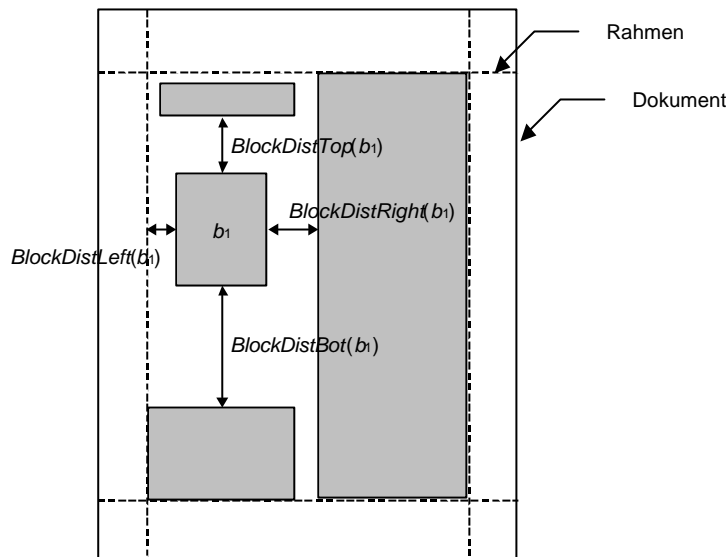


Abbildung 4-7: Bezeichnungen für die Definition von Freistellung

#### 4.2.6 Definition von Schriftgröße

Wie bereits erwähnt, bilden Blöcke auf einem Dokument kaum mehr Texturen. Ein Grund dafür sind Unterschiede in der Schriftgröße der einzelnen Blöcke. Ein Block-Texton, dessen Schriftgröße sich von der Schriftgröße der umliegenden Blöcke unterscheidet, hat damit nämlich bereits ein Merkmal, das eine gemeinsame Textur mit den umliegenden Blöcken verhindert.

Ein Maß für die Schriftgröße eines Blocks ist die mittlere Zeilenhöhe im Block. Dieses neue Merkmal des Block-Textons ergibt sich aus der Höhe der einzelnen Zeilen-Textons  $l \triangleright b$ , die den Block bilden. Die *mittlere Zeilenhöhe eines Blocks*  $b \in BLOCKS$  ist also wie folgt definiert:

$$MeanHeight(b) := \frac{\sum_{l \in \{l: l \triangleright b \wedge l \in LINES\}} (Top(l) - Base(l) + 1)}{|\{l: l \triangleright b \wedge l \in LINES\}|} \quad (4-32)$$

#### 4.2.7 Definition eines Flächenmaßes

Graphikblöcke bilden eine von Textblöcken prinzipiell verschiedene Textur auf dem Dokument. Untereinander unterscheiden sich die Graphikblöcke unter anderem durch ihre Fläche. Dieses Merkmal gibt es bei Textblöcken nicht, weil die Größe eines Blockes dort mit der Länge eines Sinnabschnittes im Text gekoppelt ist, d.h. bei der Gestaltung eines Dokuments nicht frei wählbar ist.

Ein Maß für die Fläche eines Blockes  $b \in BLOCKS$  ist

$$AreaMeas(b) := Width(b) + Height(b) \quad (4-33)$$

Dieses Maß hat eine geringere Dynamik als  $Width(b) \cdot Height(b)$  und wurde deshalb im Hinblick auf die zu berechnende Attraktivitätsbewertung gewählt.

#### 4.2.8 Definition von Schwärzungsgrad

Graphikblöcke unterscheiden sich untereinander außer durch ihre Fläche auch durch ihren Schwärzungsgrad. Der *Schwärzungsgrad eines Graphikblocks*  $b$  differenziert zwischen Rasterbildern und den weniger attraktiven Liniengraphiken. Er berechnet sich entsprechend Gleichung 4-4:

$$BlackIntens(b) := |Area(b) \cap Dom(b)| / |Area(b)| \quad (4-34)$$

### 4.3 Attraktivitätsbewertung

#### 4.3.1 Übersicht

Die beiden vorhergehenden Kapitel 4.1 und 4.2 haben gezeigt, wie visuell attraktive Objekte als Textons auf einem Dokument durch texturbasierte Segmentierung und Resegmentierung lokalisiert werden können. Nach der Grundidee der Arbeit (Stream-Modell II, Abbildung 2-2) ist damit implizit auch eine Lokalisierung relevanter Information realisiert. Auf diesem Ergebnis aufbauend kann nun in einem nächsten Schritt eine Bewertung für die visuelle Attraktivität der einzelnen Objekte und damit implizit wiederum für die Relevanz von Information berechnet werden. Die Bewertung soll also eine Aussage darüber treffen, wie attraktiv ein Objekt im Vergleich zu den anderen Objekten auf dem Dokument ist, d.h. wie stark es die Aufmerksamkeit eines Betrachters auf den ersten Blick anziehen würde.

Im folgenden wird in Hinblick auf eine spätere Informationsextraktion für jedes Wort und jeden Graphikblock eine Attraktivitätsbewertung berechnet, da dies die kleinsten inhaltlich sinntragenden Objekte des Dokuments sind. Diese Attraktivitätsbewertung wird aus den Layoutmerkmalen der jeweiligen Objekte hergeleitet. Deshalb kann man sie auch als mittelbares Merkmal der Wort- und Graphikblock-Textons des Dokuments betrachten. In diesem Sinn ist bereits in Kapitel 3.4 ein Bewertungsbegriff formal beschrieben worden, der nun die Attraktivitätsbewertung definiert.

Ausgangspunkt für die Berechnung der Attraktivitätsbewertung ist das Ergebnis der Resegmentierung, die jedes Texton bezüglich der Layoutmerkmale klassifiziert hat. Dabei wird die Zugehörigkeit eines Textons zu den Klassen eines Layoutmerkmals durch Zahlenwerte repräsentiert. Diese Zahlenwerte werden normiert und gewichtet und entsprechen dann Bewertungen einzelner Merkmale nach Gleichung 3-19. Um eine gesamte Attraktivitätsbewertung für ein Texton zu erhalten, werden dann die Bewertungen der einzelnen Layoutmerkmale addiert, d.h. die Attraktivitätsbewertung ist eine additive Bewertung nach Gleichung 3-20. Die folgenden Kapitel werden die Bewertung der einzelnen Layoutmerkmale und die Definition einer gesamten Attraktivitätsbewertung im einzelnen darstellen.

#### 4.3.2 Normierung

Wie bereits erwähnt wird die Zugehörigkeit eines Textons zu den Klassen eines Layoutmerkmals durch Zahlenwerte repräsentiert. Um daraus für jedes Texton und jedes Layoutmerkmal einzelne Bewertungen zu machen, die anschließend zu einer gesamten Attraktivitätsbewertung addiert werden können, müssen diese Zahlenwerte zunächst normiert werden.

Die Normierung erfolgt auf den Wertebereich  $[0;1]$ .

Die Klassifikationen derjenigen Layoutmerkmale, die nur zwei Klassen induzieren, können daher direkt übernommen werden.

$$\begin{aligned}
 \text{Indent}'(w) &= \text{Indent}(w) \\
 \text{Bullet}'(w) &= \text{Bullet}(w) \\
 \text{Bold}'(w) &= \text{Bold}(w) \\
 \text{Italics}'(w) &= \text{Italics}(w) \\
 \text{AllCapitals}'(w) &= \text{AllCapitals}(w)
 \end{aligned} \tag{4-35}$$

Für diejenigen Layoutmerkmale, die so viele Klassen induzieren, wie im Rahmen der Meßgenauigkeit Meßwerte entstehen, wird das errechnete Maß der Merkmalsausprägung bezüglich der maximalen Merkmalsausprägung auf dem Dokument normiert.

$$\begin{aligned}
 \text{Free}'(b) &= \text{Free}(b) / \max\{\text{Free}(b): b \in \text{BLOCKS}\} \\
 \text{MeanHeight}'(b) &= \text{MeanHeight}(b) / \max\{\text{MeanHeight}(b): b \in \text{BLOCKS}\} \\
 \text{AreaMeas}'(b) &= \text{AreaMeas}(b) / \max\{\text{AreaMeas}(b): b \in \text{BLOCKS}\} \\
 \text{BlackIntens}'(b) &= \text{BlackIntens}(b) / \max\{\text{BlackIntens}(b): b \in \text{BLOCKS}\}
 \end{aligned} \tag{4-36}$$

### 4.3.3 Gewichte

Um eine Attraktivitätsbewertung der einzelnen Layoutmerkmale zu erhalten, muß nun noch berücksichtigt werden, daß nicht alle Layoutmerkmale von einem Betrachter als gleich attraktiv empfunden werden. So ist beispielsweise allgemein bekannt, daß Kursivdruck die Aufmerksamkeit eines Betrachters weniger stark anzieht als Fettdruck, d.h. Kursivdruck weniger attraktiv ist als Fettdruck. Aus diesem Grund, werden die oben normierten Zahlenwerte mit Gewichten multipliziert, die die unterschiedliche Attraktivität einzelner Layoutmerkmale modellieren sollen.

Initialwerte für diese Gewichte wurden zunächst durch eine Befragung (Kapitel 5) geschätzt, da die Attraktivität der einzelnen Layoutmerkmale für jeden Betrachter subjektiv ist. Abbildung 5-4 zeigt die Ergebnisse der Befragung. Das sogenannte Ranking der Layoutmerkmale gibt die Reihenfolge an, in der die Merkmale auf den ersten Blick wahrgenommen werden, d.h. die Reihenfolge bezüglich der Attraktivität. Die Gewichte werden aus dieser Reihenfolge als Reziprokwerte der Rankingwerte abgeleitet und zum Teil empirisch angepaßt.

Aus dem Ranking von Textblöcken (2.82), Liniengraphiken (1.88) und Bildern (1.28) berechnet man

1. das Maß für die strukturbedingt höhere Attraktivität von Liniengraphiken und Bildern gegenüber von Text  $d_{\text{Graphik}}$ . Graphikblöcke zeigen nämlich im Gegensatz zu Textobjekten keine ausgeprägte Strich-, Wort- und Zeilenstruktur, sondern eine sehr unregelmäßige innere Struktur, denn die Bestandteile eines Graphikblocks sind von variabler Ausdehnung und Anordnung. Man berechnet
2. das Gewicht für die Bewertung des Schwärzungsgrades  $g_{\text{BlackIntens}}$ , der den Unterschied zwischen Liniengraphiken und Bildern beschreiben soll.

Wie Abbildung 4-8 zeigt, werden die Werte des Rankings dazu zunächst auf den Wert für Text normiert, so daß Text keine zusätzliche Gewichtung erhält. Von den normierten Werten bildet man dann die Reziprokwerte. Man kann nun ablesen, daß Graphik strukturbedingt eine um 0.5 höhere Attraktivität als Text hat, d.h.  $d_{\text{Graphik}} = 0.5$ . Die Differenz zwischen dem Wert für Liniengraphik und dem Wert für ein Bild ist das Gewicht  $g_{\text{BlackIntens}}$ , das ein Graphikobjekt zusätzlich zu  $d_{\text{Graphik}}$  um so höher gewichtet, je dichter es erscheint.

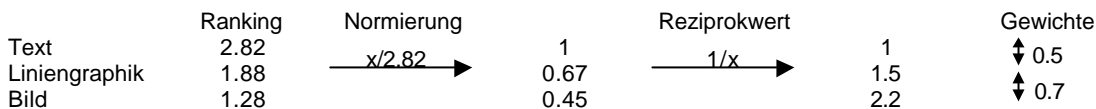


Abbildung 4-8: Zur Herleitung der Gewichte  $d_{\text{Graphik}}$  und  $g_{\text{BlackIntens}}$

Aus dem Ranking der Layoutmerkmale Freistellung (1.28), Fettdruck (3.84), Bullets (4.13), Einrückung (4.67), Großschreibung (5.36) und Kursivdruck (6.23) werden durch Bildung des Reziprokwertes die entsprechenden Gewichte.

Zwei der Gewichte werden nicht direkt aus der Befragung übernommen, nämlich  $g_{\text{MeanHeight}}$ , das Gewicht für die Schriftgröße, und  $g_{\text{AreaMeas}}$ , das Gewicht für das Flächenmaß eines Graphikblocks.

1. Das Gewicht  $g_{\text{MeanHeight}}$  ergäbe sich aus der Befragung zu  $1/2.41$ , wird aber im Laufe der Implementierung auf  $1/1.5$  erhöht, um Überschriften, die nur durch die Schriftgröße betont sind, mit einer ausreichend großen Attraktivität zu bewerten.
2. Das Gewicht  $g_{\text{AreaMeas}}$  ergäbe sich aus der Befragung zu  $1/1.72$ , wird aber auf  $1/3$  gesenkt, damit bei unterschiedlich großen Graphiken in einem Dokument die kleineren nicht unterbewertet werden.

Insgesamt ergeben sich die in Tabelle 4-4 aufgeführten Gewichte.

| Gewicht                  | Wert                  |
|--------------------------|-----------------------|
| $g_{\text{Indent}}$      | $1/4.67 \approx 0.21$ |
| $g_{\text{Bullet}}$      | $1/4.13 \approx 0.24$ |
| $g_{\text{Bold}}$        | $1/3.84 \approx 0.26$ |
| $g_{\text{Italics}}$     | $1/6.23 \approx 0.16$ |
| $g_{\text{AllCapitals}}$ | $1/5.36 \approx 0.19$ |
| $g_{\text{Free}}$        | $1/1.28 \approx 0.78$ |
| $g_{\text{MeanHeight}}$  | $1/1.50 \approx 0.67$ |
| $g_{\text{AreaMeas}}$    | $1/3.00 \approx 0.33$ |
| $g_{\text{BlackIntens}}$ | 0.7                   |
| $d_{\text{Graphik}}$     | 0.5                   |

Tabelle 4-4: Gewichte zur Bewertung der einzelnen Layoutmerkmale

### 4.3.4 Bewertungen der einzelnen Layoutmerkmale

Um eine Bewertung der einzelnen Layoutmerkmale eines Textons zu erhalten (Gleichung 3-19), werden die normierten Zahlenwerte der einzelnen Layoutmerkmale mit den entsprechenden Gewichten multipliziert. Es ergeben sich so die in Tabelle 4-5 aufgeführten Bewertungen.

| Layoutmerkmal   | Bewertung   |
|-----------------|---|
| Einrückung      | $g_{\text{Indent}} \cdot \text{Indent}'(w)$           |
| Bullet          | $g_{\text{Bullet}} \cdot \text{Bullet}'(w)$           |
| Fettdruck       | $g_{\text{Bold}} \cdot \text{Bold}'(w)$               |
| Kursivdruck     | $g_{\text{Italics}} \cdot \text{Italics}'(w)$         |
| Großschreibung  | $g_{\text{AllCapitals}} \cdot \text{AllCapitals}'(w)$ |
| Freistellung    | $g_{\text{Free}} \cdot \text{Free}'(b)$               |
| Schriftgröße    | $g_{\text{MeanHeight}} \cdot \text{MeanHeight}'(b)$   |
| Flächenmaß      | $g_{\text{AreaMeas}} \cdot \text{AreaMeas}'(b)$       |
| Schwärzungsgrad | $g_{\text{BlackIntens}} \cdot \text{BlackIntens}'(b)$ |

Tabelle 4-5: Bewertungen der einzelnen Layoutmerkmale

### 4.3.5 Attraktivitätsbewertung als additive Bewertung

Die Bewertungen der einzelnen Layoutmerkmale können dann für jedes Wort- und jedes Graphikblock-Texton zu einer gesamten Attraktivitätsbewertung zusammengefaßt werden. Dies geschieht wie bereits erläutert aufgrund neuerer elektrophysiologischer Untersuchungen ([Bac97]) in Form einer additiven Bewertung (Gleichung 3-20).

Interessant ist dabei die Behandlung der beiden Layoutmerkmale Freistellung und Schriftgröße, die Merkmale für Textblöcke, und nicht für Wörter sind. Da Text aber in Einheiten von Worten bewertet werden soll, bildet man die Bewertungen der beiden Merkmale auf die Wortebene ab und rechnet sie in die Attraktivitätsbewertung für Wörter ein. Das ist aus der Sicht der Attraktivitätsbewertung durchaus sinnvoll, da sich die Attraktivität eines Blockes auch auf die Komponenten erstrecken muß, die ihn bilden. Allerdings tragen Wörter im allgemeinen noch weitere Merkmale, die ihrerseits weitere Attraktivitätsbewertungen verursachen. Damit ist die Attraktivitätsbewertung eines Wort-Textons  $w$ , das zur Bildung eines Block-Textons  $b$  beigetragen hat ( $w \succ b$ ), normalerweise höher als die Attraktivitätsbewertung des Block-Textons  $b$ . Dies ist völlig konform zur formalen Definition der Bewertung in einer Texturhierarchie nach Gleichung 3-21, die besagt, daß Textons niedrigerer Hierarchiestufe Spezialisierungen der Textons höherer Hierarchiestufe sind und deshalb eine höhere Bewertung haben sollten.

Für ein Wort  $w \succ b$  gilt also

$$\begin{aligned} (g_{\text{Free}} \cdot \text{Free}'(w)) &= (g_{\text{Free}} \cdot \text{Free}'(b)) \\ (g_{\text{MeanHeight}} \cdot \text{MeanHeight}'(w)) &= (g_{\text{MeanHeight}} \cdot \text{MeanHeight}'(b)) \end{aligned} \quad (4-37)$$

Die Bewertungen der einzelnen Layoutmerkmale werden nun zu einer additiven Bewertung, der gesamten Attraktivitätsbewertung, verknüpft. Man definiert dabei unterschiedliche Attraktivitätsbewertungen für Wörter und Graphikblöcke, da diese Textons jeweils ganz verschiedene Layoutmerkmale tragen.

Die Attraktivitätsbewertung  $A$  eines Wortes  $w \in \text{WORDS}$  ist gegeben durch

$$\begin{aligned} A(w) &= g_{\text{Indent}} \cdot \text{Indent}'(w) + \\ &+ g_{\text{Bullet}} \cdot \text{Bullet}'(w) + \\ &+ g_{\text{Bold}} \cdot \text{Bold}'(w) + \\ &+ g_{\text{Italics}} \cdot \text{Italics}'(w) + \\ &+ g_{\text{AllCapitals}} \cdot \text{AllCapitals}'(w) + \\ &+ g_{\text{Free}} \cdot \text{Free}'(w) + \\ &+ g_{\text{MeanHeight}} \cdot \text{MeanHeight}'(w) \end{aligned} \quad (4-38)$$

Die Attraktivitätsbewertung  $A$  eines Graphikblocks  $b \in \text{BLOCKS}$  ist gegeben durch

$$\begin{aligned} A(b) &= d_{\text{Graphik}} + \\ &+ g_{\text{Free}} \cdot \text{Free}'(b) + \\ &+ g_{\text{AreaMeas}} \cdot \text{AreaMeas}'(b) + \\ &+ g_{\text{BlackIntens}} \cdot \text{BlackIntens}'(b) \end{aligned} \quad (4-39)$$

Diese Attraktivitätsbewertungen werden nun noch so auf den Wertebereich  $[0;1]$  normiert, daß das Objekt auf dem Dokument mit dem niedrigsten Attraktivitätswert die Bewertung 0

erhält, und das Objekt mit dem höchsten Attraktivitätswert die Bewertung 1. Mit den Bezeichnungen

$$\begin{aligned} AttrMax &= \max\{A(b): b \in BLOCKS, A(w): w \in WORDS\} \text{ und} \\ AttrMin &= \min\{A(b): b \in BLOCKS, A(w): w \in WORDS\} \end{aligned} \quad (4-40)$$

ergeben sich die normierten Attraktivitätsbewertungen zu

$$\begin{aligned} A'(w) &= (A(w) - AttrMin) / (AttrMax - AttrMin) \text{ und} \\ A'(b) &= (A(b) - AttrMin) / (AttrMax - AttrMin). \end{aligned} \quad (4-41)$$

Aufgrund dieser Normierung und der vorhergehenden Normierungen der Zahlenwerte, die die Klassifikationen der Textons repräsentieren, erhält man eine Bewertung, die die Attraktivität der einzelnen Objekte auf einem Dokument relativ zueinander beurteilt. Ein Vergleich von Attraktivitätsbewertungen zwischen verschiedenen Dokumenten ist also wohlgermerkt nicht möglich, allerdings auch nicht vorgesehen. Ein Dokument im automatischen Verfahren entspricht nämlich dem Blickfeld der psychologischen Texton Theorie, und nur in diesem Blickfeld läuft die „preattentive vision“ und die visuelle Aufmerksamkeit ab.

#### 4.3.6 Thematisch verwandte Arbeiten

Eglin und Emptoz stellen in [Egl97] und [Egl98] eine Methode zur Segmentierung von Dokumenten vor, die einzelne Bereiche auf einem Dokument entsprechend ihrer visuellen Auffälligkeit in unterschiedlichen Auflösungen repräsentiert. Auf Grundlage von physiologischen Erkenntnissen wird das Dokument als erstes bezogen auf einen Fixationspunkt in eine „log-polar mapping“ genannte Darstellung übertragen. Diese Darstellung wird rund um den Fixationspunkt mit Gauß-Funktionen gefiltert, wobei die räumliche Auflösung mit zunehmender Entfernung geringer wird. Danach erfolgt eine Kantendetektion und diskrete Approximation der berechneten Kanten, die unregelmäßige Blockstrukturen umgeben. In dieser Darstellung werden nun die „interessanten“ Bereiche identifiziert, indem die Fläche, die Symmetrie und Kantenkrümmung jedes Blocks berechnet wird. Auch die Textur jedes Blocks wird zu seiner Charakterisierung benutzt, allerdings nicht, um wie in dieser Arbeit Textons bezüglich der Layoutmerkmale zu klassifizieren. Vielmehr werden aus der Textur eines Blocks die allgemeineren Kennzeichen Komplexität, Kompaktheit und Lesbarkeit abgeleitet.

Auf diese Art und Weise ergibt sich ein neuer Fixationspunkt, der Ausgangspunkt für eine erneute Segmentierung ist. Ein Dokument wird also durch eine zeit- und ortsvariante Darstellung repräsentiert, die seine Segmentierung entsprechend der menschlichen Blickbewegungen wiedergibt.

Eglin und Emptoz weisen demnach den einzelnen Bereichen auf einem Dokument eine Attraktivität etwa im Sinn dieser Arbeit zu, allerdings nicht in Form eines expliziten Maßes, sondern über die Auflösung der errechneten Segmentierung. Je attraktiver ein Bereich, desto höher ist die Auflösung.





# 5 Befragung zur Wahrnehmung von Layoutmerkmalen

## 5.1 Ziel der Befragung

In Kapitel 4.3 wurde erläutert, wie diese Arbeit eine Attraktivitätsbewertung aus den Bewertungen einer Reihe von Layoutmerkmalen berechnet. Dieses Vorgehen ist von der menschlichen Wahrnehmung inspiriert. Da aber die Wahrnehmung von Layoutmerkmalen ausgesprochen subjektiv ist, ging es in der im folgenden beschriebenen Befragung darum,

- zum einen die Liste der hier vorgeschlagenen Layoutmerkmale zu verifizieren und gegebenenfalls zu ergänzen, und
- zum anderen eine initiale Schätzung für die Gewichte in den Bewertungen der einzelnen Layoutmerkmale zu erzeugen.

## 5.2 Durchführung und Teilnehmer

Man könnte versuchen, diese Fragen zu klären, indem man Testpersonen eine Reihe von realen Dokumenten präsentiert und dann mittels Blickverfolgung feststellt, welche Dokumentbereiche die Aufmerksamkeit des Betrachters anziehen. Eine derartige Untersuchung wurde in [Egl98] durchgeführt, um das dortige Verfahren zur zeit- und ortsvarianten Segmentierung von Dokumenten entsprechend der menschlichen Blickbewegung zu verifizieren. Die Untersuchung konnte allerdings nur zur Bestätigung des entwickelten Verfahrens benutzt werden, Regeln zur Implementierung des Verfahrens konnten nicht abgeleitet werden.

Auch für die hier vorliegende Arbeit hätte eine Untersuchung mittels Blickverfolgung Nachteile:

- Die einzelnen Layoutmerkmale treten nicht getrennt, sondern miteinander kombiniert auf. Dadurch kann man nicht feststellen, wie stark die Attraktivität eines Merkmals alleine ist, die für die Implementierung wichtige Schätzung der Gewichte wäre also nicht möglich.
- Selbst wenn man keine realen Dokumente verwendet, sondern Beispiele konstruiert, hängt das Ergebnis des Tests extrem von den gewählten Beispielen ab und kann durch Störfaktoren, wie Artefakte auf dem Dokument, beeinflusst werden.

Deshalb wurde es vorgezogen, zunächst eine Reihe von Personen nach ihrer Erfahrung bei der Wahrnehmung von Layoutmerkmalen zu befragen, obwohl eine Untersuchung der Augenbewegungen der Testteilnehmer eine mögliche Ergänzung der Befragung sein könnte (z.B. zur Verifikation der Ergebnisse).

Dabei beschränkte sich die Befragung im wesentlichen auf Layoutmerkmale wie sie in wissenschaftlichen Veröffentlichungen vorkommen, weil diesem Dokumenttyp aus praktischen Gründen zunächst das Hauptinteresse der Arbeit galt. So sind für wissenschaftliche Veröffentlichungen bereits Dokumentdatenbanken zu Test- und Entwicklungszwecken verfügbar (siehe Anhang C und [UoW96]), aber auch außerhalb dieser Datenbanken sind Dokumente dieses Typs in nahezu beliebiger Ausführung und Anzahl erhältlich.

Wie schon in Kapitel 2.1.3 betont, ist dem Informationsempfänger im täglichen Umgang mit Dokumenten normalerweise nicht bewußt, welche Layoutmerkmale er wahrnimmt. Allerdings können diese Merkmale auf Nachfrage hin durchaus benannt und beurteilt werden.

Die Befragung wurde in Form einer schriftlichen Umfrage stark standardisiert,

- um direkt vergleichbare Auskünfte zu erhalten,
- um eine Beeinflussung der Antworten durch den Interviewer auszuschließen,
- damit sich die Teilnehmer auf alle Fälle genug Zeit nehmen konnten, um sich die Wirkung der für gewöhnlich unbewußt benutzten Layoutmerkmale bewußt zu machen. Tatsächlich benötigte jeder Teilnehmer etwa 15 min zur Beantwortung der Fragen.

An der Befragung nahmen 46 Personen teil. Alle Teilnehmer hatten eine akademische Ausbildung und dadurch auch Erfahrung im Umgang mit wissenschaftlichen Veröffentlichungen und deren Layout. Damit verfügten Sender (Autor der Veröffentlichung) und Empfänger (Teilnehmer) zwangsläufig über die gleiche Wissensbasis, wie bei der Modellierung der Informationsübertragung (Kapitel 2.1) als unbedingt notwendig vorausgesetzt.

### 5.3 Aufbau des Fragebogens

Der Fragebogen (Anhang A) ist so aufgebaut, daß die Teilnehmer durch Frage 1 zunächst an das Thema herangeführt werden, bevor in Frage 2 nach Details gefragt wird. Frage 2 selbst kommt von der vergleichenden Betrachtung der Dokumentblöcke Text, Graphik und Bild zur Beurteilung von Text- und Graphikmerkmalen untereinander. Die letzte, offene Frage ermöglicht es noch, ohne starr vorgegebene Antworten eigene Gedanken und Kommentare anzufügen (Abbildung 5-1).

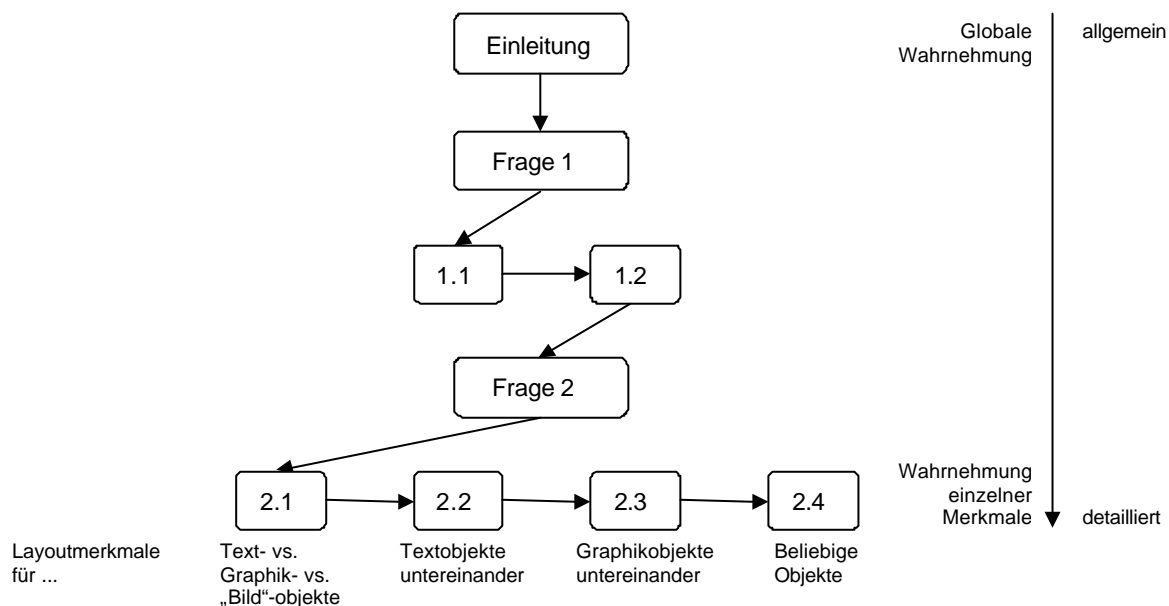


Abbildung 5-1: Aufbau des Fragebogens zur Wahrnehmung von Layoutmerkmalen

Ein Einfluß des Inhalts auf die Beurteilung ist weitestgehend ausgeschlossen, da entweder fremdsprachige Beispieldokumente benutzt werden (Frage 1) oder Text vollständig abstrahiert wird (Frage 2).

In einer *Kopfzeile* werden neben Name und Bearbeitungsdatum, Beruf und Alter des Teilnehmers abgefragt. Diese Informationen sollen bestätigen, daß die Teilnehmer Erfahrung im Umgang mit wissenschaftlichen Veröffentlichungen haben, und damit von einer gemeinsamen Wissensbasis von Sender und Empfänger ausgegangen werden kann.

Die *Einleitung* erklärt dem Teilnehmer kurz, daß es in der Befragung seine Aufgabe sein wird, Layout und einzelne Layoutmerkmale bezüglich ihrer Attraktivität zu beurteilen. Es wird an dieser Stelle besonders betont, daß es nur um die Beurteilung von Merkmalen des Layouts, nicht um die Beurteilung von Inhalt geht.

In *Frage 1* sollen die Teilnehmer an vier Beispieldokumenten angeben, wie stark ihnen einzelne Teile der Dokumente auf den ersten Blick auffallen, d.h. in welcher Reihenfolge sie sie anschauen würden. So können einerseits die Teilnehmer an die folgenden Aufgaben herangeführt werden, und andererseits kann bestätigt werden, daß es im Kreis der Testteilnehmer eine weitgehend einheitliche Wahrnehmung von Layoutmerkmalen gibt.

Die Dokumente sind so ausgewählt, daß sie möglichst viele, insbesondere alle der hier interessierenden, Layoutmerkmale enthalten und gegenüberstellen. So findet man z.B. die typographischen Merkmale Fettdruck, Kursivdruck, Großschreibung und Schriftgröße, außerdem Bullets und verschieden stark ausgeprägte Weißbereiche um die einzelnen Dokumentblöcke. Bilder unterschiedlicher Größe und innerer Struktur werden gegenübergestellt.

Die Beispieldokumente sind aus finnischen populärwissenschaftlichen Zeitungen und wissenschaftlichen Veröffentlichungen entnommen, weil man davon ausgehen konnte, daß aus dem vorgesehenen Teilnehmerkreis nur wenige Personen eine Sprache aus der finno-ugrischen Sprachgruppe beherrschen. Dadurch wurde der Einfluß des Inhalts auf die Beurteilung so weit wie möglich ausgeschaltet, weil die Teilnehmer noch nicht einmal Wortstämme oder ähnliches identifizieren können. Gleichzeitig ist der Eindruck des Schriftbildes aber immer noch der gewohnte, anders als z.B. bei arabischen oder chinesischen Dokumenten.

Frage 1.1 überprüft, ob die Teilnehmer tatsächlich kein Finnisch sprechen, weil Teilnehmer mit finnischen Sprachkenntnissen bei der Auswertung von Frage 1 nicht berücksichtigt werden.

Frage 1.2 präsentiert dann die Beispieldokumente. Dabei werden im Fragebogen verkleinerte Versionen der Dokumente mit den zu beurteilenden, markierten Teilen gezeigt, in die die Bewertung eingetragen werden soll. Die eigentliche Beurteilung sollen die Teilnehmer aber mit Hilfe der originalen Dokumentversionen durchführen, die in einem abzutrennenden Anhang in Originalgröße mitgegeben sind, damit ihr Eindruck vom Dokument nicht durch die Markierungen oder durch die falsche Größe verändert wird.

*Frage 2* ist der Kern der Umfrage. Hier geht es nicht mehr um die globale Wahrnehmung eines Dokuments, sondern um die einzelnen Merkmale, die zur Attraktivität beitragen. Prinzipiell sollen die Teilnehmer vornehmlich aus ihrer Erfahrung bei der spontanen Wahrnehmung von Dokumenten berichten, allerdings werden zu jeder Frage Gestaltungsbeispiele mit stark abstrahiertem Text gegeben. Dadurch sind die verwendeten Begriffe geklärt.

Frage 2.1 soll unsere Annahme bestätigen, daß Liniengraphiken, Textblöcke und Bilder die Aufmerksamkeit unterschiedlich stark anziehen, und eine Reihenfolge dieser drei Blockarten bezüglich der Attraktivität erzeugen. Obwohl man die Teilnehmer auffordert, aus ihrer Erfahrung zu urteilen, muß man bei den Beispielen darauf achten, daß die Liniengraphik und das Bild beide ähnlich irregulär wirken. Ein reguläres Bild setzt sich nämlich sonst nicht ausreichend gegen eine unruhige Liniengraphik ab.

In Frage 2.2 werden anhand von Beispielen sieben Layoutmerkmale für Wörter, bzw. Textblöcke, vorgeschlagen, für die die Teilnehmer eine Reihenfolge in der Attraktivität festlegen sollen

Frage 2.3 stellt zwei Layoutmerkmale für Graphikblöcke, nämlich die Größe des umgebenden Weißbereichs („Freistellung“) und die Größe des Blocks an sich, gegenüber. Die Teilnehmer sollen auch hier markieren, durch welches Merkmal die Aufmerksamkeit eines Betrachters auf den ersten Blick stärker angezogen wird.

Die bisher beschriebenen Fragen sind sogenannte geschlossene Fragen mit vorgegebenen Antwortmöglichkeiten, bei denen durch die Vorgaben eine gewisse Beeinflussungsgefahr besteht. Mit Frage 2.4 folgt dagegen eine offene Frage, deren Beantwortung die Teilnehmer völlig frei gestalten können. In dieser Frage können die Teilnehmer selbst wichtige Layoutmerkmale und deren Bewertung vorschlagen. Dadurch wird sichergestellt, daß in Umfrage und Implementierung kein wichtiges Layoutmerkmal unberücksichtigt bleibt.

Allgemeine Hinweise zu Aufbau und Gestaltung eines Fragebogens finden sich in [Ber96] und [Sud82].

## **5.4 Auswertung: Ergebnisse und Diskussion**

### **5.4.1 Kopfzeile**

Da viele Teilnehmer die Kopfzeile aus Datenschutzgründen nur lückenhaft ausfüllten, konnte eine statistische Auswertung nicht durchgeführt werden.

### **5.4.2 Frage 1**

Bei der Auswertung von Frage 1 wurden nur 45 Teilnehmer berücksichtigt, da einer der Teilnehmer in Frage 1.1 angab, Finnisch zu sprechen.

In Abbildung 5-2 sind die Dokumentbereiche, die in Frage 1.2 zu bewerten waren, in jedem Dokument von oben nach unten fortlaufend mit den Buchstaben des Alphabets bezeichnet, so daß die Teile bei der folgenden Darstellung der Ergebnisse einfach referenziert werden können.

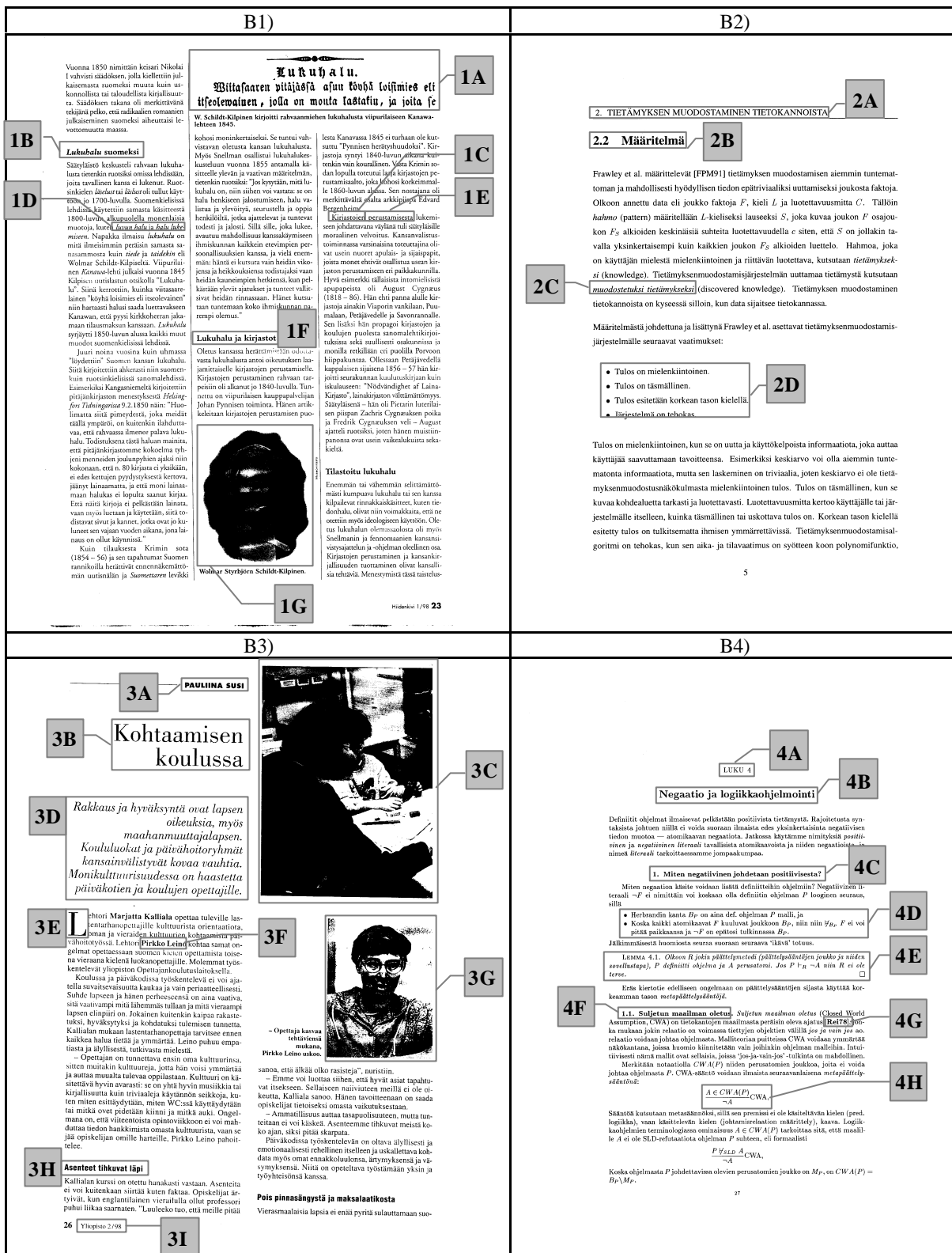


Abbildung 5-2: Bezeichnungen der zu bewertenden Dokumentbereiche

Die Testteilnehmer gaben für jedes Dokument die Reihenfolge der Dokumentbereiche bezüglich ihrer Attraktivität an, wobei der Stellenwert eines Dokumentbereiches in einem Dokument mit  $I$  zu bewertenden Bereichen durch die Zahlenwerte  $R_i \in \{1, \dots, I\}$  dargestellt wurde.

Zur Auswertung wurden zunächst für jeden bewerteten Dokumentbereich der *Mittelwert*  $\bar{R}_i$  der angegebenen Stellenwerte bezüglich der Attraktivität

$$\bar{R}_i = \frac{\sum_{j=1}^J R_{ij}}{J} \quad (5-1)$$

und die *mittlere Abweichung*  $\tilde{R}_i$  des Stellenwertes bezüglich der Attraktivität

$$\tilde{R}_i = \frac{\sum_{j=1}^J |R_{ij} - \bar{R}_i|}{J} \quad (5-2)$$

berechnet, wobei  $J = 45$  die Anzahl der Teilnehmer ist.

Aus den Mittelwerten ergibt sich für jedes Dokument eine Reihenfolge der Dokumentbereiche bezüglich ihrer Attraktivität, die man *Ranking* nennt.

Abbildung 5-3 zeigt das Ranking der Dokumentbereiche für die vier Testdokumente. In Klammern ist hinter der Bezeichnung jedes Dokumentbereichs die *mittlere Abweichung* der angegebenen Stellenwerte bezüglich der Attraktivität aufgeführt. Die mittlere Abweichung dient als Maß dafür, wie sicher sich die Teilnehmer bei der Bewertung eines Dokumentbereichs waren. Je größer die mittlere Abweichung, desto unsicherer waren die Teilnehmer. Abbildung 5-3 gibt außerdem die *mittlere Abweichung pro Dokument*

$$\bar{\tilde{R}} = \frac{\sum_{i=1}^I \tilde{R}_i}{I}, \quad (5-3)$$

mit  $I$  als Anzahl der zu bewertenden Dokumentbereiche pro Dokument, an. Sie zeigt wie leicht oder schwer ein Dokument insgesamt zu bewerten war.

Bei Dokument B1) fallen folgende Punkte besonders auf:

- Entgegen den Erwartungen und den Ergebnissen von Frage 2.1 wird nicht das Bild (1G), sondern der vergrößerte Textabschnitt (1A) im Ranking an die erste Stelle gesetzt. Hier spielt vermutlich die Lesereihenfolge, bzw. die Lage im Dokument eine Rolle, deren Einfluß die Testteilnehmer nicht vollständig ausschalten konnten.
- Die Textabschnitte folgen im Ranking mit deutlichem Abstand.
- Eine Überschrift, die einen Anteil in Kursivschrift enthält (1B), ist etwas attraktiver als eine ohne Kursivanteil (1F).
- Ein Wort nach einer Einrückung (1C) fällt etwas mehr auf als ein kursives Wort (1D), und dies wiederum mehr als ein „normales“ Wort (1E).
- Schwierigkeiten, entsprechend der hohen mittleren Abweichung, macht offensichtlich die Bewertung der Markierungen innerhalb des Textes, nämlich die Kursivschrift und die Einrückung.

Dokument B2) war insgesamt am leichtesten zu beurteilen, wohl weil hier nur wenige, deutlich verschiedene Layoutmerkmale zu sehen sind. Es fällt auf, daß

- die fettgedruckte, freigestellte Überschrift (2B) auf Rang 1 liegt, und
- die kursiv markierten Wörter im Text (2C) die geringste Attraktivität haben.

Für Dokument B3) ist die mittlere Abweichung pro Dokument bereits höher, da die Gestaltung komplexer und daher schwieriger zu beurteilen ist. Interessant sind hier folgende Anmerkungen:

- Offensichtlich fallen Bilder nicht grundsätzlich vor allen anderen Dokumentbereichen auf. Das kleinere Bild 3G steht im Ranking erst an dritter Stelle und fällt erst nach der großen Überschrift (3B) auf. Dies weist bereits darauf hin, daß große Schriftgröße ein sehr attraktives Layoutmerkmal ist (siehe Frage 2.2 und Kapitel 4.3.3).
- Die Teile 3A und 3H sind vom Layout her sehr ähnlich (gleiche Schriftgröße, gleicher Font und Fontstil) und erhalten in der Befragung folgerichtig auch ein fast identisches Ranking. 3A ist in Großbuchstaben geschrieben und daher etwas attraktiver.

Dokument B4) konnte am schlechtesten beurteilt werden. Offensichtlich setzen sich die einzelnen Dokumentbereiche nicht ausreichend von einander ab. Mit einer mittleren Abweichung von 2.34 ist insbesondere die Kapitelnummerierung (4A) überhaupt nicht mehr einzuordnen.

|   |   |  |   |
|---|---|--|---|
| Ranking<br>1<br>2<br>3<br>4<br>5<br>6<br>7<br>$\bar{R}$           | <b>Dok. B1)</b><br>- 1A (0.55)<br>- 1G (0.74)<br><br>- 1B (0.52)<br>- 1F (0.54)<br><br>= 1C (0.61)<br>1D (0.77)<br>- 1E (0.52)<br><br><b>(0.61)</b>                             | Ranking<br>1<br>2<br>3<br>4<br>$\bar{R}$                     | <b>Dok. B2)</b><br>- 2B (0.35)<br><br>- 2C (0.13)<br>- 2A (0.61)<br><br>- 2D (0.56)<br><br><b>(0.41)</b>  |
| Ranking<br>1<br>2<br>3<br>4<br>5<br>6<br>7<br>8<br>9<br>$\bar{R}$ | <b>Dok. B3)</b><br>= 3C (0.86)<br>3B (0.71)<br>- 3G (1.01)<br><br>- 3D (0.67)<br><br>= 3A (1.31)<br>3H (1.09)<br>- 3E (1.31)<br>- 3F (0.81)<br>- 3I (0.71)<br><br><b>(0.94)</b> | Ranking<br>1<br>2<br>3<br>4<br>5<br>6<br>7<br>8<br>$\bar{R}$ | <b>Dok. B4)</b><br>- 4B (0.25)<br><br>- 4C (0.86)<br>- 4H (1.31)<br><br>= 4A (2.34) 4D (1.15)<br>= 4F (1.41)<br>4E (1.14)<br><br>- 4G (0.76)<br><br><b>(1.15)</b> |

Abbildung 5-3: Auswertung der Bewertung von Dokumentbereichen

### 5.4.3 Frage 2

Frage 2 wurde von allen 46 Teilnehmern beantwortet.

Abbildung 5-4 zeigt die Auswertung von Frage 2.1 bis Frage 2.3. Analog zur Auswertung von Frage 1.2 ist hier pro Frage das Ranking der einzelnen Layoutmerkmale, d.h. ihre Reihenfolge bezüglich der Attraktivität, angegeben. Hinter jedem Layoutmerkmal ist jeweils wieder die mittlere Abweichung der angegebenen Stellenwerte bezüglich der Attraktivität angegeben.

Entsprechend der mittleren Abweichung pro Dokument ist hier unter der Bezeichnung  $\bar{R}$  jeweils die mittlere Abweichung für die ganze Frage angegeben.

|   |   |   |
|---|---|---|
| <p><b>Frage 2.1</b></p> <p>1<br/>— Bild (0.43)</p> <p>2<br/>— Liniengraphik (0.31)</p> <p>3<br/>— Textblock (0.41)</p> <p><math>\bar{R}</math> 3<br/>(0.38)</p> | <p><b>Frage 2.2</b></p> <p>1<br/>— Freistellung (0.43)</p> <p>2<br/>— Schriftgröße (0.87)</p> <p>3<br/>— Fettdruck (0.88)</p> <p>4<br/>— Bullets (1.26)</p> <p>5<br/>— Einrückung (1.24)</p> <p>6<br/>— Großschreibg. (1.10)</p> <p>7<br/>— Kursivdruck (0.86)</p> <p><math>\bar{R}</math> 7<br/>(0.95)</p> | <p><b>Frage 2.3</b></p> <p>1<br/>— Freistellung (0.40)</p> <p>2<br/>— Größe (0.40)</p> <p><math>\bar{R}</math> 2<br/>(0.40)</p> |
|---|---|---|

Abbildung 5-4: Auswertung der Fragen zur Attraktivität einzelner Layoutmerkmale

Für alle drei Fragen ergibt sich jeweils ein eindeutiges Ranking der Merkmale, das sich für eine Umsetzung in eine Attraktivitätsbewertung eignet (Kapitel 4.3). Zwar erhält man bei Frage 2.2 eine relativ hohe mittlere Abweichung, da jeder Teilnehmer die sieben Merkmale etwas anders beurteilt, eine Reihenfolge wird aber trotzdem deutlich.

Die Auswertung der offenen Frage 2.4 ist in Abbildung 5-5 zu sehen.

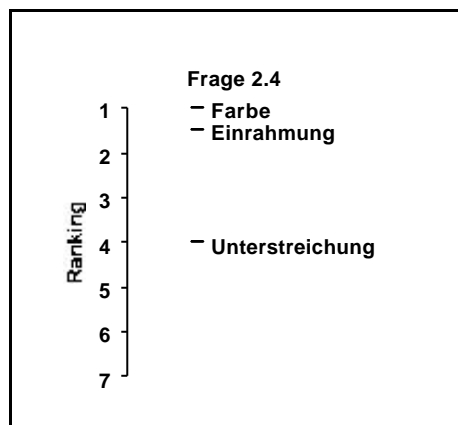


Abbildung 5-5: Auswertung der offenen Frage bezüglich weiterer Layoutmerkmale und ihrer Attraktivität

Von allen vorgeschlagenen Merkmalen sind die drei am häufigsten genannten Merkmale mit der jeweils vorgeschlagenen Bewertung aufgeführt. An erster Stelle steht dabei das Merkmal Farbe, dessen Bewertung bei Farbdokumenten sicher außerordentlich wichtig ist. Die vorliegende Arbeit befaßt sich ausschließlich mit Binärbildern, daher kann eine Bewertung der Farbe zur Zeit nicht erfolgen. Da für die Zukunft aber damit zu rechnen ist, daß in Datenbanken auch farbig gescannte Dokumente immer häufiger auftreten, sollte man bei einer erweiterten Version der Attraktivitätsbewertung auch eine Bewertung der Farbe in Erwägung ziehen. Die Hinweise auf die beiden weiteren Merkmale Einrahmung und Unterstreichung sind ebenfalls richtig. Klassifikationsverfahren für diese Merkmale sind in der aktuellen Programmversion



noch nicht berücksichtigt, da zunächst die prinzipielle Machbarkeit des Ansatzes gezeigt werden sollte. Für eine erweiterte Version der Attraktivitätsbewertung wäre es aber sinnvoll, diese Merkmale miteinzubeziehen.



# 6 Experimente und Tests

## 6.1 Experimente zur Bestimmung von Entscheidungsschwellen

Bei der Klassifikation der Textons bezüglich der Fontstile ist zur Bestimmung von Entscheidungsschwellen eine genauere Untersuchung der Merkmale nötig, die den jeweiligen Fontstil definieren, weil der Zusammenhang zwischen der Ausprägung dieser Merkmale und ihrer zahlenmäßigen Repräsentation nicht wie bei den anderen Layoutmerkmalen ohne weiteres erkennbar ist.

Bei der Klassifikation eines Textons bezüglich eines Fontstils muß das Merkmal eines Textons global mit den Merkmalen anderer Textons verglichen werden. Dieser Vergleich wird bei der Klassifikation bezüglich Fettdruck durch eine adaptive Entscheidungsschwelle realisiert, die von der Verteilung des Schwärzungsgrades auf dem ganzen Dokument abgeleitet ist. Zur Klassifikation bezüglich der Fontstile Kursivdruck (Kapitel 4.2.4.2) und Großschreibung (Kapitel 4.2.4.3) dagegen wird der großflächige Vergleich durch Experimente vorweggenommen.

Die Experimente werden auf synthetischen Dokumenten durchgeführt, um die Merkmale, die den jeweiligen Fontstil definieren, unter standardisierten Bedingungen untersuchen zu können.

Es wurden synthetische Dokumente in vier verschiedenen Schriften und drei verschiedenen Schriftgrößen, d.h. zwölf verschiedenen Fonts, benutzt, wobei für jeden dieser Fonts der Fontstil noch zwischen „normal“, Kursiv- und Fettdruck variiert wurde. Alle synthetischen Dokumente wurden auf einem Laserdrucker mit einer Auflösung von 600 dpi gedruckt und anschließend mit einer Auflösung von 300 dpi eingescannt.

Jedes synthetische Dokument enthält das Alphabet einmal in Großbuchstaben und einmal in Kleinbuchstaben, sowie die Ziffern von 0 bis 9. Die Buchstaben sind entsprechend ihrer Unter- und Oberlängen in Gruppen aufgeteilt, so daß verschieden strukturierte Buchstaben getrennt untersucht werden können. Abbildung 6-1 zeigt ein Beispiel eines synthetischen Dokuments.

A B C D E F G H I J K L M N O P R S T U V W X Y Z

Q

a c e i m n o r s u v w x z

b d f h k l t

g j p q y

1 2 3 4 5 6 7 8 9 0

Abbildung 6-1: Beispiel eines synthetischen Dokuments in Schrift Graphite Light

Eine Entscheidungsschwelle für die Klassifikation bezüglich Kursivdruck ist die untere Schranke für die mittlere Schräglage der ungefähr senkrechten Strokes in kursiv gedrucktem Text. Dieser Wert wird aus den synthetischen Dokumenten ermittelt, indem die mittlere Schräglage der ungefähr senkrechten Strokes über jedes der synthetischen Dokumente berechnet wird. Dabei verwendet man einmal die synthetischen Dokumente in Kursivdruck und einmal die in normalem Fontstil. Ein Vergleich der sich ergebenden Werte in Tabelle 6-1 läßt eine Schwelle von  $TH = 0.12$  sinnvoll erscheinen.

| Font                  | Mittlere Schräglage |        |
|-----------------------|---------------------|--------|
|                       | kursiv              | normal |
| Arial, 8pt            | 0.13                | 0.02   |
| Arial, 10pt           | 0.13                | 0.02   |
| Arial, 12pt           | 0.11                | 0.02   |
| Times New Roman, 8pt  | 0.17                | 0.01   |
| Times New Roman, 10pt | 0.19                | 0.01   |
| Times New Roman, 12pt | 0.19                | 0.01   |
| Graphite Light, 8pt   | 0.19                | 0.11   |
| Graphite Light, 10pt  | 0.25                | 0.12   |
| Graphite Light, 12pt  | 0.19                | 0.11   |
| Courier, 8pt          | 0.13                | -0.01  |
| Courier, 10pt         | 0.14                | 0.01   |
| Courier, 12pt         | 0.14                | 0.02   |

Tabelle 6-1: Mittlere Schräglage der ungefähr senkrechten Strokes auf synthetischen Dokumenten

Zur Klassifikation bezüglich Großschreibung benötigt man das Größenverhältnis der Kleinbuchstaben ohne Unter- und Oberlänge zu den Großbuchstaben. Tabelle 6-2 zeigt die für alle synthetischen Dokumente in normalem Fontstil ermittelten Werte. Der Durchschnittswert für alle Fonts ist 0.74, ein Verhältnis, das bei der Definition der tatsächlichen Entscheidungsschwelle als Anhaltspunkt dient.

| Font                  | Größenverhältnis |
|-----------------------|------------------|
| Arial, 8pt            | 0.76             |
| Arial, 10pt           | 0.76             |
| Arial, 12pt           | 0.75             |
| Times New Roman, 8pt  | 0.72             |
| Times New Roman, 10pt | 0.72             |
| Times New Roman, 12pt | 0.70             |
| Graphite Light, 8pt   | 0.67             |
| Graphite Light, 10pt  | 0.73             |
| Graphite Light, 12pt  | 0.65             |
| Courier, 8pt          | 0.83             |
| Courier, 10pt         | 0.77             |
| Courier, 12pt         | 0.76             |

Tabelle 6-2: Größenverhältnis der Kleinbuchstaben ohne Unter- und Oberlängen zu den Großbuchstaben auf synthetischen Dokumenten

## 6.2 Erkennungsraten der Klassifikation bezüglich ausgewählter Layoutmerkmale

Um die Erkennungsleistung der in Kapitel 4.2 vorgeschlagenen Verfahren zur Klassifikation von Textons zu untersuchen, vergleicht man auf einer Reihe von Testdokumenten die tatsächlichen Ergebnisse der Klassifikationsverfahren mit den theoretisch optimalen Ergebnissen. Dies ist natürlich nur für diejenigen Layoutmerkmale sinnvoll, die zwei Klassen im Merkmalsraum erzeugen, nicht jedoch für Layoutmerkmale, bei denen lediglich Meßwerte im Rahmen der Meßgenauigkeit in Klassen abgebildet werden. Bei diesen Layoutmerkmalen ist eher die Frage, ob sie im Hinblick auf eine Attraktivitätsbewertung korrekt definiert sind. Diese Frage ist nur subjektiv zu beantworten und kann deshalb auch nur mit einem interaktiven Test, also unter Beteiligung von realen Betrachtern beantwortet werden (siehe Kapitel 6.3).

Es wurden also zunächst nur die Verfahren zur Klassifikation von Textons bezüglich Einrückungen, Bullets und bezüglich der Fontstile Fettdruck, Kursivdruck und Großschreibung getestet.

Die Tests wurden auf den in Anhang C abgebildeten 10 Testdokumenten durchgeführt. Die Dokumente zeigen große Variabilität im Layout und sind in drei verschiedenen Sprachen (Finnisch, Deutsch, Englisch) geschrieben.

Da die theoretisch optimalen Ergebnisse zu den Dokumenten nicht in maschinenlesbarer Form vorliegen, d.h. es keine sogenannte „Groundtruth“ zum Testen gibt, muß die Auswertung der Klassifikationsergebnisse manuell erfolgen. Dieses ausgesprochen zeitaufwendige Verfahren erklärt die geringe Anzahl von verwendeten Testdokumenten. Selbstverständlich können die mit nur 10 Dokumenten gewonnenen Ergebnisse keinen statistischen Nachweis einer Erkennungsrate liefern, eine Abschätzung der Erkennungsleistung erscheint aber dennoch möglich.

Beim Testen wurden für jedes Wort-Texton die tatsächlichen Layoutmerkmale mit den klassifizierten Layoutmerkmalen verglichen. Dabei entsprechen die während des Segmentierungsprozesses entstehenden Wort-Textons nicht zwangsläufig den Wörtern, wie ein Mensch sie wahrnehmen würde. In der automatischen Segmentierung sind nämlich solche „realen“ Wörter zum Teil aufgetrennt, aber auch verbunden. Satzzeichen, wie Punkte oder Kommas, und i-Punkte werden im Segmentierungsprozeß im allgemeinen ebenfalls als jeweils ein Wort-Texton erzeugt. Diese Textons wurden im Test nicht berücksichtigt, da an ihnen Layoutmerkmale auch von Menschen kaum zu erkennen sind.

Tabelle 8-1 in Anhang D führt die Testergebnisse im Detail auf. Tabelle 6-3 gibt eine Zusammenfassung.

Offensichtlich besteht bei der Klassifikation bezüglich Großschreibung ein systematisches Problem darin, Zahlen von großgeschriebenen Wörtern zu unterscheiden. Ziffern haben nämlich fast die gleiche Höhe wie Großbuchstaben, die Buchstabenhöhe aber wird – nach Julesz – als einfaches, beobachtbares Merkmal zur Klassifikation benutzt. Es ist an dieser Stelle zu vermuten, daß eine rein visuelle Unterscheidung von Ziffern und Großbuchstaben nicht möglich ist, sondern inhaltliche Information auf alle Fälle berücksichtigt werden müßte.

| Layoutmerkmal  | Layoutmerkmal | Layoutmerkmal ^<br>Layoutmerkmal(w) <br>(Erkennungsrate) | Kein Layoutmerkmal | Kein Layoutmerkmal ^<br>Layoutmerkmal(w) <br>(„Fehlalarm“-Rate) |
|----------------|---------------|--|--------------------|---|
| Einrückungen   | 27            | 26 (96%)   | 6558               | 0 (0.00%)   |
| Bullets        | 7             | 7 (100%)   | 6578               | 1 (0.01%)   |
| Fettdruck      | 470           | 467 (99%)  | 6115               | 4 (0.07%)   |
| Kursivdruck    | 108           | 104 (96%)  | 6477               | 4 (0.06%)   |
| Großschreibung | 28            | 25 (89%)   | 6557               | 2 (0.03%)   |
|                |               |  |                    | + 23 Zahlen   |

Tabelle 6-3: Erkennungsraten bezüglich Einrückungen, Bullets, Fettdruck, Kursivdruck und Großschreibung auf den Testdokumenten T1) bis T10) aus Anhang C

Ein erster Test der hier präsentierten Klassifikationsverfahren bezüglich der Fontstile Fettdruck, Kursivdruck und Großschreibung wurde bereits in [Sch98a] veröffentlicht. Dieser Test wurde auf Dokumenten der UW-III Document Image Database ([UoW96]) durchgeführt (leider bietet auch diese Datenbank keine hier geeignete, wortweise Groundtruth). Es wurden die 20 Dokumente S000bin.tif bis S00kbin.tif ausgewählt, die mit einer Auflösung von 300 dpi direkt aus originalen wissenschaftlichen Veröffentlichungen gescannt sind.

Die Ergebnisse werden ausführlich in Tabelle 8-2 (Anhang D) und zusammengefaßt in Tabelle 6-4 angegeben.

| Layoutmerkmal  | Layoutmerkmal | Layoutmerkmal ^<br>Layoutmerkmal(w) <br>(Erkennungsrate) | Kein Layoutmerkmal | Kein Layoutmerkmal ^<br>Layoutmerkmal(w) <br>(„Fehlalarm“-Rate) |
|----------------|---------------|--|--------------------|---|
| Fettdruck      | 568           | 542 (95%)<br>+ 11Großgeschriebene                        | 16148              | 9 (0.06%)   |
| Kursivdruck    | 378           | 350 (93%)  | 16338              | 19 (0.12%)  |
| Großschreibung | 94            | 89 (95%)   | 16622              | 24 (0.14%)  |
|                |               |  |                    | + 103 Zahlen  |

Tabelle 6-4: Erkennungsraten bezüglich Fettdruck, Kursivdruck und Großschreibung aus [Sch98a]

Beim Vergleich der Ergebnisse mit den aktuellen Ergebnissen ist zu beachten, daß sich die ältere Definition der Klassifikationsverfahren in drei Punkten von der in der Arbeit präsentierten Version unterscheidet:

1. Ähnlichkeits- und Nachbarschaftsrelation der Segmentierung waren noch nicht für die Analyse von wissenschaftlichen Veröffentlichungen optimiert.
2. Bei der Klassifikation von Fettdruck war die Mittelzone eines Wortes  $w$  noch als

$$\begin{aligned}
 \text{MiddleZone}(w) := \\
 \{(x,y): \text{Left}(w) \leq x \leq \text{Right}(w) \wedge \text{Base}(l) \leq y \leq (\text{Base}(l) + 0.67 \times (\text{Top}(w) - \text{Base}(l) + 1))\}
 \end{aligned}
 \tag{6-1}$$

definiert. Dadurch ergab sich das Problem, daß großgeschriebene Wörter in Fettdruck oft als nicht fettgedruckt klassifiziert wurden, weil ein großer Anteil der Buchstaben außerhalb der definierten Mittelzone des Wortes lag.

3. Bei der Klassifikation bezüglich Großschreibung wurden Wörter bereits ab einer Länge von 4 Buchstaben behandelt.

Beide Tests demonstrieren die gute Erkennungsleistung der Verfahren. Obwohl keine absolut korrekte Klassifikation erzielt wird, eignen sich die Verfahren als Grundlage einer Attraktivitätsberechnung, bei der die Ergebnisse der Klassifikationen bezüglich mehrerer Layoutmerkmale gemeinsam bewertet werden. Einzelne Fehler werden dabei oft ausgeglichen.

Die festgestellten Erkennungsraten für die Klassifikation bezüglich der Fontstile liegen etwas unter den Erkennungsraten von Chaudhuri und Garain ([Cha98], Table 1), die die Struktur der Buchstaben zu Klassifikationszwecken sehr genau vermessen. Die Erkennungsraten von Zramdini und Ingold für etwa wortlange Textabschnitte ([Zra98], Table 3, Spalte „¼ L“) liegen dagegen ziemlich genau in der hier ermittelten Größenordnung, da Zramdini und Ingold ähnlich der vorliegenden Arbeit visuelle Aspekte des Texts untersuchen.

### 6.3 Interaktiver Test der Attraktivitätsbewertung

#### 6.3.1 Ziel des Tests

Der sogenannte interaktive Test sollte die für jedes Objekt auf dem Dokument berechnete Attraktivitätsbewertung (Kapitel 4.3) untersuchen. Da diese Bewertung ganz wesentlich von der Gewichtung der einzelnen Klassifikationsergebnisse abhängt, wurden implizit also die aus der Befragung zur Wahrnehmung von Layoutmerkmalen (Kapitel 5) abgeleiteten Gewichte geprüft. Außerdem wurden damit natürlich auch alle Klassifikationsverfahren bezüglich der Layoutmerkmale getestet, diesmal auch diejenigen, die Meßwerte in Klassen umsetzen.

#### 6.3.2 Durchführung und Teilnehmer

Die Attraktivität ist eine Bewertung, die subjektiv empfunden wird, also nur unter Beteiligung von realen Betrachtern getestet werden konnte.

Den Testteilnehmern wurden nacheinander fünf Dokumente und deren normierte Attraktivitätsbewertung  $A'$  nach Kapitel 4.3.5 präsentiert. Die Teilnehmer konnten dabei die Attraktivitätsbewertung derjenigen Objekte verändern, für die diese Bewertung ihrem Eindruck nach nicht korrekt war. Die von den Teilnehmern vorgenommenen Änderungen wurden automatisch protokolliert und anschließend ausgewertet.

Die verwendeten Testdokumente sind wie schon bei der Befragung (Kapitel 5) aus finnischen populärwissenschaftlichen Zeitungen und wissenschaftlichen Veröffentlichungen entnommen. Das verhindert, daß der Inhalt des Dokuments Einfluß auf die Beurteilung der Attraktivität hat. Die in Anhang B abgedruckte schriftliche Anleitung zum Test überprüft deshalb zunächst mit einer Frage, ob der jeweilige Teilnehmer Finnisch spricht.

Die Testdokumente wurden so ausgewählt, daß neben der Attraktivitätsberechnung implizit auch die Klassifikationsverfahren für alle Layoutmerkmale und die entsprechende Gewichtung der Layoutmerkmale getestet wurden. Tabelle 6-5 gibt eine Übersicht über die jeweils in den Dokumenten benötigten Klassifikationsverfahren und Gewichte. Offensichtlich wirken an der Attraktivitätsbewertung eines Dokuments immer eine ganze Reihe von Einflüssen mit, da

sie aber an unterschiedlichen Stellen im Dokument auftreten, sind sie für einen Test durchaus zu trennen.

|              | Einrückungen,<br>$g_{\text{Incident}}$ | Bullets,<br>$g_{\text{Bullet}}$ | Fettdruck,<br>$g_{\text{Bold}}$ | Kursivdruck,<br>$g_{\text{Italics}}$ | Großschreibung,<br>$g_{\text{AllCaps}}$ | Freistellung,<br>$g_{\text{Free}}$ | Schriftgröße,<br>$g_{\text{MeanHeight}}$ | Flächenmaß,<br>$g_{\text{AreaMeas}}$ | Schwärzungsgrad,<br>$g_{\text{BlackIntens}}$ | $d_{\text{Graphik}}$ |
|--------------|--|---------------------------------|---------------------------------|--------------------------------------|---|------------------------------------|--|--------------------------------------|--|----------------------|
| Dokument I1) | x                                      |                                 | x                               |                                      | x                                       | x                                  | x  | x                                    | x  | x                    |
| Dokument I2) | x                                      | x                               | x                               | x                                    | x                                       | x                                  | x  |                                      |  |                      |
| Dokument I3) | x                                      |                                 | x                               | x                                    | x                                       | x                                  | x  | x                                    | x  | x                    |
| Dokument I4) | x                                      |                                 | x                               | x                                    | x                                       | x                                  | x  |                                      |  |                      |
| Dokument I5) | x                                      |                                 | x                               | x                                    |   | x                                  |  | x                                    | x  | x                    |

Tabelle 6-5: Testbarkeit der Klassifikationsverfahren für Layoutmerkmale und der entsprechenden Gewichte auf den Testdokumenten I1) bis I5) im interaktiven Test

Abbildung 6-2 zeigt die Attraktivitätsbewertungen der fünf Testdokumente. Die Attraktivitätsbewertung  $A'$  ist farbcodiert so dargestellt, daß gelbe Farbe niedrige Attraktivität und rote Farbe hohe Attraktivität bedeutet.

Die schriftliche Anleitung zum Test zeigt in einem Anhang die Originalversionen der Dokumente, d.h. die Dokumente sind auf Papier, in Originalgröße und ohne die an sich schon wieder attraktive Farbcodierung dargestellt. Mit Hilfe dieser Originale konnten sich die Teilnehmer über die Attraktivität der Dokumente klar werden, bevor sie die Anzeige der Attraktivitätsbewertung auf dem Bildschirm sahen und beurteilten.

Am Test beteiligten sich 16 Personen. Analog zur Befragung hatten alle Teilnehmer eine akademische Ausbildung und dadurch auch Erfahrung im Umgang mit wissenschaftlichen Veröffentlichungen und deren Layout. Damit verfügten Sender (Autor der Veröffentlichung) und Empfänger (Testteilnehmer) zwangsläufig über die gleiche Wissensbasis, wie bei der Modellierung der Informationsübertragung (Kapitel 2.1) vorausgesetzt.







### 6.3.3 Auswertung: Ergebnisse und Diskussion

Keiner der 16 Teilnehmer sprach Finnisch.

Eine Attraktivität haben Wörter  $w \in WORDS$  und Graphikblöcke  $b \in BLOCKS$ . Für jedes dieser Textons, bzw. Objekte, gibt es nach dem Test

1. eine präsentierte Attraktivität  $A'_p(w, b)$ , die der berechneten Attraktivitätsbewertung  $A'(w, b)$  entspricht, und
2. für jeden von  $K$  Teilnehmern eine korrigierte Attraktivität  $A'_{ck}(w, b)$ .

Für jedes Objekt kann man damit den *Mittelwert der korrigierten Attraktivitäten*  $M(w, b)$  berechnen:

$$M(w, b) = \sum_{k=1}^K A'_{ck}(w, b) / K \quad (6-2)$$

Damit ergibt sich für jedes Objekt eine *Differenz zwischen der präsentierten Attraktivität und dem Mittelwert der korrigierten Attraktivitäten*  $D(w, b)$ :

$$D(w, b) = A'_p(w, b) - M(w, b) \quad (6-3)$$

Dieses Maß für die Größe und Richtung der vorgenommenen Korrekturen ist in Abbildung 6-3 farbcodiert dargestellt. Grüne Farbe bedeutet dabei, daß die präsentierte Attraktivität deutlich zu klein war, hellblaue Farbe bedeutet geringe Korrekturen bei der Attraktivität, und dunkelblaue Farbe bedeutet, daß die präsentierte Attraktivität deutlich zu hoch war. Zusätzlich ist für jedes Objekt in der linken oberen Ecke des Objektes der Wert  $\lfloor 100 \cdot D(w, b) \rfloor$  angegeben. Objekte, bei denen kein Teilnehmer die Attraktivität korrigiert hat, erscheinen in der gelb-roten Farbe ihrer Attraktivitätsbewertung.

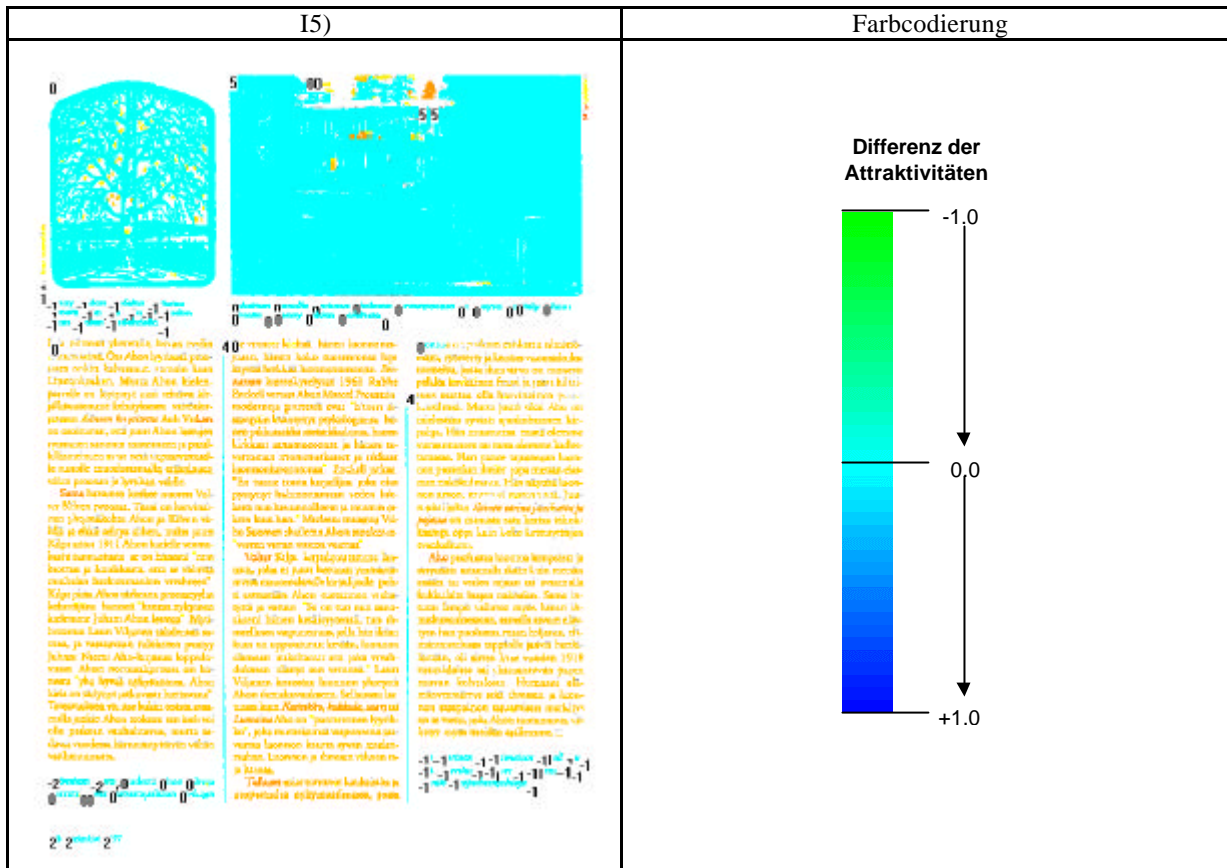
Neben der Differenz  $D(w, b)$  kann man für jedes Objekt auch noch die *mittlere Abweichung der korrigierten Attraktivitäten*  $V(w, b)$  zum Mittelwert  $M(w, b)$  berechnen:

$$V(w, b) = \sum_{k=1}^K |A'_{ck}(w, b) - M(w, b)| / K \quad (6-4)$$

Dieser Wert ist ein Maß dafür, wie groß die Übereinstimmung der Teilnehmer bei der Korrektur eines Attraktivitätswertes war. Abbildung 6-4 zeigt die mittlere Abweichung farbcodiert an. Dabei ist prinzipiell die gleiche Codierung wie zur Darstellung der Differenz  $D(w, b)$  benutzt worden, allerdings nimmt  $V(w, b)$  keine negativen Werte an, weshalb der grünliche Bereich der Farbtabelle nicht genutzt wird. Auch in dieser Darstellung gibt es in der linken oberen Ecke jedes Objekts den Wert  $\lfloor 100 \cdot V(w, b) \rfloor$ . Objekte, bei denen kein Teilnehmer die Attraktivität korrigiert hat, erscheinen in der gelb-roten Farbe ihrer Attraktivitätsbewertung.



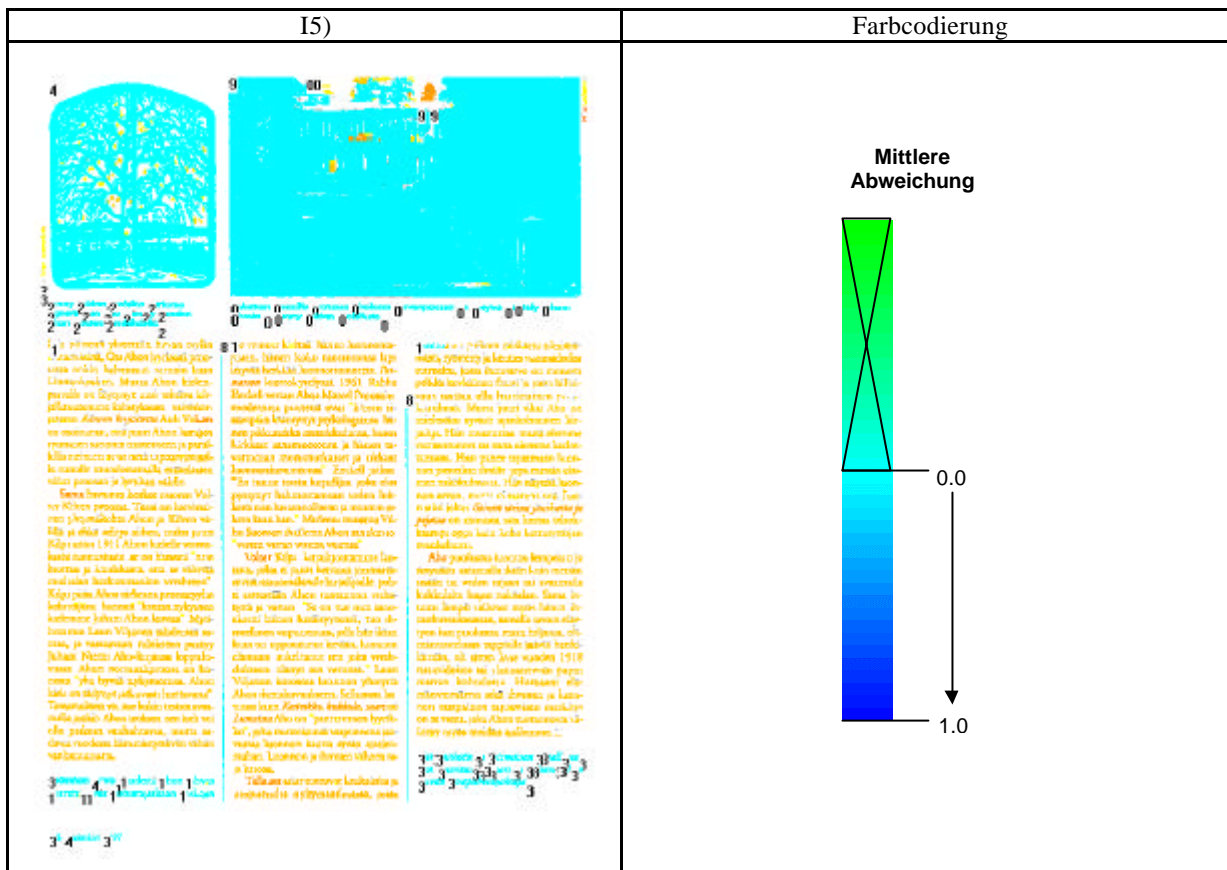
Abbildung 6-3: Differenz der präsentierten Attraktivität und des Mittelwertes der korrigierten Attraktivitäten  $D(w, b)$  im interaktiven Test



Fortsetzung zu Abbildung 6-3: Differenz der präsentierten Attraktivität und des Mittelwertes der korrigierten Attraktivitäten  $D(w, b)$  im interaktiven Test







Fortsetzung zu Abbildung 6-4: Mittlere Abweichung der korrigierten Attraktivitäten  $V(w, b)$  im interaktiven Test

Betrachtet man die Differenz zwischen der präsentierten Attraktivität und dem Mittelwert der korrigierten Attraktivitäten  $D(w, b)$  in Abbildung 6-3, so fällt zunächst auf, daß die Korrekturen im Mittel gering ausfallen. Nur bei wenigen Objekten übersteigt die Korrektur den Wert  $|\lfloor 100 \cdot D(w, b) \rfloor| = 10$ , d.h. den tatsächlichen Wert  $|D(w, b)| = 0.1$ . Dies ist ein Zehntel der maximal möglichen Korrektur von  $|D(w, b)| = 1.0$ , d.h. selbst diese Änderung kann noch als gering angesehen werden.

Man bemerkt insgesamt die folgenden Punkte:

- In Dokument I1), I3) und I4) sind die Überschriften in der präsentierten Version zu niedrig bewertet und wurden daher von einigen Teilnehmern nach oben korrigiert. Bilder wurden im Gegensatz dazu offensichtlich von einigen Teilnehmern als weniger attraktiv eingeschätzt und daher nach unten korrigiert.

Dieser Effekt ist zum einen wohl durch das Maß für die strukturbedingt höhere Attraktivität von Bildern gegenüber Text  $d_{\text{Graphik}} = 0.5$  verursacht. Eine Verkleinerung des Wertes würde Text und damit Überschriften höher, Bilder dagegen niedriger bewerten.

Einen ähnlichen Effekt hätte das Verkleinern der Gewichte  $g_{\text{AreaMeas}}$  und  $g_{\text{BlackIntens}}$ . Allerdings würden dann auch die Unterschiede zwischen den Bildern untereinander geringer. Ein großes Bild wäre also nicht mehr attraktiver als ein kleines, und ein visuell dichtes nicht mehr attraktiver als eine Liniengraphik.

Da in Dokument I4) die Überschrift auch ohne gleichzeitige Anwesenheit eines Bildes nach oben korrigiert wurde, scheint das Problem aber nicht alleine durch  $d_{\text{Graphik}}$  verursacht zu sein. Die korrigierten Überschriften zeichnen sich durch die Layoutmerkmale Fettdruck (Dokument I1) und I4)) und Großschreibung (Dokument I3)) aus. Eine etwas höhere Gewichtung dieser Layoutmerkmale wäre also eventuell angebracht.

- Daß einige der Teilnehmer sowieso eine höhere Gewichtung von Fettdruck ( $g_{\text{Bold}}$ ) vorschlagen, ist in Dokument I3) in der ersten Zeile des Haupttextes zu erkennen. Hier wurden zwei Wörter in Fettdruck trotz einer bereits erhöhten Attraktivität noch stärker bewertet.

Die Korrektur der Attraktivität des fettgedruckten Literaturverweises im ersten Abschnitt von Dokument I2) ist übrigens in diesem Zusammenhang nicht so aussagekräftig, weil hier ein Klassifikationsfehler für Fettdruck vorliegt. Der Literaturverweis wurde fälschlicherweise nicht als fettgedruckt erkannt. Dieser Fehler wurde von einer Reihe von Teilnehmern entdeckt.

Betrachtet man nun die mittlere Abweichung der korrigierten Attraktivitäten  $V(w, b)$  (Abbildung 6-4), relativieren sich die obigen Schlußfolgerungen allerdings etwas. Offensichtlich ist die mittlere Abweichung bei allen Änderungen relativ hoch, die Testteilnehmer waren sich also bei den Korrekturen sehr uneinig. Tatsächlich ist bereits während der Durchführung des Tests aufgefallen, daß jeder Teilnehmer die Dokumente etwas anders sieht, und daher bei den selben Objekten unterschiedliche Korrekturen vorgenommen wurden, oder sowieso verschiedene Objekte korrigiert wurden. Dabei spielt auch eine Rolle, daß die Teilnehmer offensichtlich unterschiedlich genau beobachteten.



Ein gutes Beispiel dafür ist das linke Bild auf Dokument I5). Zwar ist  $D(w, b) = 0$ , aber  $V(w, b) \approx 0.04$ . Das heißt, daß mehrere Teilnehmer die Attraktivität dieses Objekts in genau entgegengesetzte Richtungen korrigiert haben. Nach mündlichen Aussagen der Teilnehmer während des Tests beruht dies auch auf verschiedenen wahrgenommenen Merkmalen. Die stärkere Irregularität des Bildes im Vergleich zu dem rechten Bild auf Dokument I5) sollte nach Meinung einiger Versuchsteilnehmer eine höhere Attraktivität des linken Bildes bewirken. Es gibt für die Irregularität eines Graphikblock im vorgestellten System bis jetzt noch kein Klassifikationsverfahren, dies sollte aber möglicherweise in Betracht gezogen werden.

Zusammenfassend kann man feststellen, daß der Test das bestehende Verfahren zur Attraktivitätsbewertung und die darin enthaltenen Gewichte im wesentlichen bestätigt, die berechnete Attraktivitätsbewertung der Wahrnehmung des Menschen also recht gut entspricht. Die oben diskutierten Änderungen, die durch die Testergebnisse angeregt werden, sollte man mit Hilfe einer größeren Anzahl von Testdokumenten untersuchen. Insbesondere eine genaue Zahlenangabe für die Änderung der Gewichte kann nicht aus den vorliegenden Ergebnissen abgelesen werden, sondern muß empirisch an einer größeren Anzahl von Dokumenten ermittelt werden.

Eine ergänzende Möglichkeit, das Verfahren zu verifizieren, wäre ein Test durch Blickverfolgung. Dabei würde man den Testpersonen nacheinander die verschiedenen Testdokumente im Original präsentieren. Mit Hilfe einer Augenkamera könnte man dann den Blick jeder Testperson verfolgen und dadurch die Fixationspunkte auf dem Dokument und die Reihenfolge ihrer Fixierung ermitteln. Eine Bestätigung des Verfahrens zur Attraktivitätsbewertung ergäbe sich aus diesem Test, wenn die Fixationspunkte mit überdurchschnittlich attraktiven Dokumentbereichen zusammenfallen würden, und die Reihenfolge der Fixierungen die Attraktivitätsbewertung widerspiegeln würde. Dies wäre allerdings nur eine grobe Überprüfung des Verfahrens und nicht mit dem in Lokalisierung und Bewertung sehr viel höher auflösenden interaktiven Test vergleichbar.

In ähnlichem Zusammenhang wurde Blickverfolgung bereits in [Egl98] benutzt, um das dortige, physiologisch begründete Verfahren zur zeit- und ortsvarianten Segmentierung von Dokumenten zu verifizieren.



# 7 Extraktion „attraktiver“ Information

Das eigentliche Ziel der bisher dargestellten Attraktivitätsbewertung von Dokumenten ist die Extraktion relevanter Information aus den Dokumenten. Nach der hier eingenommenen Sichtweise sind die relevanten Informationen gerade diejenigen, die in der texturbasierten Attraktivitätsanalyse als attraktiv bewertet wurden.

Daß diese Annahme tatsächlich zu einer sinnvollen Informationsextraktion führt, wird durch eine Programmergänzung demonstriert. Sie realisiert eine von mehreren möglichen Weiterverarbeitungen der Attraktivitätsbewertung eines Dokuments, indem sie *Schlüsselsätze des Textes* aus dem Dokument extrahiert. Dabei werden die Graphikobjekte auf dem Dokument nicht berücksichtigt.

Grundlage für die Extraktion ist die Attraktivitätsbewertung  $A'(w)$  für jedes Wort  $w \in WORDS$  und der daraus abgeleitete *Attraktivitätsmittelwert für jeden Textblock*  $b$   $A'(b)$

$$A'(b) = \frac{\sum_{w \in \{w: w > b \wedge w \in WORDS\}} A'(w)}{|\{w: w > b \wedge w \in WORDS\}|} \quad (7-1)$$

Dieser Attraktivitätsmittelwert ist ein mittelbares Merkmal des Textblock-Textons  $b$ , das sich aus den Merkmalen der Textons niedriger Ebene, nämlich aus den Attraktivitätsbewertungen der Wort-Textons  $w$ , ergibt. Die Bezeichnung entspricht der Bezeichnung der Attraktivitätsbewertung für Graphikblöcke; da aus dem Zusammenhang aber immer klar ist, ob Text- oder Graphikblöcke behandelt werden, besteht keine Verwechslungsgefahr.

Aus dem Histogramm der Attraktivitätsbewertungen der Wort-Textons im Text, kurz der Wortattraktivitäten, bestimmt man eine adaptive Entscheidungsschwelle, die attraktive, zu extrahierende Textons, bzw. Objekte, von nicht attraktiven trennt:

Sei  $1 \leq i, j \leq N_{bin}$ , dann ist das *Histogramm der Wortattraktivitäten im Text* gegeben durch

$$H(i) = \left| \left\{ w: w \in WORDS \wedge i = \lfloor N_{bin} \cdot A'(w) \rfloor \right\} \right| \quad (7-2)$$

und die abgeleitete *Entscheidungsschwelle* durch

$$Th = (\max\{i : \forall j : H(i) \geq H(j)\} + d) / N_{bin}, \quad (7-3)$$

mit den empirisch hergeleiteten Werten  $N_{bin} = 100$  und  $d = 10$ .

- Alle Wörter  $w$ , für die gilt  $w \in WORDS \wedge A'(w) > Th$  und
  - alle Textblöcke  $b$ , für die gilt  $b \in (Text-)BLOCKS \wedge A'(b) > Th$ ,
- werden entsprechend ihrer Attraktivitätsbewertung sortiert. Ausgehend von jedem dieser Objekte wird nun mit Hilfe eines OCR-Programms ein Schlüsselsatz, zum Teil auch mehrere Schlüsselsätze, extrahiert. Dabei beginnt man mit der Bearbeitung bei dem Objekt mit der höchsten Attraktivitätsbewertung.

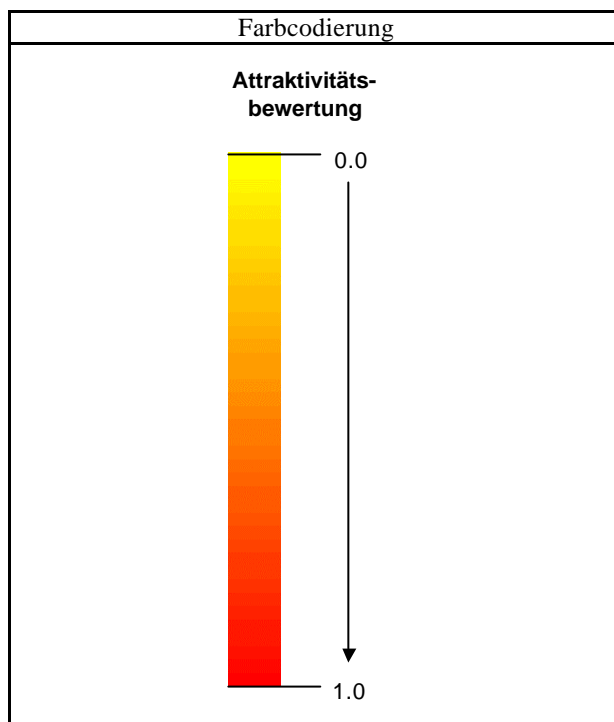
Ist dieses attraktive Objekt ein Wort, so wird der Satz, in dem sich das Wort befindet, ermittelt und als Schlüsselsatz am Monitor angezeigt. Sind ganze Textblöcke als attraktiv bewertet, so werden sie in ihrer Gesamtheit von der OCR übersetzt und angezeigt. Abbildung 7-1 zeigt zwei Beispiele.

Die Programmergänzung gibt einen guten Eindruck einer aufmerksamkeits- bzw. attraktivitätsbasierten Extraktion. Sie weist aber auch auf bestehende Probleme hin, die bei einer Vertiefung der Arbeit in Richtung Extraktion gelöst werden müßten:

- Enthält ein Satz mehrere attraktive Wörter, so entstehen zunächst mehrere identische Schlüsselsätze. Diese Duplikate werden zur Zeit aufgrund ihrer identischen räumlichen Lage im Dokument erkannt, und der Schlüsselsatz wird nur einmal, bei seinem ersten Auftreten, ausgegeben. Dieses Verfahren funktioniert nur für absolut identische Duplikate. Ist dagegen ein ermittelter Schlüsselsatz Teil eines ebenfalls extrahierten Blocks, funktioniert dieses Verfahren nicht mehr.
- Die Erkennungsleistung der OCR wird durch die Verwendung von sprachspezifischen Lexika verbessert. Die zu verwendende Sprache muß für die beschriebene Programmergänzung vom Benutzer über ein Menü eingestellt werden. Für eine vollständig automatische Informationsextraktion müßte auch die Sprache automatisch bestimmt werden.
- Die bisher realisierte Extraktion produziert eine Zusammenfassung, die dem Leser durch Hinweise auf den Textinhalt entscheiden hilft, ob er das Originaldokument lesen soll oder nicht. Eine schwierigere Aufgabe ist dagegen die Erstellung einer Zusammenfassung, die so ausgewogen, vollständig und zusammenhängend ist, daß der Leser das Originaldokument gar nicht mehr lesen muß ([Pai90]). Die Probleme Ausgewogenheit, Vollständigkeit und Zusammenhang sind in dieser Arbeit noch völlig außer Acht gelassen worden, und es muß vermutet werden, daß sie alleine auf Grundlage der Attraktivitätsbewertung eines Dokuments nicht gelöst werden können.

| Attraktivitätsbewertung auf Testdokument T4)   | Extrahierte Schlüsselsätze  |
|--|---|
|  <p><b>Verfahren und Systeme der Computertomographie</b><br/> <b>Röntgen - Magnetische Resonanz - Ultraschall</b><br/> <b>E. Hacht und G. Schwierz</b><br/>     Springer AG, München</p> <p><b>Zusammenfassung:</b> Die bildgebenden Systeme der medizinischen Diagnostik erfahren zur Zeit eine sehr rasche und gleichzeitige Entwicklung, durch die Digitalisierung der Meßsignale und die rechnergesteuerte, elektronische Verarbeitung, Röntgen, Magnetische Resonanz und Ultraschall ermöglichen drei Systeme der Computertomographie, die Einzelbilder des zu untersuchenden Objektes mit hoher Qualität, Auflösung und Aussagekraft liefern. In diesem Bereich werden die physikalischen Prinzipien, die Algorithmen zur Rekonstruktion, die Meßsysteme sowie der klinische praktische Aufbau erläutert.</p> <p><b>Schlüsselwörter:</b> Bildgebende Verfahren, Computertomographie, Magnetische Resonanz, Medizinische Diagnostik</p> <p><b>Summary:</b> Imaging systems of medical diagnosis are developing very rapidly and especially which is caused by the operations of measuring signals and computerized processed signals. X-ray, magnetic resonance and ultrasound enable three systems of computerized tomography which produce excellent views of the object being of high quality, high resolution and high diagnostic power. The paper discusses the physical principles, the algorithms for reconstruction, the measuring systems and the processing performance.</p> <p><b>Key words:</b> Imaging systems, Computerized tomography, Image reconstruction, Medical diagnosis</p> <p>Computing Review Classification: J.3</p> <p><b>1. Einleitung</b><br/>     Die bildgebenden Systeme der medizinischen Diagnostik erfahren zur Zeit eine sehr rasche und gleichzeitige Entwicklung. Einem derart raschen physikalischen Pro-</p> | <p><b>SCHLÜSSELSÄTZE:</b><br/>     *****<br/> <b>Verfahren und Systeme der Computertomographie</b><br/> <b>Röntgen - Magnetische Resonanz - Ultraschall</b><br/>     *****<br/> <b>1. Einleitung</b><br/>     *****<br/> <b>Informaik-Spektrum (1985)8: 273282</b><br/>     *****<br/> <b>E. Hacht und G. Schwierz</b><br/> <b>Siemens AG, München</b><br/>     *****<br/>     ©<br/>     *****<br/> <b>Computing Reviews Classification: J. 3</b><br/>     *****<br/> <b>Key words: Imaging Systems, Computerized tomography, Image reconstruction, Medical diagnosis</b><br/>     *****<br/> <b>Zusammenfassung,</b><br/>     *****<br/> <b>Summary,</b><br/>     *****<br/> <b>Schlüsselwörter: Bildgebende Verfahren, Computertomographie, Bildrekonstruktion, Medizinische Diagnostik</b><br/>     *****<br/> <b>Die Röntgen-Computer-tomographie war das erste System der bildgebenden medizinischen Technik, bei dem konsequent alle Meßwerte digitalisiert und rein digital weiterverarbeitet werden.</b><br/>     *****<br/> <b>Tomographie bedeutet Abbildung einer dünnen Schicht des dreidimensionalen Körpers.</b></p>   |
| Attraktivitätsbewertung auf Testdokument T7)   | Extrahierte Schlüsselsätze  |
|  <p><b>Object Orientation in Heterogeneous Distributed Computing Systems</b><br/> <b>John R. Nicol, C. Thomas Wilkes, and Frank A. Manola</b><br/>     GTE Laboratories Inc.</p> <p><b>A least-common-denominator approach to object orientation is emerging as a key strategy for flexibly coordinating and integrating networked information processing resources.</b></p> <p>June 1993</p>   | <p><b>SCHLÜSSELSÄTZE:</b><br/>     *****<br/> <b>Object Orientation in Heterogeneous Distributed Computing Systems</b><br/>     *****<br/> <b>John R. Nicol, C. Thomas Wilkes, and Frank A. Manola</b><br/>     GTE Laboratories Inc.<br/>     *****<br/> <b>008-9 162193/0600-0057803.00 © 1993 IEEE</b><br/>     *****<br/>     C<br/>     *****<br/> <b>A least-common denominator approach to object orientation is emerging as a key strategy for flexibly coordinating and integrating networked information processing resources.</b><br/>     *****<br/> <b>C omputer usage is spreading into all functions of large organizations, as can be seen in the widespread appearance of powerful PCs for desktop use and powerful workstations for more specialized use. Because computers are becoming commonplace in routine business functions, future information-processing environments will likely consist of vast networks of heterogeneous, autonomous, and distributed computing resources, including computers (from mainframe to personal), information-intensive applications, and data (fil and databases). The telecommunications infrastructure provides a clear example of such an environment.</b><br/>     *****<br/> <b>There is a growing need for technology to flexibly coordinate integrating computing resources to create networkwide integrated distributed systems that address the changing and heterogeneous resource requirements. In some cases, these systems might combine resources belonging to different individuals or organizations (for example, resources between networked university systems and regular college courses).</b><br/>     *****<br/> <b>The first phase in creating integrated, large-scale, distributed information systems is interconnectivity. Large, centrally used computers have made it difficult to coordinate with other systems. Two or more computing resources are interconnected if they can exchange messages. However, common security guarantees may be unobtainable, not negotiable.</b><br/>     *****<br/> <b>For interconnectivity, the client-server model is the most pervasive.</b><br/>     *****<br/> <b>There is a growing need for technology to flexibly coordinate distributed computing resources to create networkwide integrated distributed systems that address challenging new information processing requirements. In some cases, these systems might combine resources belonging to different individuals or organizations (for example, resources between networked university systems and regular college courses).</b><br/>     *****<br/> <b>The first phase in creating integrated, large-scale, distributed information systems is interconnectivity.</b></p> |

Abbildung 7-1: Attraktivitätsbewertung der Testdokumente T4) und T7) und das Ergebnis der davon ausgehenden Extraktion von Schlüsselsätzen



Fortsetzung zu Abbildung 7-1: Farbcodierung der Attraktivitätsbewertung der Testdokumente T4) und T7)

### Thematisch verwandte Arbeiten

Verfahren zur Informationsextraktion aus Dokumenten lassen sich abhängig von der Art der extrahierten Information in zwei Gruppen einteilen, nämlich in

- Verfahren zur Extraktion des *Dokumenttyps* und
- Verfahren zur Extraktion des *Dokumentinhaltes*.

Bei der Bestimmung des Dokumenttyps ergibt sich eine grobe Einordnung des Dokuments, wie durch folgende Beispiele verdeutlicht:

1. Auf Briefumschlägen treten die bei der Postsortierung zu lesenden Adreßfelder je nach Absender in verschiedenen Anordnungen auf. Jede Anordnung entspricht einem Dokumenttyp, der durch Modellierung des Umschlag-Layouts identifiziert werden kann. Walischewski ([Wal97]) repräsentiert das Layout durch einen gerichteten Graphen, der die Lage und Eigenschaften der einzelnen visuellen Objekte auf dem Umschlag enthält.
2. Bei der Verteilung von geöffneten Geschäftsbriefen gehen Briefe eines einzelnen Absenders oftmals an einen einzelnen zuständigen Sachbearbeiter innerhalb des Unternehmens. Der Absender eines Dokuments ist am Dokumenttyp erkennbar, wobei der Dokumenttyp bestimmt wird durch die Anordnung von Adreßfeldern und Logos auf den Briefen. Ein entsprechendes System ist in [Brü96] beschrieben, das dort benutzte Verfahren zur Erkennung von Logos in [Sud97].
3. So wie es im Posteingang einer Firma je nach Absender verschiedene Typen von Briefen gibt, so gibt es je nach Verlag in einer Dokumentdatenbank ebenfalls verschiedene Dokumenttypen. Nagy ([Nag92]) beschreibt ein System, daß die Typen mit Hilfe einer Blockgrammatik des Layouts identifiziert und zur sinnvollen Darstellung des Dokuments beim Browsing benutzt.

4. In seiner Übersicht zum Thema „Retrieval von Dokumenten aus Datenbanken“ gibt Doermann ([Doe97]) in Abschnitt 4 weitere Beispiele zur Extraktion des Dokumenttyps aus der Layout-Struktur des Dokuments.
5. Während der Dokumenttyp in den bisher erwähnten Arbeiten im wesentlichen durch die geometrische Anordnung der einzelnen Dokumentbereiche bestimmt war, ist bei Cullen, Hull und Hart in [Cul97] die Textur eines Dokuments verantwortlich für seinen Typ, der zum Suchen innerhalb einer Dokumentdatenbank dient.

Bei der Extraktion des *Dokumentinhalts* geht es darum, detaillierte Information zu gewinnen, also z.B. Schlüsselwörter im Text zu finden, oder eine Zusammenfassung des Dokuments zu erzeugen. Dabei sind in der Vergangenheit unterschiedliche Verfahren entstanden, je nachdem ob das Dokument in elektronischer Form oder als Papierdokument vorliegt:

1. Zur Extraktion inhaltlicher Information aus fehlerfreiem, elektronischem Text, also z.B. aus einer Folge von ASCII-Zeichen, sind unter dem Schlagwort „Information Retrieval“ eine Reihe von Methoden entstanden. Eine Form des „Information Retrieval“ ist das Generieren von Zusammenfassungen. In [Pai90] nennt Paice sieben Methoden zum Auswählen relevanter Wörter und Sätze aus einem Text:
  - „Frequency-keyword approach“: Nach dem Aussondern von unbedeutenden Füllwörtern werden die restlichen Wörter nach der Häufigkeit ihres Auftretens sortiert. Schlüsselwörter sind diejenigen, die am häufigsten auftreten. Sie markieren relevante Sätze.
  - „Title-keyword approach“: Wie beim „frequency-keyword approach“ werden Schlüsselwörter benutzt, um wichtige Sätze zu markieren. Allerdings werden die Schlüsselwörter nach dem Aussondern von Füllwörtern nur aus den Überschriften des Dokuments extrahiert.
  - „Location method“: Erste und letzte Sätze des Dokuments und jedes Absatzes sind relevanter als andere.
  - „Syntactic criteria“: Ein vom Autor als wenig erfolgversprechend beurteilter Ansatz, der davon ausgeht, daß die grammatikalische Struktur eines Satzes mit seiner Relevanz verknüpft ist.
  - „Cue method“: Diese Methode benutzt Indikatorworte wie „greatest“ oder „significant“, die in wesentlichen Sätzen enthalten sein sollen.
  - „Indicator-phrase method“: Es werden ähnlich wie bei der „cue method“ Indikatoren benutzt, nun allerdings ganze Ausdrücke wie „The main aim of the present paper is to describe...“.
  - „Relational criteria“: Die Grundidee dieser Methode ist, daß die wichtigsten Sätze im Dokument diejenigen sind, die mit der größten Anzahl anderer Sätze verknüpft sind. Eine Variante des „Frequency-keyword approach“ wird von Salton, Allan, Buckley und Singhal ([Sal94]) eingeführt, um erfolgreich u.a. Zusammenfassungen von Texten zu erzeugen.
2. Die für elektronische Texte entwickelten Methoden kann man nur begrenzt anwenden, wenn Information aus einem Papierdokument extrahiert werden soll. Aus dem eingescannten Bild des Dokumentes kann zwar mit Hilfe eines OCR-Programmes ein elektronischer Text erzeugt werden, allerdings sind die Ergebnisse dieser relativ zeitintensiven Übersetzung immer fehlerbehaftet. So kann bei der Erstellung von Zusammenfassungen mit fehlerhaften OCR-Ergebnissen weder eine korrekte Analyse noch eine gute Lesbarkeit des

Resultats gewährleistet werden. Doermann erwähnt in Kapitel 2 seines Überblicks [Doe97] lediglich eine Methode, die aus fehlerbehafteten OCR-Texten Indizes für eine Datenbank erzeugen kann, aber eben keine Zusammenfassung des Textes.

Da OCR also fehlerbehaftet und vergleichsweise langwierig ist, gab es Ansätze, die Form des Dokuments zur Extraktion aus Papierdokumenten auszunutzen. In Kapitel 3 des Überblicks [Doe97] beschreibt Doermann zunächst zwei derartige Methoden zur Extraktion, bzw. zum Lokalisieren von Schlüsselwörtern:

- „Characterizing Text“ ([Doe97], Kapitel 3.1): Wörter sind Schlüsselwörter, abhängig davon ob sie großgeschrieben sind oder nicht, abhängig von ihrer Lage im Satz, von ihrer Länge und der Länge und syntaktischen Kategorie der vorhergehenden und nachfolgenden Wörter.
- „Keyword Spotting“ ([Doe97], Kapitel 3.2): Vorgegebene Schlüsselwörter werden durch ihre Kontur oder durch ihren „shape code“ repräsentiert und in dieser Form mit den Wörtern der zu untersuchenden Dokumente verglichen. Der „shape code“ entsteht, indem die einzelnen Buchstaben aufgrund ihrer Struktur (z.B. mit oder ohne Unter- und Oberlängen) auf eine kleine Menge von Symbolen abgebildet werden.

[Doe97], Kapitel 3.3, beschreibt außerdem, wie Zusammenfassungen aus einer Modellierung des Textbildes generiert werden können („Automated Abstracting“). Chen und Bloomberg ([Che97a]) werden mit ihrer Arbeit zitiert, die auf Ebene des Dokumentbildes die Extraktion aus elektronischem Text nachbildet. Ohne OCR wird im Text mit dem vorherrschenden Font die Häufigkeit von Wörtern durch Vergleich von Wort-Bildern gemessen. Dieses Merkmal wird zusammen mit Wortbreite, Lage und Leserichtung zur Bewertung und Extraktion von Sätzen benutzt.

Die oben entwickelte aufmerksamkeitsbasierte Extraktion des Dokumentinhalts hat diesen Verfahren gegenüber einige Vorteile:

- Die OCR wird zum spätestmöglichen Zeitpunkt eingesetzt, der Einfluß von Fehlern durch die OCR ist also minimiert.
- Das Verfahren ist im Bereich lateinischer Schrift unabhängig von der verwendeten Sprache, ganz im Gegenteil zu den obigen Verfahren „Syntactic criteria“, „Cue method“ oder „Indicator-phrase method“.
- Das aufmerksamkeitsbasierte Verfahren berücksichtigt zur Extraktion von Schlüsselsätzen nicht nur den Text im vorherrschenden Font, wie das an sich leistungsfähige Verfahren von Chen und Bloomberg, da wichtige Information oftmals gerade in Dokumentbereichen mit nicht vorherrschendem Font, z.B. in Überschriften, steckt.
- Das Verfahren könnte leicht so modifiziert werden, daß statt Schlüsselsätzen Schlüsselwörter extrahiert werden. Dabei wäre das hier vorgestellte Verfahren deutlich flexibler als die oben genannten, da z.B. Schlüsselwörter nicht wie beim „Keyword Spotting“ vorgegeben werden müssen oder nur durch restriktive Regeln wie bei der „Characterizing Text“-Methode ausgewählt werden können.



# 8 Zusammenfassung und Ausblick

Das in dieser Arbeit vorgestellte aufmerksamkeitsbasierte Verfahren lokalisiert relevante Information auf einem Dokument und bewertet die Relevanz jeder Information im Vergleich zu den anderen Informationen auf dem Dokument, wobei dies rein auf der Bildinformation des Dokuments geschieht.

Ausgangspunkt der Entwicklung war eine neue Sichtweise auf Dokumente, in der die Erzeugung eines Dokuments durch den Autor und die Verarbeitung der darin enthaltenen Information durch den Leser im Mittelpunkt steht. Zwei psychologisch motivierte Sender-Empfänger-Modelle der Informationsübertragung zwischen zwei Menschen erklärten deshalb zunächst, wie Autor und Leser eines Dokuments bei der Informationsübertragung zusammenwirken, und wie die Begriffe Aufmerksamkeit und Attraktivität mit der Relevanz von Dokumentbereichen verknüpft sind.

Nach diesen Modellen ist eine Information in einem Dokument relevant, wenn der Autor des Dokuments sie als wichtig erachtet. Es wurde angenommen, daß der Autor Informationen, die er für wichtig hält, beim Erstellen des Dokuments besonders kennzeichnet. Er weist einzelnen Dokumentbereichen unterschiedliche Attraktivität zu, indem er sie mit verschiedenen gestalterischen Merkmalen, den sogenannten Layoutmerkmalen, markiert. Diese Layoutmerkmale ziehen die Aufmerksamkeit des Informationsempfängers (Lesers) auf den ersten Blick an und ermöglichen ihm so eine effiziente Informationsextraktion.

Um eine effiziente Informationsextraktion auch im Rahmen der automatischen Dokumentanalyse zu ermöglichen, hat die vorliegende Arbeit nun die wahrgenommene Attraktivität von Dokumentbereichen modelliert. Diese Modellierung hätte von der Senderseite oder von der Empfängerseite her durchgeführt werden können. Die Arbeit hat die Modellierung aus dem Blickwinkel des Informationsempfängers durchgeführt, da erst auf dieser Seite eine Attraktivität wahrgenommen wird, die die Aufmerksamkeit des Empfängers steuert. Außerdem ist zu bedenken, daß man bei der heutigen weiten Verbreitung von elektronischen Textverarbeitungssystemen auch nicht mehr davon ausgehen kann, daß nur ausgebildete Setzer und Designer Dokumente gestalten, man also auf Senderseite feste, abfragbare Regeln zur Modellierung ausnützen könnte.

Grundlegend für die Entwicklung einer Attraktivitätsbewertung für die einzelnen Objekte auf einem Dokument war deshalb eine Befragung von Testpersonen, in der untersucht wurde, welche Layoutmerkmale Leser wie stark wahrnehmen. Zum einen wurde dadurch verifiziert, bezüglich welcher Layoutmerkmale die Dokumentbereiche klassifiziert werden müssen. Das waren für die Textobjekte im Dokument Einrückungen, Bullets, Fettdruck, Kursivdruck, Großschreibung, Freistellung und Schriftgröße und für Graphikobjekte Freistellung, Schwärzungsgrad und ein Flächenmaß. Obwohl in der Befragung noch weitere wichtige Layoutmerkmale wie Farbe, Einrahmung und Unterstreichung vorgeschlagen wurden, wurden zunächst Klassifikationsverfahren nur für die genannten Layoutmerkmale realisiert, da sie am häufigsten in dem in dieser Arbeit vorwiegend untersuchten Dokumenttyp der wissenschaftlichen Veröffentlichungen vorkommen. Für eine erweiterte, noch flexiblere Version der Attraktivitätsbewertung wäre es aber sinnvoll, weitere Layoutmerkmale miteinzubeziehen, insbe-

sondere auch ein Merkmal, das die Irregularität der inneren Struktur eines Graphikobjekts beschreibt. Aus der Befragung ergab sich zum anderen eine initiale Schätzung für die Gewichte, mit denen die normierten Ergebnisse der Klassifikationen in einer gewichteten Summe verknüpft werden, um eine Attraktivitätsbewertung zu erhalten.

Da man davon ausgeht, daß ein Mensch Text auf den ersten Blick als eine Ansammlung von verschiedenen Texturen wahrnimmt, wurde die Attraktivitätsanalyse mit einer texturbasierten Methode realisiert. Dazu wurde ein von der Juleszschen Texton Theorie inspirierter Texturbegriff verwendet, der Texturen als Anhäufungen von diskreten Wahrnehmungseinheiten, den Textons, beschreibt. Für eine technische Umsetzung dieser psychologisch fundierten Beschreibung wurden die Begriffe Textur und Texton formalisiert, wobei eine Ähnlichkeits- und eine Nachbarschaftsrelation die Beziehungen zwischen den Textons einer Textur beschreiben. In einer Weiterentwicklung der Juleszschen Vorstellung wurde außerdem eine Texturhierarchie eingeführt, die sich speziell dafür eignet, die Verwobenheit von Texturen auf Dokumenten zu beschreiben. Auch der Begriff der Bewertung wurde an dieser Stelle formal definiert. Formal definiert wurden außerdem die Begriffe Segmentierung und Resegmentierung. Eine Segmentierung ist die Zerlegung des Blickfeldes in Texturen, eine Resegmentierung ist die erneute Segmentierung eines bereits segmentierten Bildes, wobei allerdings ausschließlich bereits erzielte Ergebnisse verändert werden, um mehr Details herauszuarbeiten.

Im entwickelten technischen Verfahren entstehen in einem Segmentierungsprozeß aus den Strichen auf einem Dokument Buchstaben, Wörter, Zeilen und Blöcke, d.h. der Segmentierungsprozeß bildet auf jeder Stufe der Texturhierarchie Texturen. Diejenigen Teile der Texturen, die eines oder mehrere Layoutmerkmale tragen, werden dann durch eine Resegmentierung der bestehenden Segmentierung identifiziert. Attraktive Dokumentbereiche, und damit implizit auch relevante Informationen, sind so lokalisiert.

Die Bewertung der Attraktivität, die implizit eine Bewertung der Relevanz ist, wird dann aufbauend auf dieser Lokalisierung berechnet. Durch die Lokalisierung ist nämlich jedes Wort und jeder Graphikblock auf dem Dokument bezüglich der Layoutmerkmale klassifiziert. Die Zugehörigkeit zu einer Klasse wird dabei durch einen Zahlenwert repräsentiert, der zunächst für jedes Layoutmerkmal einzeln normiert und gewichtet wird. Dadurch ergeben sich Bewertungen bezüglich der einzelnen Layoutmerkmale, die dann für jedes Wort und jeden Graphikblock addiert werden, so daß die endgültige Attraktivitätsbewertung entsteht.

Um das Ergebnis der Attraktivitätsbewertung zu testen, wurde eine Untersuchung durchgeführt, bei der Testpersonen ihre Wahrnehmung von Attraktivität auf einigen Testdokumenten mit der berechneten Attraktivitätsbewertung vergleichen und dort etwaige Unterschiede korrigieren konnten. Dieser Test des Attraktivitätsmaßes zeigte, daß das aktuelle Verfahren sehr plausible Ergebnisse für eine Attraktivitätsbewertung liefert. Bei einer Überarbeitung wäre allerdings die Bewertung von Graphik im Vergleich zu Text noch einmal zu überdenken, da Graphiken zur Zeit offensichtlich etwas zu hoch bewertet werden.

Das Verfahren zur Attraktivitätsbewertung wurde außerdem auch durch die Programmergänzung zur Extraktion von Schlüsselsätzen bestätigt. Die aufgrund der Attraktivitätsbewertungen aus dem Dokument extrahierten Sätze ergeben nämlich sinnvolle Zusammenfassungen des Dokumentinhalts, die einem Leser entscheiden helfen können, ob er das gesamte Dokument lesen sollte.

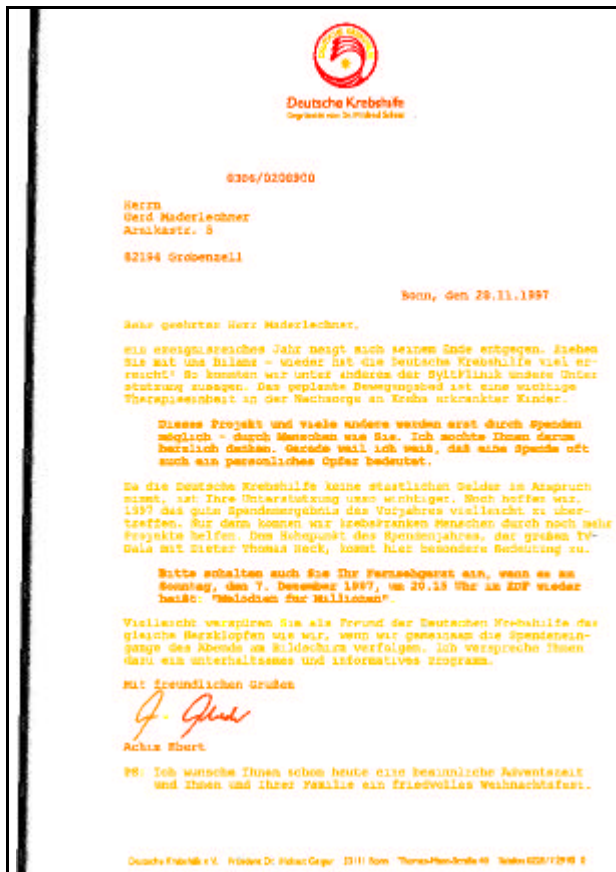
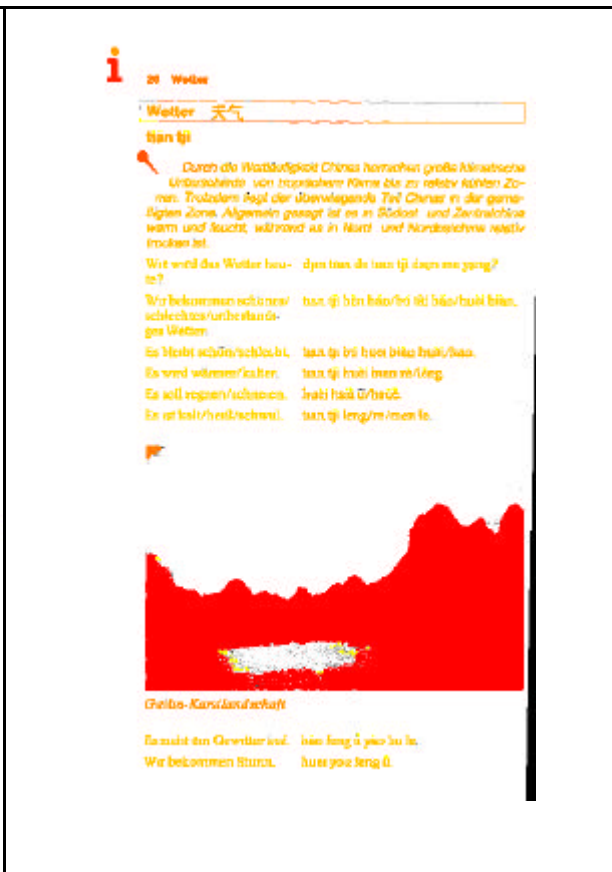
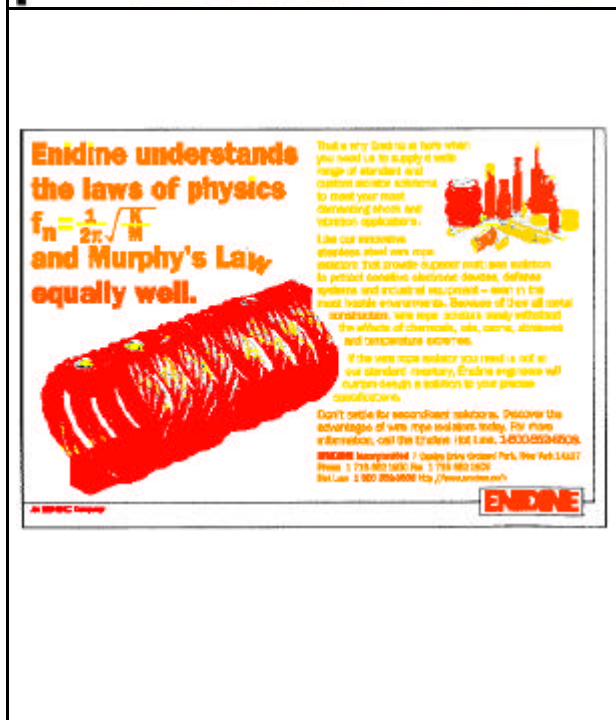
|  |  |
|--|--|
|  <p>Logo: Deutsche Krebshilfe<br/>Copyright von Dr. Hubert Geyer</p> <p>8306/0206900</p> <p>Herrn<br/>Heidi Maderlechner<br/>Amnikaserstr. 5<br/>82194 Grobenzell</p> <p>Bonn, den 28.11.1997</p> <p>Lieber geehrtes Herr Maderlechner,</p> <p>ein ereignisreiches Jahr neigt sich seinem Ende entgegen. Stehen Sie mit uns Bilanz - wieder hat die Deutsche Krebshilfe viel erreicht! So konnten wir unter anderem das GUT-Klinik unsere Unterstützung zusagen. Das geplante Bewegungsbad ist eine wichtige Therapieeinheit in der Nachsorge an Krebs erkrankter Kinder.</p> <p>Dieses Projekt und viele andere wurden erst durch Spenden möglich - durch Menschen wie Sie. Ich möchte Ihnen darum herzlich danken. Gerade weil ich weiß, daß eine Spende oft auch ein persönliches Opfer bedeutet.</p> <p>Da die Deutsche Krebshilfe keine staatlichen Gelder in Anspruch nimmt, ist Ihre Unterstützung umso wichtiger. Noch hoffe ich, 1997 das gute Spendenverhältnis des Vorjahres vielleicht zu übertrafen. Nur dann können wir lebensbranken Menschen durch noch mehr Projekte helfen. Dem Höhepunkt des Spendenjahres, der großen TV-Gala mit Dieter Thomae Heck, kommt hier besondere Bedeutung zu.</p> <p>Bitte erhalten auch Sie Ihr Fernsehgerät am, wenn es am Sonntag, den 7. Dezember 1997, um 20.15 Uhr im ZDF wieder heißt: "Mach dich für Millionen".</p> <p>Vielleicht veripiren Sie als Freund der Deutschen Krebshilfe das gleiche Herzklopfen wie wir, wenn wir gemeinsam die Spendenkampagne des Abends am Bildschirm verfolgen. Ich verspreche Ihnen dazu ein unterhaltsames und informatives Programm.</p> <p>Mit freundlichen Grüßen</p> <p><i>J. Geyer</i><br/>Hubert Geyer</p> <p>PS: Ich wünsche Ihnen schon heute eine besinnliche Adventszeit und Ihnen und Ihrer Familie ein friedvolles Weihnachtsfest.</p> <p>Deutsche Krebshilfe e.V. Präsident Dr. Hubert Geyer   53111 Bonn   Thomas-Meyer-Straße 49   Telefon 0228/12979 8</p> |  <p>20 Wetter</p> <p>Wetter 天气</p> <p>Nian ti</p> <p>Durch die Vielfaltigkeit Chinas herrschen große klimatische Unterschiede von tropischem Klima bis zu relativ kaltem Zonen. Trotzdem liegt der überwiegende Teil Chinas in der gemäßigten Zone. Allgemein gesagt ist es in Südost- und Zentralchina warm und feucht, während es in Nord- und Nordostchina relativ trocken ist.</p> <p>Wie wird das Wetter sein? - Yin tian de tian ti zhe mo gang? ta?</p> <p>Wir bekommen schönes/ schlechtes/überstandiges Wetter</p> <p>Es bleibt schön/schlecht. - tian ti bi hui bi ke hui/kuo.</p> <p>Es wird wärmer/kälter. - tian ti hui ren re/heng.</p> <p>Es soll regnen/schneien. - hai hai yu/hu.</p> <p>Es ist kalt/hot/schneit. - tian ti leng/wen le.</p> <p>Gelbes-Karstlandschaft</p> <p>Es macht dem Wetter kalt. - hui lang ti yao lu lu.</p> <p>Wir bekommen Sturm. - hui yao leng lu.</p> |
|  <p>ENIDINE understands the laws of physics</p> <p><math>f_n = \frac{1}{2\pi} \sqrt{\frac{K}{M}}</math></p> <p>and Murphy's Law equally well.</p> <p>That's why ENIDINE is here when you need us to supply a wide range of standard and custom wire solutions to meet your most demanding flow and vibration applications.</p> <p>Like our innovative stainless steel wire rope solutions that provide superior rust-resistance to protect sensitive electronic devices, defense systems and critical equipment - even in the most hostile environments. Because of the all metal construction, wire rope solutions easily withstand the effects of chemicals, oils, acids, solvents and temperature extremes.</p> <p>If the wire rope solution you need is not in our standard inventory, ENIDINE engineers will custom-design a solution to your precise specifications.</p> <p>Don't settle for second-rate solutions. Discover the advantage of wire rope solutions today. For more information, call the ENIDINE Hot Line, 1-800-853-4806.</p> <p>ENIDINE Incorporated   2000 First Street Park, New York, NY 10027<br/>Phone: 1-718-607-6000 Fax: 1-718-607-6000<br/>Hot Line: 1-800-853-4806 http://www.enidine.com</p> <p>ENIDINE</p>   | <p>Farbcodierung</p> <p>Attraktivitätsbewertung</p>  <p>0.0</p> <p>1.0</p>   |

Abbildung 8-1: Flexibilität des Verfahrens zur Attraktivitätsbewertung, demonstriert an drei verschiedenen Dokumenttypen

Dabei ist das eingeführte Verfahren zur Lokalisierung und Bewertung relevanter Information sehr allgemein. Obwohl es anhand von wissenschaftlichen Veröffentlichungen entwickelt wurde, ist es

- nicht auf einen bestimmten Dokumenttyp und
- nicht auf eine bestimmte Sprache beschränkt.

Die Beispieldokumente in den vorigen Kapiteln hatten zwar alle den gleichen Dokumenttyp, aber die drei verschiedenen Sprachen Deutsch, Englisch und Finnisch. Abbildung 8-1 zeigt nun noch drei weitere Dokumenttypen, nämlich einen Geschäftsbrief, einen Ausschnitt aus einem Reisewörterbuch und eine Anzeige. Der Ausschnitt aus dem Wörterbuch ist besonders interessant, weil hier eine weitere „Sprache“, nämlich eine deutsche Lautschrift des Chinesischen, auftaucht.

Eine Einschränkung des Verfahrens ist lediglich durch Dokumente gegeben, bei denen Information ausschließlich durch den Einsatz bewußter Aufmerksamkeit zugänglich wird (Abbildung 8-2). Dies ist der Fall, wenn

- relevante Dokumentbereiche nicht visuell markiert werden, wie in Erzählungen oder Romanen, oder wenn
- bei der Erzeugung des Dokuments versucht wurde, alle Dokumentbereiche visuell als relevant zu markieren, so daß alle Dokumentbereiche eine hohe, aber sehr ähnliche Attraktivitätsbewertung haben, wie in den „Gelben Seiten“.

Bei derartigen Dokumenten gelingt es auch einem Betrachter nicht, auf den ersten Blick zu erkennen, in welchem Dokumentbereich relevante Information zu finden ist. Eine Lokalisierung, Bewertung und damit auch Extraktion relevanter Information ist dann nur durch eine inhaltliche Analyse des Dokuments möglich.

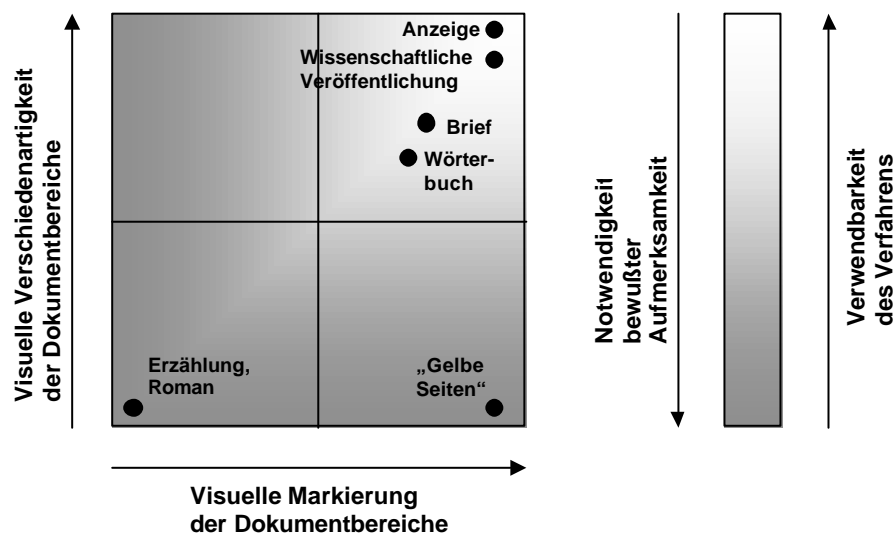


Abbildung 8-2: Verwendbarkeit des Verfahrens abhängig von der notwendigen bewußten Aufmerksamkeit: bewußte Aufmerksamkeit ist zur Extraktion relevanter Information umso notwendiger, je gleichartiger, bzw. je weniger visuell markiert Dokumentbereiche sind.

Es soll außerdem noch einmal betont werden, daß das Verfahren in keiner Weise auf Papierdokumente beschränkt ist, sondern bei entsprechender Ergänzung der untersuchten Layoutmerkmale und Änderung der Klassifikationsverfahren auch für elektronische Dokumente wie E-Mails und WWW-Seiten geeignet ist. In diesem Fall sind selbstverständlich die texturbasierten Klassifikationsverfahren für Layoutmerkmale nicht mehr nötig, da in elektronischen Dokumenten die Layoutmerkmale direkt zugänglich sind. Gegebenenfalls müßten aber weite-

re Layoutmerkmale berücksichtigt werden. Auf WWW-Seiten findet man häufig Bewegung als ein Layoutmerkmal, das die Aufmerksamkeit des Lesers anziehen soll. Außerdem spielt die Farbgestaltung eine wichtige Rolle. Erkenntnisse zur Attraktivität verschiedener Farben müßten berücksichtigt werden, oder Versuche dazu durchgeführt werden.

Aufgrund der großen Flexibilität ergibt sich eine Vielzahl von Anwendungsmöglichkeiten des Verfahrens, insbesondere in Dokumentdatenbanken mit sehr unterschiedlichen Dokumenten:

- Die Verteilung attraktiver Bereiche auf den Dokumenten kann als Dokumenttyp interpretiert werden. Entsprechend ihrer Attraktivitätsverteilungen könnte man die Dokumente einer Datenbank in Gruppen ähnlichen Typs vorsortieren. Die so entstehende Strukturierung der Datenbank hilft einem Benutzer beim Durchsuchen der Datenbank, wenn er bereits eine Vorstellung vom Typ des gesuchten Dokuments hat.
- Die Attraktivitätsbewertung eines Dokuments bietet eine weitere Möglichkeit, einen Benutzer beim Durchsuchen einer Dokumentdatenbank zu unterstützen, indem man ihm auf dem Monitor nicht immer das ganze Dokument, sondern nur dessen relevanteste Teile präsentiert. Der Benutzer kann so schneller und bequemer entscheiden, ob das fragliche Dokument für ihn interessant ist oder nicht. Nur wenn er sich näher für das Dokument interessiert, wird es auch komplett auf dem Monitor angezeigt.

Diese Vorgehensweise ist insbesondere auch dann sinnvoll, wenn der Benutzer über ein Netz auf die Dokumentdatenbank zugreift. In diesem Fall muß immer nur ein kleiner Teil des ganzen Dokuments über das Netz übertragen werden. Das erhöht für den Benutzer die Zugriffsgeschwindigkeit auf die Datenbank und entlastet das Netz durch eine verringerte Datenübertragung.

- Wenn die Attraktivitätsbewertung als Grundlage einer tatsächlichen Informationsextraktion mittels OCR benutzt wird, bieten sich eine Reihe weiterer Möglichkeiten. So können aus den attraktivsten Dokumentteilen mittels OCR Schlüsselwörter gewonnen werden, die sowohl zum Indexieren des Dokuments innerhalb der Datenbank, als auch zum Suchen in einer nicht indexierten Datenbank benutzt werden können.
- Wird an die Attraktivitätsbewertung eine OCR gekoppelt, können auch Zusammenfassungen von Dokumenten erstellt werden. Eine Möglichkeit zur Extraktion von Sätzen wurde in Kapitel 7 demonstriert. Dort werden nur attraktive Dokumentteile einer OCR zugeführt. Die vorgestellte Methode könnte man weiterentwickeln, so daß eine so gute Zusammenfassung des Dokuments entsteht, daß der Leser das Originaldokument gar nicht mehr lesen muß. Allerdings bestehen dabei Probleme bezüglich Ausgewogenheit, Vollständigkeit und Zusammenhang der Zusammenfassung, und wahrscheinlich können diese Probleme alleine auf Grundlage der Attraktivitätsbewertung eines Dokuments nicht gelöst werden. In Zukunft ist also gegebenenfalls die Lesereihenfolge als nicht mehr rein aufmerksamkeits- und texturbasierter Vorgang mit einzubeziehen.

Eine völlig andere Anwendungsmöglichkeit des Verfahrens bietet sich in der Werbebranche. Hier wird bereits jetzt die Gestaltung von neuen Anzeigen, Werbebroschüren, Katalogen, TV-Spots und Produktverpackungen daraufhin untersucht, ob sie dem Betrachter die gewünschte Information optimal vermittelt. Die Gesellschaft für angewandtes Direktmarketing (GAD) legt dazu die zu untersuchenden Objekte einer Reihe von Versuchspersonen vor ([GAD96]).

Mittels Blickverfolgung wird dann ermittelt, welche Teile des Objekts in welcher Reihenfolge wie lange fixiert werden, d.h. welche Teile für die Testteilnehmer wie interessant sind. Dieses aufwendige Verfahren könnte durch die hier vorgestellte, automatische Attraktivitätsbewertung von Dokumenten teilweise ersetzt und ergänzt werden. Ein vollständiger Austausch erscheint zunächst nicht möglich, da das Verfahren der GAD nicht nur die unbewußte, visuelle Aufmerksamkeit der „preattentive vision“, sondern auch alle höheren kognitiven Prozesse, untersucht. In höheren kognitiven Prozessen spielen aber auch Einflüsse wie Modeströmungen eine Rolle, so daß sich das Seh- und Rezeptionsverhalten mit der Zeit ändern kann. Die Modellierung höherer kognitiver Prozesse und einer damit verbundenen zeitlichen Änderung ist im hier präsentierten Ansatz nicht vorgesehen und könnte höchstens durch ständige, auch wieder aufwendige, Befragungen versucht werden.

## Literatur

- [And88] J. R. Anderson: *Kognitive Psychologie*, Spektrum der Wissenschaft Verlag, Heidelberg, 1988
- [App90] W. Appelt: *Dokumentaustausch in Offenen Systemen, Einführung in die ISO-Norm 8613: Office Document Architecture (ODA) and Interchange Format*, Springer Verlag, Berlin, Heidelberg, 1990
- [Bac97] M. Bach: *Visuelle Textursegmentierung - elektrophysiologische und psychophysische Untersuchungen*, in Experimentelle und theoretische Hirnforschung II. Sinneswahrnehmung, sensomotorische Koordination, neuronale Informationsverarbeitung, Freiburger Universitätsblätter, März 97, 1997, pp. 27 – 36
- [Bal82] D. H. Ballard, C. M. Brown: *Computer Vision*, Prentice-Hall, Englewood Cliffs, New Jersey, 1982
- [Ber96] L. Berekoven, W. Eckert, P. Ellenrieder: *Marktforschung*, Gabler, Wiesbaden, 1996
- [Blo91] D. S. Bloomberg: *Multiresolution Morphological Approach to Document Image Analysis*, Proc. First International Conference on Document Analysis and Recognition, 1991, pp. 963 – 971
- [Bro88] *Brockhaus-Enzyklopädie*, Brockhaus, Mannheim, 1988
- [Bro91] I. N. Bronstein, K. A. Semendjajew: *Taschenbuch der Mathematik*, B. G. Teubner Verlagsgesellschaft Stuttgart, Leipzig und Verlag Nauka, Moskau, 1991
- [Brü96] T. Brückner, P. Suda, H. U. Block, G. Maderlechner: *In-house Mail Distribution by Automatic Address and Content Interpretation*, Proc. Fifth Annual Symposium on Document Analysis and Information Retrieval, 1996, pp. 67 – 75
- [Cha98] B. B. Chaudhuri, U. Garain: *Automatic Detection of Italic, Bold and All-Capital Words in Document Images*, Proc. International Conference on Pattern Recognition, 1998, pp. 610 – 612
- [Che92] F. R. Chen, M. Withgott: *The Use of Emphasis to Automatically Summarize a Spoken Discourse*, Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 1992, pp. 228 – 232
- [Che97a] F. R. Chen, D. S. Bloomberg: *Extraction of Indicative Summary Sentences from Imaged Documents*, Proc. Fourth International Conference on Document Analysis and Recognition, 1997, pp. 227 – 232

- [Che97b] J.-L. Chen: *A Simplified Approach to the HMM Based Texture Analysis and its Application to Document Segmentation*, Pattern Recognition Letters 18, 1997, pp. 993 – 1007
- [Cho95] P. A. Chou, G. E. Kopec: *A Stochastic Attribute Grammar Model of Document Production and its Use in Document Image Decoding*, SPIE, Vol. 2422, 1995, pp. 66 – 73
- [Cla86] M. Clark, A. C. Bovik: *Texture Discrimination Using a Model of Visual Cortex*, Proc. IEEE International Conference on Systems, Man and Cybernetics, 1986, pp. 1425 – 1430
- [Cla89] M. Clark, A. C. Bovik: *Experiments in Segmenting Texton Patterns Using Localized Spatial Filters*, Pattern Recognition, Vol. 22, No. 6, 1989, pp. 707 – 717
- [Cul97] J. F. Cullen, J. J. Hull, P. E. Hart: *Document Image Retrieval and Browsing Using Texture Analysis*, Proc. Fourth International Conference on Document Analysis and Recognition, 1997, pp. 718 – 721
- [Doe97] D. Doermann: *The Retrieval of Document Images: A Brief Survey*, Proc. Fourth International Conference on Document Analysis and Recognition, 1997, pp. 945 – 949
- [Doe98] D. Doermann, E. Rivlin, A. Rosenfeld: *The Function of Documents*, Image and Vision Computing 16, 1998, pp. 799 – 814
- [Dor96] D. Dori, D. Doermann, C. Shin, R. Haralick, I. Phillips, M. Buchman, D. Ross: *The Representation of Document Structure: a Generic Object-Process Analysis*, in P. S. P. Wang, H. Bunke (eds.): *Handbook of Optical Character Recognition and Document Image Analysis*, World Scientific Publishing Company, 1996
- [Dut96] C. Dutt, M. Schmithäuser: *Der neue Desktop Knigge*, Franzis, 1996
- [Egl97] V. Eglin, H. Emptoz: *Logarithmic Spiral Grid and Gaze Control for the Development of Strategies of Visual Segmentation on a Document*, Proc. Fourth International Conference on Document Analysis and Recognition, 1997, pp. 689 – 692
- [Egl98] V. Eglin: *Contributions à la structuration fonctionnelle des documents imprimés*, Doktorarbeit, Institut National des Sciences Appliquées, Lyon, 1998
- [Ete97] K. Etemad, D. Doermann, R. Chellappa: *Multiscale Segmentation of Unstructured Document Pages Using Soft Decision Integration*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 1, 1997, pp. 92 – 96
- [Fos77] J. Foster, P. Coles: *An Experimental Study of Typographic Cueing in Printed Text*, Ergonomics, Vol. 20, No. 1, 1977, pp. 57 – 66



- [GAD96] Gesellschaft für angewandtes Direktmarketing mbH: *Verstehen, was wir sehen. Die Blick-Registration mit der Augenkamera*, 1996
- [Har94] R. M. Haralick: *Document Image Understanding: Geometric and Logical Layout*, Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1994, pp. 385 – 390
- [Itt98] L. Itti, C. Koch, E. Niebur: *A Model of Saliency-Based Visual Attention for Rapid Scene Analysis*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, 1998, pp. 1254 – 1259
- [Jai92] A. K. Jain, S. Bhattacharjee: *Text Segmentation Using Gabor Filters for Automatic Document Processing*, Machine Vision and Application, Vol. 5, 1992, pp. 169 – 184
- [Jai96] A. K. Jain, Y. Zhong: *Page Segmentation Using Texture Analysis*, Pattern Recognition, Vol. 29, No. 5, 1996, pp. 743 – 770
- [Jul83] B. Julesz, J. R. Bergen: *Textons, The Fundamental Elements in Preattentive Vision and Perception of Textures*, The Bell System Technical Journal, Vol. 62, No. 6, 1983, pp. 1619 – 1645
- [Jul91] B. Julesz: *Early Vision and Focal Attention*, Reviews of Modern Physics, Vol. 63, No. 3, 1991, pp. 735 – 772
- [Kli92] T. S. Klitz, J. S. Mansfield, G. E. Legge: *Font "Pop Out" in Text Images*, OSA Annual Meeting Technical Digest 23 (Optical Society of America, Washington, D.C.), 1992, pp. 170
- [Koc85] C. Koch, S. Ullman: *Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry*, Human Neurobiology 4, 1985, pp. 219 – 227
- [Kho96] S. Khoubyari, J. J. Hull: *Font and Function Word Identification in Document Recognition*, Computer Vision and Image Understanding, Vol. 63, No. 1, 1996, pp. 66 – 74
- [Lin81] P. H. Lindsay, D. A. Norman: *Einführung in die Psychologie: Informationsaufnahme und -verarbeitung beim Menschen*, Springer Verlag, Berlin, Heidelberg, New York, 1981
- [Mad99] G. Maderlechner, A. Schreyer, P. Suda: *Information Extraction from Document Images Using Attention Based Layout Segmentation*, Online Proc. Document Layout Interpretation and its Applications Workshop (Fifth International Conference on Document Analysis and Recognition), 1999, Ib
- [Mad00] G. Maderlechner, A. Schreyer, P. Suda: *Extraction of Relevant Information from Document Images Using Measures of Visual Attention*, Proc. International Conference on Pattern Recognition, 2000

- [Mal90] J. Malik, P. Perona: *Preattentive Texture Discrimination with Early Vision Mechanisms*, J. Opt. Soc. Am. A, Vol. 7, No. 5, 1990, pp. 923 – 932
- [Mor92] R. A. Morris: *Classification of Digital Typefaces Using Spectral Signatures*, Pattern Recognition, Vol. 25, No. 8, 1992, pp. 869 – 876
- [Nag92] G. Nagy, S. Seth, M. Viswanathan: *A Prototype Document Image Analysis System for Technical Journals*, IEEE Computer, Vol. 25, No. 7, 1992, pp. 10 – 24
- [Nor89] P. Norrish: *Semantic Structures of Text*, in J. Andre, R. Furuta, V. Quint: *Structured Documents*, Cambridge University Press, 1989, pp. 143 – 159
- [OGo93] L. O’Gorman: *The Document Spectrum for Page Layout Analysis*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 15, No. 11, 1993, pp. 1162 – 1173
- [Pai90] C. D. Paice: *Constructing Literature Abstracts by Computer: Techniques and Prospects*, Information Processing & Management, Vol. 26, No. 1, 1990, pp. 171 – 186
- [Rea85] T. C. Rearick: *A Texture Analysis Algorithm Inspired by a Theory of Preattentive Vision*, Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1985, pp. 312 – 217
- [Sal94] G. Salton, J. Allan, C. Buckley, A. Singhal: *Automatic Analysis, Theme Generation, and Summarization of Machine-Readable Texts*, Science, Vol. 264, No. 3, 1994, pp. 1421 – 1426
- [Sch98a] A. Schreyer, P. Suda, G. Maderlechner: *Font Style Detection Using Textons*, Proc. Document Analysis Systems Workshop, 1998, pp. 99 – 108
- [Sch98b] A. Schreyer, P. Suda, G. Maderlechner: *The Idea of Attention-Based Document Analysis*, Proc. Document Analysis Systems Workshop, 1998, pp. 214 – 217
- [Sch99] A. Schreyer, P. Suda, G. Maderlechner: *A Formal Approach to Textons and its Application to Font Style Detection*, LNCS, Springer Verlag, 1999
- [Shi97] H. Shi, T. Pavlidis: *Font Recognition and Contextual Processing for more Accurate Text Recognition*, Proc. Fourth International Conference on Document Analysis and Recognition, 1997, pp. 39 – 44
- [Sud82] S. Sudman, N. M. Bradburn: *Asking Questions*, Jossey-Bass Publishers, San Francisco, 1982
- [Sud97] P. Suda, C. Bridoux, B. Kämmerer, G. Maderlechner: *Logo and Word Matching Using a General Approach to Signal Registration*, Proc. Fourth International Conference on Document Analysis and Recognition, 1997, pp. 61 – 65

- [Sou88] R. Southall: *Visual Structure and the Transmission of Meaning*, in J.C. van Vliet (ed.): *Document Manipulation and Typography*, Cambridge University Press, 1988, pp. 35 – 45
- [Tan98] T. N. Tan: *Rotation Invariant Texture Features and Their Use in Automatic Script Identification*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 7, 1998, pp. 751 – 756
- [Tre80] A. Treisman, G. Gelade: *A Feature-Integration Theory of Attention*, Cognitive Psychology 12, 1980, pp. 97 – 136
- [UoW96] *UW-III English/Technical Document Image Database*, CD-ROM, University of Washington, 1996
- [Voo88] H. Voorhees, T. Poggio: *Computing Texture Boundaries from Images*, Nature, Vol. 333, 1988, pp. 364 – 367
- [Wal97] H. Walischewski: *Automatic Knowledge Acquisition for Spatial Document Interpretation*, Proc. Fourth International Conference on Document Analysis and Recognition, 1997, pp. 243 – 247
- [Wan89] D. Wang, S. Srihari: *Classification of Newspaper Image Blocks Using Texture Analysis*, Computer Vision, Graphics and Image Processing 47, 1989, pp. 327 – 352
- [Wat93] R. J. Watt: *The Visual Analysis of Pages of Text*, in R. Sassoon (ed.): *Computers and Typography*, Intellect Books, Oxford, 1993, pp. 178 – 201
- [Zra98] A. Zramdini, R. Ingold: *Optical Font Recognition Using Typographical Features*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 8, 1998, pp. 877 – 882



## Verzeichnis der Formelzeichen

|                                  |   |
|----------------------------------|---|
| $A(b)$                           | Attraktivitätsbewertung eines Graphikblockes $b$  |
| $A'(b)$                          | Normierte Attraktivitätsbewertung eines Graphikblockes $b$ oder Attraktivitätsmittelwert eines Textblockes $b$        |
| $A(w)$                           | Attraktivitätsbewertung eines Wortes $w$  |
| $A'(w)$                          | Normierte Attraktivitätsbewertung eines Wortes $w$  |
| $A'_{ck}(w, b)$                  | Von Testteilnehmer $k$ korrigierte Attraktivität eines Wortes $w$ oder eines Graphikblocks $b$ im interaktiven Test   |
| $A'_p(w, b)$                     | Präsentierte Attraktivität eines Wortes $w$ oder eines Graphikblockes $b$ im interaktiven Test, entspricht $A'(w, b)$ |
| $AllCapitals(w)$                 | Layoutmerkmal eines Wortes $w$ (Großschreibung)   |
| $AllCapitals'(w)$                | Normiertes Layoutmerkmal eines Wortes $w$ (Großschreibung)  |
| $Area(t)$                        | Fläche eines Textons $t$  |
| $AreaMeas(b)$                    | Layoutmerkmal eines Graphikblockes $b$ (Flächenmaß)   |
| $AreaMeas'(b)$                   | Normiertes Layoutmerkmal eines Graphikblockes $b$ (Flächenmaß)  |
| $AspectRatio(w)$                 | Aspektverhältnis eines Wortes $w$   |
| $AttrMax/Min$                    | Höchste/niedrigste Attraktivitätsbewertung $A(w)$ bzw. $A(b)$ auf einem Dokument                                      |
| $b$                              | Block-Texton, kurz auch „Block“   |
| $Base(w, l)$                     | Basislinie eines Wortes $w$ oder einer Zeile $l$  |
| $BlackInt(b)$                    | Schwärzungsgrad eines Blockes $b$ , abgeleitet aus $BlackInt(w)$  |
| $BlackInt(w)$                    | Schwärzungsgrad eines Wortes $w$ , bezogen auf $MiddleZone(w)$  |
| $BlackIntens(b)$                 | Layoutmerkmal eines Graphikblockes $b$ (Schwärzungsgrad)  |
| $BlackIntens'(b)$                | Normiertes Layoutmerkmal eines Graphikblockes $b$ (Schwärzungsgrad)   |
| $BlackIntens(w)$                 | Schwärzungsgrad eines Wortes $w$ , bezogen auf $Area(w)$  |
| $BlockDistTop/Bot/Left/Right(b)$ | Abstände eines Blockes $b$ zu den nach oben, unten, links und rechts nächstgelegenen Blöcken                          |
| $BLOCKS$                         | Menge aller Block-Textons auf einem Dokument  |
| $Bold(w)$                        | Layoutmerkmal eines Wort-Textons $w$ (Fettdruck)  |
| $Bold'(w)$                       | Normiertes Layoutmerkmal eines Wort-Textons $w$ (Fettdruck)   |
| $Bot(t)$                         | Unterer Rand des Textons $t$ (y-Koordinate)   |
| $Bullet(w)$                      | Layoutmerkmal eines Wort-Textons $w$ (Bullet)   |
| $Bullet'(w)$                     | Normiertes Layoutmerkmal eines Wort-Textons $w$ (Bullet)  |
| $c$                              | Buchstaben-Texton, kurz auch „Buchstabe“  |

|                                   |   |
|-----------------------------------|---|
| $C(t)$                            | Klassifikation eines Textons $t$  |
| $Capitals(b)$                     | Merkmal für einen großgeschriebenen Block $b$   |
| $Capitals(w)$                     | Merkmal für ein großgeschriebenes Wort $w$  |
| $Cent(r)$                         | Horizontale Mitte eines Runs $r$  |
| $CHARS$                           | Menge aller Buchstaben-Textons auf einem Dokument   |
| $d_{\text{Graphik}}$              | Maß für die strukturbedingt höhere Attraktivität von Graphikobjekten im Vergleich zu Textobjekten   |
| $D(w, b)$                         | Differenz zwischen der präsentierten Attraktivität und dem Mittelwert der korrigierten Attraktivitäten eines Wortes $w$ oder eines Graphikblocks $b$ im interaktiven Test |
| $Dom(t)$                          | Träger des Textons $t$  |
| $\langle e_1, \dots, e_n \rangle$ | Merkmalsvektor in $E$   |
| $E, E^*$                          | Merkmalsräume   |
| $E_i$                             | Menge der möglichen Ausprägungen des Merkmals $i$   |
| $FrameTop/Bot/Left/Right$         | Oberer/Unterer/Linker/Rechter Rand des Rahmens um alle Blöcke eines Dokuments   |
| $Free(b)$                         | Layoutmerkmal eines Blocks $b$ (Freistellung)   |
| $Free'(b)$                        | Normiertes Layoutmerkmal eines Blocks $b$ (Freistellung)  |
| $g_{\text{AllCapitals}}$          | Gewicht für Layoutmerkmal Großschreibung $AllCapitals'(w)$  |
| $g_{\text{AreaMeas}}$             | Gewicht für Layoutmerkmal Flächenmaß $AreaMeas'(b)$   |
| $g_{\text{BlackIntens}}$          | Gewicht für Layoutmerkmal Schwärzungsgrad $BlackIntens'(b)$   |
| $g_{\text{Bold}}$                 | Gewicht für Layoutmerkmal Fettdruck $Bold'(w)$  |
| $g_{\text{Bullet}}$               | Gewicht für Layoutmerkmal Bullet $Bullet'(w)$   |
| $g_{\text{Free}}$                 | Gewicht für Layoutmerkmal Freistellung $Free'(b)$   |
| $g_{\text{Indent}}$               | Gewicht für Layoutmerkmal Einrückung $Indent'(w)$   |
| $g_{\text{Italics}}$              | Gewicht für Layoutmerkmal Kursivdruck $Italics'(w)$   |
| $g_{\text{MeanHeight}}$           | Gewicht für Layoutmerkmal Schriftgröße $MeanHeight'(b)$   |
| $Graphic(t)$                      | Layoutmerkmal eines Textons $t$ (Graphik)   |
| $H(i)$                            | Wert im Histogramm für Abschnitt $i$  |
| $Height(t)$                       | Höhe des Textons $t$  |
| $Indent(w)$                       | Layoutmerkmal eines Wortes $w$ (Einrückung)   |
| $Indent'(w)$                      | Normiertes Layoutmerkmal eines Wortes $w$ (Einrückung)  |
| $Italics(w)$                      | Layoutmerkmal eines Wortes $w$ (Kursivdruck)  |
| $Italics'(w)$                     | Normiertes Layoutmerkmal eines Wortes $w$ (Kursivdruck)   |
| $l$                               | Zeilen-Texton, kurz auch „Zeile“  |

|  |   |
|--|---|
| $Left(t)$                                | Linker Rand des Textons $t$ (x-Koordinate)  |
| $LINES$                                  | Menge aller Zeilen-Textons auf einem Dokument   |
| $M$                                      | Meßraum   |
| $M(w, b)$                                | Mittelwert der korrigierten Attraktivitäten eines Wortes $w$ oder eines Graphikblockes $b$ im interaktiven Test |
| $MeanHeight(b)$                          | Layoutmerkmal eines Blockes $b$ (Schriftgröße), auch mittlere Zeilenhöhe eines Blockes $b$                      |
| $MeanHeight'(b)$                         | Normiertes Layoutmerkmal eines Blockes $b$ (Schriftgröße)   |
| $MeanSlant(b)$                           | Mittlere Schräglage der Strokes innerhalb eines Blockes $b$   |
| $MeanSlant(w)$                           | Mittlere Schräglage der Strokes innerhalb eines Wortes $w$  |
| $Middle(w)$                              | Mittellinie eines Wortes $w$  |
| $MiddleZone(w)$                          | Mittelzone eines Wortes $w$   |
| $N$                                      | Nachbarschaftsrelation  |
| $N_h$                                    | Nachbarschaftsrelation auf einer höheren Hierarchiestufe  |
| $O$                                      | Objektraum  |
| $OverHor(b, b_1),$<br>$OverVert(b, b_1)$ | Horizontale und vertikale Überlappung zweier Blöcke $b$ und $b_1$   |
| $p$                                      | Pixel-, „Texton“, kurz auch „Pixel“   |
| $Pred(w)$                                | Direktes Vorgänger-Wort eines Wortes $w$  |
| $Prop(t)$                                | Merkmalsvektor des Textons $t$  |
| $r$                                      | Run-, „Texton“, kurz auch „Run“   |
| $R_i$                                    | Stellenwert eines Dokumentbereiches/Layoutmerkmals in der Befragung   |
| $\bar{R}_i$                              | Mittelwert der angegebenen Stellenwerte   |
| $\tilde{R}_i$                            | Mittlere Abweichung der angegebenen Stellenwerte  |
| $\overline{\tilde{R}_i}$                 | Mittlere Abweichung der angegebenen Stellenwerte pro Dokument/Frage   |
| $Right(t)$                               | Rechter Rand des Textons $t$ (x-Koordinate)   |
| $s$                                      | Stroke-Texton, kurz auch „Stroke“   |
| $Slant(s)$                               | Schräglage eines Strokes $s$  |
| $STROKES$                                | Menge aller Stroke-Textons auf einem Dokument   |
| $Suc(w)$                                 | Direktes Nachfolger-Wort eines Wortes $w$   |
| $t_h$                                    | Texton einer höheren Hierarchiestufe (verglichen mit $t_n$ )  |
| $t_n$                                    | Texton einer niedrigeren Hierarchiestufe (verglichen mit $t_h$ )  |
| $TEX_{O,E,F}$ , oder $TEX$               | Menge aller Textons in $TS=(O,E,F)$   |

|                        |   |
|------------------------|---|
| $TEXTURES$             | Menge aller $\mathcal{S}$ - $N$ -Texturen in $TS$   |
| $Th, Th_1, Th_2, Th_3$ | Entscheidungsschwellen bei der Klassifikation von Textons   |
| $Top(t)$               | Oberer Rand des Textons $t$ (y-Koordinate)  |
| $Top(w, l)$            | Oberlinie eines Wortes $w$ oder einer Zeile $l$   |
| $TS$                   | Textonstruktur  |
| $V(w, b)$              | Mittlere Abweichung der korrigierten Attraktivitäten eines Wortes $w$ oder eines Graphikblocks $b$ im interaktiven Test |
| $Verticals(w)$         | Menge der ungefähr senkrechten Strokes im Wort $w$  |
| $w$                    | Wort-Texton, kurz auch „Wort“   |
| $Width(t)$             | Breite des Textons $t$  |
| $WordDist(w)$          | Abstand des Wortes $w$ zum nach rechts nächstgelegenen Wort   |
| $WORDS$                | Menge aller Wort-Textons auf einem Dokument   |
| $X$                    | Menge der Punkte des Raumes   |
| $D$                    | Abbildung $D:TEXTURES \rightarrow E^*$ in einer Texturhierarchie  |
| $F, F^*$               | Registrierungsfunktionen  |
| $m, m^*$               | Bewertungen   |
| $m _{TEX_j}$           | Bewertung, beschränkt auf die Hierarchiestufe $j$ der Texturhierarchie  |
| $g$                    | Objekt aus $G$  |
| $G, G^*$               | Mengen aller Objekte in $O$   |
| $Q$                    | Abbildung $Q:TEXTURES \rightarrow G^*$ in einer Texturhierarchie  |
| $S$                    | Ähnlichkeitsrelation  |
| $S_K$                  | Teil der Ähnlichkeitsrelation $S$ (explizite Klassen)   |
| $S_R$                  | Teil der Ähnlichkeitsrelation $S$ (implizite Klassen)   |



## Verzeichnis wichtiger Begriffe

|                                    |  |
|------------------------------------|--|
| Ähnlichkeitsrelation <b>S</b>      | In dieser Arbeit eine reflexive, symmetrische und transitive Relation auf dem Merkmalsraum, die die Ähnlichkeit von Textons beschreibt   |
| Attentive vision                   | Nach der psychologischen Texton Theorie ([Jul83]) einer der beiden Modi der menschlichen, visuellen Informationsverarbeitung: ein Prozeß, der beschränkt auf eine kleine Blende eine serielle Suche realisiert, wie bei der Formerkennung                                  |
| Attraktivität                      | In dieser Arbeit die Eigenschaft eines Dokumentbereichs, die mit der Aufmerksamkeit eines Betrachters korrespondiert   |
| Attraktivitätsbewertung            | In dieser Arbeit ein Maß für die Attraktivität eines Dokumentbereiches, das einen Rückschluß darauf zuläßt, wie stark der Dokumentbereich die Aufmerksamkeit eines Betrachters auf den ersten Blick anziehen würde; implizit eine Bewertung der Relevanz von Informationen |
| Attraktivitätsverteilung           | Ergibt sich in dieser Arbeit, indem jedem einzelnen Wort oder Graphikblock eines Dokuments seine Attraktivitätsbewertung zugeordnet wird   |
| Aufmerksamkeit                     | In dieser Arbeit eine unbewußte Reaktion auf visuelle Eindrücke nach Julesz ([Jul83]), im Gegensatz zur höheren kognitiven Leistung der bewußten Aufmerksamkeit ([And88], [Lin81])   |
| Aufmerksamkeitsbasiertes Verfahren | In dieser Arbeit ein Verfahren, das die menschliche Aufmerksamkeit übertragen auf ein Dokument als Attraktivität von Dokumentbereichen modelliert; diese Modellierung dient als Grundlage weiterer Schritte der Dokumentanalyse.   |
| Betonung                           | In den Stream-Modellen Hervorhebung von Anteilen des ein- oder zweidimensionalen Streams, damit der Informationsempfänger die Gedankenstrukturen des Empfängers möglichst korrekt und effizient aus dem Stream extrahieren kann  |
| Betrachter                         | Informationsempfänger (Leser), der nur die visuellen Signale decodiert, den Stream selbst aber nicht wirklich liest  |
| Bewertung relevanter Information   | In dieser Arbeit ein Maß dafür, wie relevant eine Information im Vergleich zu anderen Informationen auf einem Dokument ist   |
| Block                              | Objekt, das einem Block-Texton zugrunde liegt  |
| Block-Texton                       | Texton   |
| Block-Textur                       | Eine Textur bestehend aus Block-Textons, auf der nächsthöheren Hierarchiestufe ein „Superblock“-Texton   |
| Buchstabe                          | Objekt, das einem Buchstaben-Texton zugrunde liegt   |
| Buchstaben-Texton                  | Texton   |
| Buchstaben-Textur                  | Eine Textur bestehend aus Buchstaben-Textons, auf der nächsthöheren Hierarchiestufe ein Wort-Texton  |

|                      |   |
|----------------------|---|
| Bullet               | Layoutmerkmal (kleines Symbol, um die einzelnen Punkte einer Aufzählung zu markieren)   |
| Codierung            | Im Stream-Modell II eine deutliche Änderung der Informationsdarstellung:<br>eindimensionaler Stream → zweidimensionaler Stream,<br>sprachliche Betonung → Layoutmerkmale  |
| Connected Components | Zusammenhangskomponenten, d.h. Gebiete von 8-zusammenhängenden schwarzen Runs   |
| Decodierung          | Im Stream-Modell II eine deutliche Änderung der Informationsdarstellung:<br>zweidimensionaler Stream → eindimensionaler Stream,<br>Layoutmerkmale → sprachliche Betonung  |
| Dokument             | Nach der Brockhaus Enzyklopädie ([Bro88]) „das zur Belehrung über etwas oder zur Erhellung von etwas Dienliche“, d.h. ein Medium, das Information zum Zweck der Informationsübertragung speichert;<br>in dieser Arbeit vor allem Bezeichnung für ein Papierdokument, auch für das Binärbild eines eingescannten Papierdokuments |
| Dokumentbereich      | In dieser Arbeit ein Gebiet auf dem Bild des eingescannten Papierdokuments  |
| Dokumenttyp          | Grobe Einordnung des Dokuments, z.B. Einteilung in verschiedene Kategorien von Layout   |
| Dokumentinhalt       | Detaillierte Information aus dem Dokument, z.B. Schlüsselwörter des Textes oder eine Zusammenfassung des Dokuments  |
| Einrückung           | Layoutmerkmal (weißer Einzug in der ersten Zeile eines Absatzes)  |
| Empfänger            | Siehe (Informations-)Empfänger  |
| Extraktion           | Umsetzung der (relevanten) Information auf einem Dokument in eine symbolische Repräsentation zur informationstechnischen Weiterverarbeitung   |
| Fettdruck            | Layoutmerkmal (z.B. <b>xxxxxxx</b> )  |
| Flächenmaß           | Layoutmerkmal (Maß für die Ausdehnung eines Graphikblocks)  |
| Font                 | Schrift, Schriftgröße und Fontstil  |
| Fontstil             | z.B. Fettdruck, Kursivdruck   |
| Freistellung         | Layoutmerkmal (Größe des Weißbereichs, der einen Block umgibt)  |
| Gewichte             | In dieser Arbeit Faktoren in den Bewertungen der einzelnen Layoutmerkmale, die die unterschiedliche Attraktivität einzelner Layoutmerkmale modellieren sollen   |
| Graphikobjekt        | Enthält nur bildhafte Darstellungen   |
| Großschreibung       | Layoutmerkmal (z.B. XXXXXXXX)   |
| Information          | Nach der Brockhaus Enzyklopädie ([Bro88]) eine „Mitteilung, Nachricht, Auskunft“  |

|                                      |   |
|--------------------------------------|---|
| (Informations-) Empfänger            | In den Stream-Modellen I und II derjenige Partner bei der Informationsübermittlung, der die Information aufnimmt, d.h. im Stream-Modell II der Leser des Dokuments  |
| (Informations-) Sender               | In den Stream-Modellen I und II derjenige Partner bei der Informationsübermittlung, der die Information abgibt, d.h. im Stream-Modell II der Autor des Dokuments  |
| Klassifikation                       | In dieser Arbeit die Zuordnung eines Textons zu einer durch die Ähnlichkeitsrelation $\mathbf{S}$ gebildeten Äquivalenzklasse im Merkmalsraum   |
| Kursivdruck                          | Layoutmerkmal (z.B. xxxxxx)   |
| Layout                               | Abmessungen und geometrische Anordnung von Dokumentbereichen  |
| Layoutmerkmale                       | In dieser Arbeit eine Zusammenfassung der Begriffe Layout und Typographie   |
| Lokalisierung relevanter Information | In dieser Arbeit das Finden derjenigen Dokumentbereiche auf dem Dokument, die relevante Information enthalten   |
| Nachbarschaftsrelation $N$           | In dieser Arbeit eine reflexive und symmetrische Relation auf der Menge aller Textons, die den räumlichen Zusammenhang von Textons beschreibt   |
| Objekt                               | Nach der Formalisierung in dieser Arbeit eine nicht-leere Menge von Punkten im Meßraum, d.h. konkreter eine nicht-leere Menge von schwarzen Pixeln auf dem Dokument; Objekte sind paarweise disjunkt.   |
| OCR                                  | Automatisches Verfahren, das durch Mustererkennung gedruckten (oder handgeschriebenen) Buchstaben eine entsprechende symbolische Repräsentation, z.B. das entsprechende ASCII-Zeichen, zuordnet   |
| Preattentive vision                  | Nach der psychologischen Texton Theorie ([Jul83]) einer der beiden Modi der menschlichen, visuellen Informationsverarbeitung: ein Prozeß, der augenblicklich, ohne genaues Hinschauen, ein großes Gesichtsfeld analysiert, und zwar in Parallelverarbeitung unabhängig von der Anzahl der Muster; beruht auf Texturwahrnehmung. |
| Ranking                              | In dieser Arbeit die Reihenfolge der Dokumentbereiche oder Layoutmerkmale bezüglich ihrer Attraktivität   |
| Relevanz                             | Bezeichnung dafür, wie wichtig oder interessant eine Information ist; in dieser Arbeit werden diejenigen Informationen als relevant betrachtet, die der Sender als wichtig erachtet.  |
| Resegmentierung                      | In dieser Arbeit die erneute Segmentierung einer bereits bestehenden Segmentierung durch Verfeinerung der Ähnlichkeitsrelation $\mathbf{S}$ und der Nachbarschaftsrelation $N$  |
| Run                                  | Eine zusammenhängende Folge von schwarzen oder weißen Pixeln innerhalb einer Bildzeile  |
| Schriftgröße                         | Layoutmerkmal (mittlere Zeilenhöhe eines Blocks)  |

|                                      |   |
|--------------------------------------|---|
| Schwärzungsgrad                      | Layoutmerkmal (Maß für die visuelle Dichte des Blocks, das zur groben Unterscheidung von Rasterbildern und Liniengraphiken benutzt wird); auch: Merkmal eines Textons                               |
| Segmentierung                        | In dieser Arbeit die Menge aller $S-N$ -Texturen über der Menge aller Textons   |
| Segmentierungsprozeß                 | In dieser Arbeit der Prozeß zur Bestimmung aller $S-N$ -Texturen über der Menge aller Textons   |
| Sender                               | Siehe (Informations-) Sender  |
| Stream                               | In dieser Arbeit die Bezeichnung für eine in der Zeit lineare Wortkette   |
| Stream-Modell I                      | Beschreibt in dieser Arbeit die Informationsübertragung von Mensch zu Mensch, wenn als Medium gesprochene Sprache benutzt wird  |
| Stream-Modell II                     | Beschreibt in dieser Arbeit die Informationsübertragung von Mensch zu Mensch, wenn als Medium ein Dokument benutzt wird   |
| Stroke                               | Objekt, das einem Stroke-Texton zugrunde liegt  |
| Stroke-Texton                        | Texton  |
| Stroke-Textur                        | Eine Textur bestehend aus Stroke-Textons, auf der nächsthöheren Hierarchiestufe ein Buchstaben-Texton   |
| „Superblock“                         | Objekt, das einem „Superblock“-Texton zugrunde liegt  |
| „Superblock“-Texton                  | Texton  |
| Textobjekt                           | Enthält nur Text  |
| Texton                               | Bei Julesz ([Jul83]) eine diskrete Wahrnehmungseinheit mit einfachen, direkt beobachtbaren Eigenschaften; in dieser Arbeit formalisiert als Paar aus Objekt und Merkmalsvektor                      |
| Texton Theorie                       | Von B. Julesz ([Jul83]) entwickeltes Modell der menschlichen visuellen Wahrnehmung (siehe auch Preattentive vision, Attentive vision, Texton und Textur)  |
| Textur                               | Bei Julesz ([Jul83]) eine Anhäufung von ähnlichen, benachbarten Textons; in dieser Arbeit formalisiert als maximale Menge $S-N$ -zusammenhängender Textons  |
| Texturbasierte Attraktivitätsanalyse | In dieser Arbeit die zusammenfassende Bezeichnung für Segmentierung, Resegmentierung und Attraktivitätsbewertung  |
| Texturhierarchie                     | In dieser Arbeit formal dadurch definiert, daß eine Textur auf der nächsthöheren Hierarchiestufe ein Texton bildet, das dann wiederum zu einer Textur dieser höheren Hierarchiestufe beitragen kann |
| Typographie                          | Gestalterische Merkmale wie Schriftgröße und Font   |
| Wort                                 | Objekt, das einem Wort-Texton zugrunde liegt  |
| Wort-Texton                          | Texton  |

|               |  |
|---------------|--|
| Wort-Textur   | Eine Textur bestehend aus Wort-Textons, auf der nächsthöheren Hierarchiestufe ein Zeilen-Texton  |
| Zeile         | Objekt, das einem Zeilen-Texton zugrunde liegt   |
| Zeilen-Texton | Texton   |
| Zeilen-Textur | Eine Textur bestehend aus Zeilen-Textons, auf der nächsthöheren Hierarchiestufe ein Block-Texton |



# A Fragebogen zur Wahrnehmung von Layoutmerkmalen

Bitte zurück an:  
 Angela Schreyer, ZT IK 5, Otto-Hahn-Ring 6, 81730 München (Mch P, 10-399), Tel. 089/636-42283

The image shows four numbered examples of document layouts. Each example has small boxes with lines pointing to specific layout features. Example 1 shows a page with a large image and text columns. Example 2 shows a page with a large image and text columns. Example 3 shows a page with a large image and text columns. Example 4 shows a page with a large image and text columns.

Bitte zurück an:  
 Angela Schreyer, ZT IK 5, Otto-Hahn-Ring 6, 81730 München (Mch P, 10-399), Tel. 089/636-42283

Vorname, Name: ..... Datum: .....  
 Beruf: ..... Alter: .....

## Liebe Teilnehmerin, lieber Teilnehmer,

vielen Dank, daß Sie sich Zeit nehmen, uns ein paar Fragen zu einem interessanten Problem zu beantworten.

In unserer Forschungsarbeit zur Dokumentanalyse fragen wir uns, welche Dokumentbereiche die Aufmerksamkeit eines Betrachters auf den ersten Blick anziehen. Insbesondere interessiert uns, wie stark die Attraktivität der einzelnen Dokumentbereiche ist, wenn man davon ausgeht, daß die Attraktivität **nur** durch Merkmale des **Layouts**, und **nicht** durch den **Inhalt**, gegeben ist.

### Frage 1

Im Anhang sehen sie vier Dokumente. Es handelt sich um Kopien aus wissenschaftlichen Veröffentlichungen, bzw. Zeitschriften. Diese Art von Dokumenten steht bei unserer Untersuchung im Vordergrund.

Zur weiteren Bearbeitung von Frage 1 können Sie die vier Dokumente abtrennen.

#### Frage 1.1

Beherrschen Sie die Sprache, in der diese Dokumente verfaßt sind?

ja  nein

Wenn ja, welche Sprache ist das? .....

#### Frage 1.2

Auf der nächsten Seite sehen sie Verkleinerungen der vier Dokumente, in denen einzelne Dokumententeile markiert und mit Kästchen versehen sind

Bitte tragen Sie in den entsprechenden Kästchen bei Dokument

- 1) eine Zahl von 1 (zieht Aufmerksamkeit sehr stark an) bis 7 (zieht Aufmerksamkeit nur wenig an)
- 2) eine Zahl von 1 (zieht Aufmerksamkeit sehr stark an) bis 4 (zieht Aufmerksamkeit nur wenig an)
- 3) eine Zahl von 1 (zieht Aufmerksamkeit sehr stark an) bis 9 (zieht Aufmerksamkeit nur wenig an)
- 4) eine Zahl von 1 (zieht Aufmerksamkeit sehr stark an) bis 8 (zieht Aufmerksamkeit nur wenig an)

ein, wobei Sie **jede Zahl nur ein Mal** benutzen. Sie soll ein Maß dafür sein, wie stark der jeweilige Dokumentbereich **auf dem Originaldokument aus dem Anhang** Ihre Aufmerksamkeit auf den ersten Blick anziehen würde, bzw. in welcher Reihenfolge Sie die Bereiche bei einem ersten Blick betrachten würden.

Bitte zurück an:  
 Angela Schreyer, ZT IK 5, Otto-Hahn-Ring 6, 81730 München (Mch P, 10-399), Tel. 089/636-42283

**Frage 2**




Nun geht es um die Eigenschaften der Dokumentbereiche, die zu deren Attraktivität beitragen.

Wir werden Sie im folgenden fragen, wie stark verschiedene Eigenschaften Ihrer Meinung nach die Aufmerksamkeit eines Betrachters anziehen. Dabei werden wir als Anhaltspunkt Beispiele für die Anwendung von Eigenschaften zeigen, wir bitten Sie jedoch, für Ihre Antworten **auch Ihre eigene Erfahrung** bei der spontanen Wahrnehmung von Dokumenten zu berücksichtigen.

**Frage 2.1**

Wir nehmen an, daß Blöcke mit unterschiedlichem Erscheinungsbild die Aufmerksamkeit eines Betrachters unterschiedlich stark anziehen.

Wie stark fällt Ihnen ein Block - etwa der innere Block im unten stehenden Beispiel - gegenüber dem umliegenden Text auf, wenn er eine Liniengraphik, ein Textblock oder ein Rasterbild ist? Bitte vergeben sie Werte von 1 (zieht Aufmerksamkeit sehr stark an) bis 3 (zieht Aufmerksamkeit nur wenig an), wobei Sie **jeden Wert nur ein Mal** benutzen.

|               |   |                          |
|---------------|---|--------------------------|
| Liniengraphik |    | <input type="checkbox"/> |
| Textblock     |   | <input type="checkbox"/> |
| Rasterbild    |  | <input type="checkbox"/> |

Bitte zurück an:  
 Angela Schreyer, ZT IK 5, Otto-Hahn-Ring 6, 81730 München (Mch P, 10-399), Tel. 089/636-42283

**Frage 2.2**

Wir haben nachfolgend 7 Eigenschaften bezeichnet, von denen jede mehr oder weniger zur Attraktivität eines Textbereiches beiträgt.

Bitte tragen Sie in die Kästchen neben den Beispielen jeweils eine Zahl zwischen 1 (sehr starker Beitrag zur Attraktivität) bis 7 (schwacher Beitrag zur Attraktivität) ein, wobei Sie **jede Zahl nur ein Mal** vergeben.

| Eigenschaft    | Beispiel  | Wert                     |
|----------------|---|--------------------------|
| Schriftgröße   | abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd <b>abcd</b> abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd  | <input type="checkbox"/> |
| Einrückung     | abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd   | <input type="checkbox"/> |
| Fettdruck      | abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd <b>abcd</b> abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd  | <input type="checkbox"/> |
| Freistellung   | abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd  | <input type="checkbox"/> |
| Kursivschrift  | abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd   | <input type="checkbox"/> |
| Großschreibung | abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd ABCD abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd   | <input type="checkbox"/> |
| Bullets        | abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd<br>abcd abcd abcd abcd abcd abcd | <input type="checkbox"/> |



Angela Schreyer, ZT IK 5, Otto-Hahn-Ring 6, 81730 München (Mch P, 10-399), Tel. 089/636-42283

**Frage 2.3**

Bitte zurück an:  
Unten sind 2 Eigenschaften angegeben, die die Attraktivität eines Graphikbereiches bestimmen.  
Bitte tragen Sie in die Kästchen neben den Beispielen entweder 1 (starker Beitrag zur Attraktivität) oder 2 (schwacher Beitrag zur Attraktivität) ein, wobei Sie **jeden Wert nur ein Mal** vergeben.

| Eigenschaft  | Beispiel für geringe Ausprägung | Beispiel für stärkere Ausprägung | Wert                     |
|--------------|---------------------------------|----------------------------------|--------------------------|
| Freistellung |                                 |                                  | <input type="checkbox"/> |
| Größe        |                                 |                                  | <input type="checkbox"/> |

Angela Schreyer, ZT IK 5, Otto-Hahn-Ring 6, 81730 München (Mch P, 10-399), Tel. 089/636-42283

**Frage 2.4**

Bitte zurück an:  
Für Textbereiche wurden oben die Eigenschaften Schriftgröße, Einrückung, Fettdruck, Freistellung, Kursivschrift, Großschreibung und Bullets als diejenigen angegeben, die die Aufmerksamkeit auf einen Bereich lenken.

Wenden Sie aus Ihrer eigenen Erfahrung mit Dokumenten weitere Layout-Eigenschaften für Textbereiche an?

ja  nein

Wenn ja, welche und mit welcher Bewertung?.....

.....

.....

.....

Für Graphikbereiche wurden oben die Eigenschaften Freistellung und Größe als diejenigen angegeben, die die Aufmerksamkeit auf einen Bereich lenken.

Wenden Sie aus Ihrer eigenen Erfahrung mit Dokumenten weitere Layout-Eigenschaften für Graphikbereiche an?

ja  nein

Wenn ja, welche und mit welcher Bewertung?.....

.....

.....

.....

Bitte zurück an:  
Angela Schreyer, ZT IK 5, Otto-Hahn-Ring 6, 81730 München (Mich P, 10-399), Tel. 089/636-42283

### Anhang

Vuonna 1850 nimettiin kesäsi Nikolai I vahvon sisäiseen, jolla kiellettiin julkaista suomesta muuta kuin uskonnollista tai taloudellista kirjallisuutta. Saadokseen aikana oli merkittävää tekijänsä pelko, että neollisten romanssien julkaiseminen suomenkielisiä lehti-voimistusta maassa.

#### Lukuhalu suomiksi

Stjärnförnin leikkurilla rahvaan lukuhaluja varten on otettu omia lehdissä, joita esillä on kansa ei lukeut. Ruotsinkielinen *Åkerblad* oli valitun käyttöön jo 1700-luvulla. Suomalaiskielisiä lehdissä käytettiin samaa käsitteistä 1800-luvun alkupuolella monenlaisia muotoja, kuten *läsa* ja *läsa läsbok*. Näpakkä ilmasta *läsbok* on niitä ilmeisimmän peräisin samaa saksanmuotoa kuin *läsa* ja *läsbok* eli Wolmar Schilde-Kilpinen. Viipurilainen *Kansalehti* julkaisi vuonna 1845 Kälpien uutislaulun otsikolla "Lukuhalu". Sitä kerrotaan, kuinka viitassaarilainen "kyösti loimies eli toisevalainen" niin harvasta halusi saada luettavaan *Kansaan*, että pyysi kirjokohertaan jakamaan tilaamattomien *Kansaan*. *Lukuhalu* syntyi 1850-luvun alussa koki muut moniort suomenkielisiä lehdissä.

Juuri noita vuosia kuin ukinassa "loimies" Suomen kansan lukuhalu. Sitä kirjoittimien alberti niin suomen kuin ruotsinkielisissä saannalehdissä. Eimieliksi Kangasemendä kirjoitettiin puuttakirjaston menestyksestä *Helsingin Tiedotus* 9.2.1850 näin: "Huoletta siitä pimeydestä, joka meidän mailla ympäröi, on kuitenkin ilahduttavaa, että rahvaassa ilmenee paha lukuhalu. Todistuksena tästä haluan mainita, että pitipankirjoitukseen kotoolema ryhjä menneiden jouluun ajaksi niin kokonaan, että n. 80 kirjasta ei yksikään ei edes kettujen pyydystyksestä kettova, jäättyä hainamatta, ja että moni laitsaan hielukas ei koptu sanut kirja. Etä tästä kirjoje ei pelkätään lainata, vaan myös laistaa ja käyvästi, siitä rodisavat stur ja kanteet, jotka ovat jo kulanee sen väkään vuoden aikana, jota laisuus on ollut käynnissä."

Kuin äläkerrä Krimin sota (1854 - 56) ja sen tapahumar Suomen rannikoilla herättivät suomenkielisenä main uutislaulun ja *Suomenmaa* leikki

### K u r u h a l u.

**Stjärnförnin pitäjässä ajun teöhä loimies eli itfeolempainen, jolla on monta lastafia, ja joita le-**

W. Schilde-Kilpinen kirjoitti rahvaanlehen lukuhaluista viipurilaiseen *Kansaan*-lehteen 1845.

kohosi moninkertaiseksi. Se tunsi vahvisavan olerusta kansan lukuhaluista. Myös Sowellman oallireit lukuhalukundeun vuonna 1855 aannalla käsitteellä yleistä ja vaativan määrätelmän, tennin ruotoksi: "Jos löydetään, mitä lukuhalu on, niin siihen voi vastata se on halu herätyksen jalostumiseen, halu väliä ja ylevyyttä, seurasella ja oppoa henkilöillä, jotka ajattelevat ja tuovat todevi ja joloti. Sillä sille, joka lukee, aravuu mahdolllisuus kansakäytteen ihmiskunnan kaikkein etevimpien perisonallisuksien kanssa, ja vielä enemmän: häntä ei kuremua vain heidän viokirja ja heikkouksiensa roidatavaksi vaan heidän launteimpien heikentä, kun pelkäävät yleistä ajatukse; ja tuoneet vallitsevat heidän rinnassa. Hänet kutsutaan tunnetmaan koko ihmiskunnan parimpi olemus."

Kirjastojen perustamisesta lukemiseen johdettavista otelista tuli sarytalle moralainen velvoitus. Kansanvalistustoiminnassa vasiinaiassa tositajista oli usein nuoret apulais- ja sijapaipit, joita monet ohivat osallistua usean kirjaston perustamiseen eri paikkakunnilla. Hyvä esimerkki tällaisista ihmismielisistä apulaisista oli August Cygnæus (1818 - 86). Hän ehti panua alalle kirjastoja ainakin Vaapotin veskiään. Puolmaalain, Peisjävedelle ja Svanrammalle. Sen lisäksi hän propagoi kirjastojen ja koulujen puolesta ainoamielikirjoituksia sekä suullisesti osakunnissa ja monilla reikillä eri puolilla Perivon hiippakuntaa. Ollessaan Peisjävedellä kappalaisen sijaisena 1856 - 57 hän kirjoitti seuraavaan luuhalukirjaan kuin lähtölauseen: "Nöterandighe af Lains Kirjaso", hinkokirjaston välttämättömyys. Säätyläisen - hän oli Petarin luterilaisen puigan Zachas Cygnæuksen poika ja Fredrik Cygnæuksen velti - August ajatteli tuoksiki, jona hänen muistina pannaan ovat ueti valiteluista sekakieki.

#### Lukuhalu ja kirjastot

Oletus kansassa heikentyneen oduttavasta lukuhalusta antoi oikeutuksen laajentamiselle kirjastojen perustamiselle. Kirjastojen perustaminen rahvaan taitpein oli alkanut jo 1800-luvulla. Tunnettu viipurilainen kauppiaspöytäkirja Johan Pynninen toiminta. Hänen arkistokolehtaan kirjastojen perustamiseen puo-



Wolmar Stjärnförnin Schilde-Kilpinen.

#### Tilastoitu lukuhalu

Enemmän tai vähemmän selvätarotitusti kumpuava lukuhalu tai sen kanssa käpälävät runakkauskäsitteet, kuten tie-önnhalu, olivat niin voimakkaita, että ne oerettiin myös ideologisen käytön. Oletus lukuhalun olemassaolosta oli myös Sowellmanin ja fennomaanien kansainvälisessä ajattelun ja ohjelman oleellinen osa. Kirjastojen perustaminen ja kansainvälisyyden tuostaminen olivat kannollisista tehovi. Menestymistä tässä taistelussa

## 2. TIETÄMYKSEN MUODOSTAMINEN TIETOKANNOISTA

### 2.2 Määritelmä

Frawley et al. määrittelevät [FPM91] tietämyksen muodostamisen aiemmin tunnettomana ja mahdollisesti hyödyllisen tiedon epätriviaaliksi uuttamiseksi joukosta faktoja. Olkoon annettu data eli joukko faktoja  $F$ , kieli  $L$  ja luotettavuusmitta  $C$ . Tällöin *hahmo* (pattern) määritellään  $L$ -kieliseksi lauseeksi  $S$ , joka kuvaa joukon  $F$  osajoukon  $F_S$  alkioiden keskinäisiä suhteita luotettavuudella  $c$  siten, että  $S$  on jollakin tavalla yksinkertaisempi kuin kaikkien joukon  $F_S$  alkioiden luettelo. Hahmoa, joka on käyttäjän mielestä mielenkiintoinen ja riittävän luotettava, kutsutaan *tietämykseksi* (knowledge). Tietämyksenmuodostamisjärjestelmän uuttamaa tietämystä kutsutaan *muodostetuksi tietämykseksi* (discovered knowledge). Tietämyksen muodostaminen tietokannoista on kyseessä silloin, kun data sijaitsee tietokannassa.

Määritelmästä johdettuna ja lisätynä Frawley et al. asettavat tietämyksenmuodostamisjärjestelmälle seuraavat vaatimukset:

- Tulos on mielenkiintoinen.
- Tulos on täsmällinen.
- Tulos esitetään korkean tason kielellä.
- Järjestelmä on tehokas.

Tulos on mielenkiintoinen, kun se on uutta ja käyttökelpoista informaatiota, joka auttaa käyttäjää saavuttamaan tavoitteensa. Esimerkiksi keskiarvo voi olla aiemmin tuntematonta informaatiota, mutta sen laskeminen on triviaalia, joten keskiarvo ei ole tietämyksenmuodostusnäkökulmasta mielenkiintoinen tulos. Tulos on täsmällinen, kun se kuvaa kohdealuetta tarkasti ja luotettavasti. Luotettavuusmitta kertoo käyttäjälle tai järjestelmälle itselleen, kuinka täsmällinen tai uskottava tulos on. Korkean tason kielellä esitetty tulos on tulkitsematta ihmisen ymmärrettävissä. Tietämyksenmuodostamisalgoritmi on tehokas, kun sen aika- ja tilavaatimus on syötteen koon polynomifunktio,

PAULIINA SUSI

## Kohtaamisen koulussa

*Rakkaus ja hyväksyntä ovat lapsen oikeuksia, myös maahanmuuttajalapsen. Koululuokat ja päivähoitoryhmät kansainvälistyvät kovaa vauhtia. Monikulttuurisuudessa on haastetta päiväkotien ja koulujen opettajille.*

**L**ehroni Marjatta Kalliala opettaa tuleville lastentarhanopettajille kulttuurista orientaatiota, oman ja vieraiden kulttuurien kohtaamista päivähoitoyössä. Lehroni Pirkko Leino kohtaa samat ongelmat opettaessaan suomen kielen opettamista toisena vieraana kieltenä luokanopettajille. Molemmat työskentelevät ylitöistä Opettajankoulutuslaitoksella.

Koulussa ja päiväkodissa työskentelevä ei voi ajatella suvaitsevaisuutta kaukaa ja vain periaatteellisesti. Suhde lapseen ja hänen perheeseensä on aina vaativia. Sitä vaativampi mitä lähemmäs tullaan ja mitä vieraampi lapsen elinympäristö on. Jokainen kuitenkin kaipaa rakastuksi, hyväksyntä ja kohdatuksi tulemisen tunnetta. Kallialan mukaan lastentarhanopettajia tarvitsee ennen kaikkea halua tietää ja ymmärtää. Leino puhuu empatiasta ja älyllisestä, tulkivasta mielestä.

– Opettajan on tunnettava ensin oma kulttuurinsa, sitten muutakin kulttuureja, jotta hän voisi ymmärtää ja auttaa muualta tulevaa oppilastaan. Kulttuuri on käsiteltävä hyvin avarasti: se on yhtä hyvin musiikkia tai kirjallisuutta kuin triviaaleja käytännön selkkoja, kuten miten esiintydään, miten WC:ssä käytäydään tai mitkä ovet pidetään kiinni ja mitkä auki. Ongelmana on, että viiteentoista opintoviikkoon ei voi mahduttaa tiedon hankkimista omasta kulttuurista, vaan se jää opiskelijan omille harteille, Pirkko Leino painottelee.

### Asenteet tiukuvat läpi

Kallialan kurssi on otettu hanakasti vastaan. Asenteita ei voi kuitenkaan siirtää kuten faktaa. Opiskelijat arvioivat, kun englantilainen vierailija ollut professori puhui liikkua saarnaten: ”Luuleeko uo, että minne pitää



– Opettaja kasvaa tehtävänsä mukana, Pirkko Leino uskoo.

sanoa, että älkää olko rasisteja”, nuristiin. – Enme voi luottaa siihen, että hyvät asiat tapahtuvat itseksensä. Sellaiseen näivitteeseen meillä ei ole oikeutta, Kalliala sanoo. Hänen tavoitteenaan on saada opiskelijat tietoiseksi omasta vaikutuksestaan.

– Anmatillisuus auttaa tasapuolisuuteen, mutta tuntee, että ei voi käskä. Asenteemme tiukuvat meistä koko ajan, siksi pitää skampata.

Päiväkodissa työskentelevän on oltava älyllisesti ja emotionaalisesti rehellinen itselleen ja uskallettava kohtaa myös omat emakkoluulonsa, ärymyksensä ja väärinymmärryksensä. Niitä on opeteltava työsiämään yksin ja työyhteisönsä kanssa.

### Pois pinnasängystä ja maksalaatikosta

Vierasmaalaisia lapsia ei enää pyritä sulauttamaan suo-

## Negaatio ja logiikkaohjelmointi

Definiitit ohjelmat ilmaisevat pelkästään positiivista tietämystä. Rajoitetusta syntaksista johdun niillä ei voida suoraan ilmaista edes yksinkertaisinta negatiivisen tiedon muotoa — atomikaavan negaatiota. Jatkossa käytämme nimityksiä *positiivinen* ja *negatiivinen literaali* tavallisista atomikaavoista ja niiden negaatioista, ja nimeä *literaali* tarkoitteessamme jompaakumpaa.

### 1. Miten negatiivinen johdetaan positiivisesta?

Miten negaation käsite voidaan lisätä definiitteihin ohjelmiin? Negatiivinen literaali  $\neg F$  ei nimittäin voi koskaan olla definiitin ohjelmaan  $F$  looginen seuraus, sillä

- Herbrandin kanta  $B_P$  on aina def. ohjelmaan  $P$  malli, ja
- Koska kaikki atomikaavat  $F$  kuuluvat joukkoon  $B_P$ , niin niin  $V_{B_P} F$  ei voi pitää paikkaansa ja  $\neg F$  on epätosi tulkinnassa  $B_P$ .

Jälkimmäisestä huomiosta seuraa suoraan seuraava 'ikävä' totuus.

LEMMA 4.1. *Olkkoon  $R$  jokin päättelymetodi (päättelysääntöjen joukko ja niiden soveltamistapa),  $P$  definiitti ohjelma ja  $A$  perusatomi. Jos  $P \vdash R \neg A$  niin  $R$  ei ole terve.*  $\square$

Etsä kiertotie edelliseen ongelmaan on päättelysääntöjen sijasta käyttäessä korkeamman tason *metapäättelysääntöjä*.

**1.1. Suljetun maailman oletus.** *Suljetun maailman oletus* (Closed World Assumption, CWA) on tietokantojen maailmasta peräisin oleva ajatus [Rei78], jonka mukaan jokin relaatio on voimassa tiettyjen objektien välillä *jos ja vain jos* ao. relaatio voidaan johtaa ohjelmasta. Malliteorian puutteissa CWA voidaan ymmärtää näkökantana, joissa huomio kiinnitetään vain joihinkin ohjelman malleihin. Intuitiivisesti nämä mallit ovat sellaisia, joissa 'jos'-ja-'vain-jos'-tulkinta on mahdollinen.

Merkitään notaatiolla  $CW A(P)$  niiden perusatomien joukkoa, joita ei voida johtaa ohjelmasta  $P$ . CWA-sääntö voidaan ilmaista seuraavanlaisena *metapäättelysääntönä*:

$$\frac{A \in CW A(P)}{\neg A} \text{CWA},$$

Sääntöä kutsutaan *metasääntöksi*, sillä sen premissi ei ole käsiteltävän kielen (pred. logiikka), vaan käsiteltävän kielen (johtamisrelaation määrittely), kaava. Logiikkaohjelmien terminologiassa ominaisuus  $A \in CW A(P)$  tarkoittaa sitä, että maallile  $A$  ei ole SLD-refutaatioita ohjelman  $P$  suhteen, eli formaalisti

$$\frac{P \not\vdash_{SLD} A}{\neg A} \text{CWA},$$

Koska ohjelmasta  $P$  johdettavissa olevien perusatomien joukko on  $M_P$ , on  $CW A(P) = B_P \setminus M_P$ .

## B Anleitung zum interaktiven Test

Bitte zurück an:

Angela Schreyer, ZT IK 5, Otto-Hahn-Ring 6, 81730 München (Mch P. 10-399), Tel. 089/636-44949

Vorname, Name: .....

Datum: .....

Beruf: .....

Alter: .....

### Liebe Teilnehmerin, lieber Teilnehmer,

vielen Dank, daß Sie sich Zeit nehmen, unser Programm zur Attraktivitätsbewertung von Dokumenten zu testen.

In unserer Forschungsarbeit zur Dokumentanalyse fragen wir uns, welche Dokumentbereiche die **Aufmerksamkeit** eines Betrachters **auf den ersten Blick** anziehen. Insbesondere interessiert uns, wie stark die Attraktivität der einzelnen Dokumentbereiche ist, wenn man davon ausgeht, daß die Attraktivität **nur** durch Merkmale des **Layouts**, und **nicht** durch den **Inhalt**, gegeben ist.

### Aufgabenbeschreibung

Sie sehen auf dem Bildschirm zwei Fenster. Im linken Fenster wird die vom Programm errechnete Bewertung des Dokuments dargestellt. Der Attraktivitätswert eines Dokumentbereichs wird dabei durch seine Farbe angezeigt. Die Farbskala reicht von gelb (niedriger Attraktivitätswert) bis rot (hoher Attraktivitätswert).

Ihre Aufgabe ist es nun, an einigen Beispielen zu beurteilen, ob die errechnete Attraktivität mit Ihrer eigenen Bewertung übereinstimmt.

Im Anhang sehen sie die fünf Beispieldokumente. Es handelt sich um Kopien aus wissenschaftlichen Veröffentlichungen, bzw. Zeitschriften. Diese Art von Dokumenten steht bei unserer Arbeit im Vordergrund.

Beherrschen Sie die Sprache, in der diese Dokumente verfaßt sind?

Ja  Nein

Wenn ja, welche Sprache ist das? .....

Zur Durchführung des Tests trennen Sie den Anhang bitte ab. Dadurch haben sie die Möglichkeit, jedes Dokument zunächst ohne die überlagerte Attraktivitätsbewertung zu betrachten.

Bitte überlegen Sie sich dann, wie stark die einzelnen Bereiche der Dokumente Ihre Aufmerksamkeit auf den ersten Blick anziehen, d.h. wie hoch ihre **Attraktivität aufgrund von Layout und typographischen Merkmalen** ist.

Sollte Ihre Bewertung von der angezeigten Bewertung abweichen, ändern sie bitte Ihrer Meinung nach nicht korrekte Attraktivitätswerte wie nachfolgend beschrieben.

Bitte zurück an:

Angela Schreyer, ZT IK 5, Otto-Hahn-Ring 6, 81730 München (Mch P. 10-399), Tel. 089/636-44949

### Ändern von Attraktivitätswerten

Markieren Sie mit der LINKEN MAUSTASTE Wörter oder Blöcke, deren Attraktivitätswert Sie verändern wollen:

- Markieren eines Wortes durch Klick in das Innere des Wortes
- Markieren eines Blocks durch Klick auf den Rand des Blocks
- Beim ersten Klick auf das Bild des Dokuments erscheinen braune Rahmen für die Wörter und Blöcke. Ein markiertes Objekt wird dagegen blau umrandet.

Verändern Sie den Attraktivitätswert eines Wortes oder eines Blocks durch Bedienen der CURSOR-TASTEN auf der Tastatur:

- Erhöhen des Attraktivitätswertes mit der noch oben zeigenden CURSOR-TASTE ↑
- Verkleinern des Attraktivitätswertes mit der nach unten zeigenden CURSOR-TASTE ↓

Die von Ihnen vorgenommenen Änderungen werden durch Drücken der RETURN-TASTE „j“ in eine Datei abgespeichert.

Anschließend erscheint im rechten Fenster ein „WEITER“-BUTTON, mit dem sie das nächste Dokument anzeigen und bearbeiten können.

## Tärkeät opetusansiot

Olen jo useaan otteeseen sattunut olemaan Yhdysvalloissa tutkimustyössä siinä vaiheessa, kun kohdalleni on osunut vuorota pääkirjoituksesta. Hieman pitempiäkin oleskelun yhteydessä tulee vaihtaneeksi varsin perusteellisesti ajatuksia amerikkalaisten kollegoiden kanssa maiden yliopistojärjestelmien eroista. Kerran puhueenaiheeksi nousi isäntälaitokseni erään apulaisprofessorin tapaus, jossa tämä oli vähintään normaalina pidettyjen virkavuosien jälkeen anonut ylennystä "täydeksi" professoriksi. Edellyksenä oli hyvä tutkimus- ja tutkimusjohtaminen, johon pakosti kuuluu myös kansainvälinen yhteistyö, sekä hallintotehtävien osallistuminen. Nämä suoritukset eivät olleet kompastuskivenä, mutta sellaiseksi muodostui kolmas edellytys eli menestyminen opetustehävissä. Asiantomaisella ei ollut tältä alueelta juurikaan näytöjä. Ylennys jäi saamatta, ja viestini oli, että uusintaa on turha edes yrittää ilman selkeitä kohennusta opetusansioissa.

Käymäni keskustelu sai aikaan monenmoista mielennyllerrystä, kun yritin suhteuttaa tapausta meillä vallitsevaan tilanteeseen. Järjestökuvioiden eniten huolta opetukseen parnostamisen merkityksestä on kantanut Yliopistonlehtöiden liitto. Muiden yliopistojärjestöjen piirissä opetustehävät ovat saaneet ilmeisesti eniten huomiota ns. kokonaistyöaikakuvion puitteissa, sillä pelkona on ollut, että yliopistot yrittävät määrärahanuukkuuden valitessa paikat opetusresurssissa koettua vajasta lisäämällä ensimmäinkin opetustehävien osuutta varsinaisen opetushenkilöstön piirissä, sekä toisaalta laajentamalla opetus sellaisen henkilöstöryhmien toimenkuvaa, joita ei ole luettu varsinaiseen opetushenkilöstöön. Jälkimmäisiä ryhmiä ovat ainakin tutkijat ja kirjastohenkilöstö. Nämä pelot eivät suinkaan ole aiheettomia, ja mainitut trendit ovat jo havaittavissa.

Sanomattakin lienee selvää, että ymmärrän järjestöpuolen pelot opetustehävien kohtuuttomasta lisääntymisestä, etenkin nykyisessä

vaikeassa määrärahatilanteessa. Alussa mainitsenani tapausten viesti kiteytyykin siihen, että opetuksena tunnettu toiminto on joka tapauksessa yksi yliopistojen kaikkein keskeisimmistä tehtävistä. Tieteentekijöiden liitto on hellinnyt ajatusta siitä, että meillemkin tulisi saada aikaan yliopistojen opetushenkilöstön virkajärjestelmä, jossa joillakin edellytyksillä urakehitys riippuu enemmän yksilön suorituksista kuin (sattumoisin) avautuvista virkaposteista. Jos rälläinen tilanne joskus tulee todellisuudeksi, on - tai ainakin tulisi olla - edellytyksenä, että yksi henkilön suorituksen arvioinnin osa liittyy opetukseen. Ise asiassa nykyisenkin järjestelmän puitteissa opetusansioille on saatettu antaa huomattava merkitys nimitysprosessissa.

Jos ja kun opetusansiot tulevat aiempaa merkittävämpään osaan yliopistoväen urakehityksen edellytyksinä, saattaa hyvin tulla tutkimustehävien väkiseisyys yllättävällä tavalla kostaantua. Tästä syystä saattaa tontekijän näkökulmasta olla lyhytnäköistä luoda niin korostetun tutkimus- ja tutkimustehävien väkiseisyyden vaikutuksia kuin mitä jokien aikojen julkistetuissa tutkimusurien kehittämissuunnitelmissa on kaavailtu. Ellei sitten ryhmän ehdotuksessa on kaavailtu. Ellei sitten olla todella valmiita luomaan opetusvirkaehtojen rinnalle selkeitä vaihtoehtoja työelämässä vietetyn ajan mittaisesta tutkimusurasta. Tällaista valmiutta on kuitenkin syytä epäillä. Niiden, jotka huvittelevat yliopistouraa, on varauduttava ennemmin tai myöhemmin kilpailemaan tutkijapettajainviroista. Jo yksin tässä suuressa on kuitenkin syytä epäillä. Niiden, jotka huvittelevat yliopistouraa, on varauduttava ennemmin tai myöhemmin kilpailemaan tutkijapettajainviroista. Jo yksin tässä suuressa on kuitenkin syytä epäillä. Niiden, jotka huvittelevat yliopistouraa, on varauduttava ennemmin tai myöhemmin kilpailemaan tutkijapettajainviroista. Jo yksin tässä suuressa on kuitenkin syytä epäillä.



Tällainen järjestely saattaisi olla itseasiassa myös tärkeä "overhead"-rahan vastine. Eritoten olennaisista tällaisissa järjestelyissä olisi kuitenkin, että opetus on todella osa työtehtäviä, ei jokin palkaton työllisyys, joka ei mitenkään näy arvioitavissa tutkimustyön vaatimissa aikalta.

Kari Pitkänen

Bitte zurück an:

Angela Schreyer, ZT IK 5, Otto-Hahn-Ring 6, 81750 München (Mich. P. 10-399), Tel. 089/636-44949

Anhang

## Predikaattilogiikka

Predikaattilogiikka [CL73] on *deklaratiivisen tietämyksen esittämissymbolisminä*. Sana 'deklaratiivinen' on peräisin lingvistiikasta, jossa deklaratiivinen lause on lause, joka voi olla joko tosi tai epätosi (vrt. kysymykset ja käskyt). Predikaattilogiikka kykenee ilmaisemaan vain deklaratiivista tietämystä. Tietämys esitetään väittämällä, että tietyt *relaatiot* (tai niiden looginen yhdelmä) ovat tai eivät ole voimassa joidenkin tarkastelujoukkomme (maailma, universumi) *yksilöiden* välillä.

Määritteliemme aluksi predikaattilogiikan syntaksin ja (denotatiivisen) semantiiikan.

### 1. Aakkosto

**1.1. Mitä aakkostolta vaaditaan.** Predikaattilogiikan kielen aakkoston symbolien tulee kyetä esittämään ainakin

- tarkastelujoukkomme yksilöitä (objekteja) ja
- niiden välisiä relaatioita.

Yksilöitä edustavia symboleja kutsutaan *vakioiksi* ja relaatioita edustavia *predikaattisymboleiksi*. Koska eri relaatiot saattavat kohdistua erisuuruiseen määrään yksilöitä, liittyy jokaiseen predikaattisymboliin sen *ariteetti*. Ariteetti on luonnollinen luku, joka kertoo, kuinka monta argumenttia relaatioon kuuluu.

Predikaatti- ja vakiosymbolien eroteluun tarvitaan jottaakin *tsäsymboleita*; va- litsemme niiksi sulut ja pilkku. Tietämys "Miuku inhoaa Mustia" voidaan nyt formalisoida muodossa *inhoa(miuku,musti)*.

Miten mallintaa väittämä "kaikki kissat inhoavat koiria" lyhyemmin kuin huet- telemalla kaikki kissa-koira -parit? Tämä onnistuu laajentamalla aakkostamme *muuttujilla* ja *kvanttorilla*. Muuttujalla esitetään tuntematonta (mutta kuitenkin yksikäsitteistä) objektia. Kvanttorit liittyvät aina johonkin muuttujaan:

- *universaalkvanttorilla*  $\forall$  ilmaistaan, että jokin relaatio (tai useamman yhdelmä) on voimassa kaikilla mahdollisilla oo. kvanttorin liittyvän muuttujan arvoilla, ja
- *eksistentiaalkvanttorilla*  $\exists$  ilmaistaan, että jokin relaatio on voimassa jollakin (sis ainakin yhdellä) mahdollisella muuttujan arvolla.

Yksittäisen relaation avulla ei voida mallintaa (ainakaan luonnollisella tavalla) monimutkaisempia luonnollisen kielen väittämiä. Miten esim. ilmaistaisiin tieto siitä, että jokin relaatio *ei* ole voimassa; tai että relaatio *p* on aina voimassa kunhan relaatio *q* on. Aakkostamme pitääkin laajentaa *loogisilla konnektiiveilla*. Kuvassa 1 on esitelty yleisimmät käytetyt konnektiivit. Konnektiivi  $\neg$  on unaarinen, ja sitä käytetään prefix-operaattorina (so. argumenttinsa edessä), muut ovat binäärisiä ja niitä käytetään infix-muotoisesti. Mahdolliset epäselvyydet vältetään sulkujen avulla.

Maaillamme objektit ovat usein *rakenteisia*, ja väittämät perustuvat joskus koko objektin sijasta johonkin (tai joihinkin) sen alirakenteisiin. Vakioiden lisäksi

# JOURNALISMIN MONIMUTKAINEN ARKKI



Pieni kierros eri alojen ihmisten parissa paljastaa, että journalismissa vallitsevat lait ja säännöt ovat monimutkaisemman kehityksen tulosta: alan toimintatavat eivät selity pelkästään nais- mies-akselilla.

Tiedostaja Olli Pärnänen Suomen Pääasesta Rististä sanoo aiheen olleen mielestä erityisesti Jugoslavian sodan takia: "Kului vuosi puolitoista ennen kuin kirjoittajille alkoi olla muuta kuin sotajälkitilanteen raportointia. Vasta vankileirien avaututtua alkoi paljastua siviileihin kohdistuneet kauheudet."

Kuitenkin esimerkiksi muslimien raskaat unohduttavat osterit koostuivat paan. "Eletään uutisen armolla ja uutiseen tarvitaan aina jotain uutta. Tyypillisten ihmisten elämä ei ole uutisarvoista."

Vuote aikana ovat Ruusudan tapahtumat olleet kiihottavampia. Pärnä- nen arvoin mukaan meitä ei ollut ollut niin välillä ettei stella olisi ollut eu- rooppalaisia. "Meidän tiedonvälitys- kulttuurissamme aiheen pitää jollain lailla liittyä Eurooppaan ja eurooppa-

laisiin ennen kuin se on raportoin- sen arvoista."

Toistamiseen valmentä minikään uutisarvoa. Bangladeshin kaksi ker- taa vudessa tuostavat tulvat ovat ny- kyisin rutiinjuttu. Luontokatastro- fin mielenkiinto pummitaan sen ma- kaan kunakin muutama kymmeniä tuhatta ihmistä se vahingoittaa.

Pärnänen ei usko, että yksi nainen tekisi kesän, tarvitaan suurempi jouk- ko, porukka, muutamaosaan välttämä- Nainen pitää kättä kättäkin ei yhdessä jössä muuta mitään, tämä sakupolvi on sisästänyt nykyiset uutiskriteerit. "Tämä on kasvatus- ja kulttuurisoidin- nainen jutu..."

# Uutta arvoa opetusansioille

*Yliopistojen toiminnalle on tunnusomaista tieteellinen tutkimus ja siihen perustuva opetus. Niiden avulla välitetään uutta tietoa opiskelijoille ja koko tiedeyhteisölle, ja ammattikorkeakouluilla ja muilla oppilaitoksilla on vastaavasti myös ammatillisen tutkimuksen tehtävä.*

Yliopistolaitoksen rakenteen ja toimintatapojen muutokset ovat lähentäneet tutkimusta ja opetusta. Monet opettajat ovat lisänneet tutkimustyötään ja moni tutkija on vastaavasti alkanut opettaa.

Akateemista uraa ajatellen tätä kehitystä on kuitenkin häirinnyt opetuksen vähämpi arvostus tutkimukseen nähden. Tästä johtuen on usein päätely tutkimuksen edistävän, mutta opetuksen hidastavan tieteellisellä uralla etenemisestä, ja valittu tämän mukaisesti työtehtävät. Tilanne on kuitenkin toinen koko yliopiston kannalta, sillä ilman tasokasta opetusta tieteellinen tieto välittyy heikosti opiskelijoille. He pettyvät ja yliopiston voimavaroja määrittelevät tutkimustavoitteetkin jäävät täyttymättä. Opiskelijat saavat enemmän hyvältä tutkijalta, jos hän on myös hyvä opettaja. Yliopiston kaikkien opetusvirtojen täytössä kannattaa siis ottaa huomioon opetusansiot.

## Opetusansioiden ottaminen huomioon

Opetustehävien asema vaihtelee henkilöstöryhmän mukaan. Lehtorit muodostavat keskeiset opettajaryhmän, jolla opetusansiot vaikuttavat jo jossain määrin ammattiarvon kehitykseen. Sen sijaan tutkimustyön lisääminen tuottaa heille usein työnjako-ongelmia, vaikka se muuten tukeekin opettajan uralla etenemistä. Myös viranmääritys, käytetty kriteeristö vaihtelee usein tilanteiden mukaan, joten se antaa vain viitteellisen perspektiivin opetukselliselle pätevyymiselle. Muita opetusvirkoja täytenä käytössä on vielä enemmän horjuttavaa.

Ainakin Jyväskylän ja Oulun yliopistoissa on

hijattain täsmennetty opetusansioihin liittyviä menettelyä niin käytön kuin sisällönkin suhteen.

Opetusansiot otetaan huomioon opetusvirkojen täytössä tieteellisen pätevyyden nimellä. Uusi menettely koskee sekä väitöksiä että professuureja. Oleellista on se, että opetusansioiden merkitys ilmaistaan se, että hakulomakkeissa ja että tutkimus- opetus- ja muun toiminnan arviointikriteerit suhteutetaan etukäteen keskenään. Tutkimus- ja muut ansiot esitetään vakautuneen tavun mukaisesti mm. julkaisun ja ansioleuelloin.

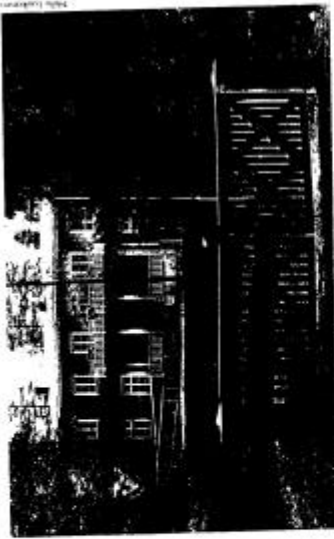
Opetusansiot esitetään tähän tarkoitukseen sopivan portfolion avulla, joka on henkilön itsensä kokonaan ”työnajassaikkuri” tai opetusansioita painottava ansioleuelloin osa.

Portfolioon kootaan opetusansiot itse kirjoittavien opettajien ja valitsijoidenkin vaivoja ajatellen. Sinne kuuluvat ainakin opetusansioita edistävät opinnot, opetuksen kehittämistyö, oppimateriaalin valmistaminen, opetuskokeemus, opintojen ja oppimäyten ohjaaminen sekä mahdolliset opetusansioista saadut arvot ja tunnukset. Valitsija laati näistä opetusansioista yhteenvetoon, jota käytetään asiointuntijalausunnon tapaan yhdessä valmistajineen.

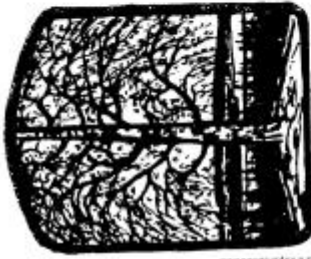
## Opetusansioiden hankinta

Korkeakoulupedagoginen kehitys ja opetusansiojen täytävyyden monipuolistuminen lisäävät opetusansioiden merkitystä ja käyttöä. Yksilön kannalta niiden hankkiminen on suunnitelmallisempaa ja monipuolisempaa sekä merkitys valintakriteerinä keskeisempää kuin aikaisemmin. Samalla opetuksen satsaaminen edistää koko yliopiston tavoitteiden saavuttamista ja osoittaa opiskelijoille sen, että heidän opintojensa edistymisestä kannetaan myös vastuu.

Tällainen opetusansioiden arvonnousu tuo myös ongelmia erityisesti niille, jotka ovat uransa alkutaipaleilla loppuvaiheessa hakeutuessaan opetusvirkaan. Alusta opetustehävät paineilevat useammin leivän kuin opetusansioiden kokoaamista, jolloin voi vähemmän valikoitua tarjontaan suhteen. Mikäli opetuksen sisältö jää etäälle



Akkemiin yleisöllä onnoso Aholassa Järvenpäässä on pyryvä nilymy Juhani Aholasta ja Venny Saldon Bröfölinin.



Venny Saldon Bröfölinin pirtistä viennyt Juhani Ahol on nilymy Aholan teoston nimikohdella.

lisa oiretta vliereitin hyvin tylin sootokasta. Osa Aholan lyraista proosa onkin lehtemästä, sitoin kuin Linnankosken. Murta Aholan kielien parille on loylyyt uusi tehtävä kirjallisuutemme kehityksessä: väitöskirjassaan Juhani Aholan kirjallisuus ja usoitaminen: etiä juuri Aholan lausujen rytmien suoraan esiintyminen ja parhaimmista avoista tiede vapautymiselle rannolle muodostamalla eräänlaisen sillan puoiaan ja lyriikan välille.

Sama havainto koskee muuten Volter Kilven proosaa. Täsi on harvinaisen yhtymäköitä Aholan ja Kilven välillä ja ehiä kehitys siihen, mikä juuri Kalpi antoi 1911 Aholan leuelle voimakasta tunnususta: se on hänessä ”niin hienoa ja kuulhaasta, etiä se välittää miellään herkimänsäkin vivahten”.

Kilpi jouti Aholaa arkeina prooaryylin kehittämistä, hänessä ”kantaa nykyinen kielimehi Juhani Aholan leimaa”. Myrsköhermian Lauri Viljanen lähdeää se maia, ja vastaavaan tulokseen päätyy Juhani Niemi. Ahol-kirjasta loppuvaiheessa: Aholan normaalioproosa on luonnosta ”yhä hyvää nykyäänään. Aholin kieli on sullyyt järkeväsi luettavana”. Toimittajien voi itse kukaan etsiä vastaavalla jorhan Aholan leuelloin: sen heeli voi olla paikoitin vanhahtavaa, mutta se on drossa vuodeska hammasystävään vahan vanhentamista.

**Luonnon arvo.** Viidettä Aholan vahoja piltteeri eivat vastustajekkaan koskaan

ole voinnet kieltä: hänen luonnontutkimus- ja kielitieteellisen tutkimuksen lisäksi hänellä on ollut merkittävä rooli kirjallisuuden tutkimuksessa. Hän on ollut mukana myös kirjallisuuden tutkimuksen ja kielitieteiden tutkimuksen parissa. Hän on ollut mukana myös kirjallisuuden tutkimuksen ja kielitieteiden tutkimuksen parissa. Hän on ollut mukana myös kirjallisuuden tutkimuksen ja kielitieteiden tutkimuksen parissa.

Alto puolistaa luonnon henkellistä ja aistillista, avoimella likään kuin meidän vianan tai veden rajan tai avoimella kulkulata laian rakoon. Sama in- tääni lämpö välittise myös hänen ih- miskaavauksessaan, sunalla tavoin elä- ryen hän puolesta maan hiljaisia elä- mäntutkimusta tappoille häviä beki- kundan, oli sitten kysy vuoden 1918 ritaridissa tai yksinäisyään papin rouvan kohtalosta. Humaanit elä- mäntutkimus sekä äänisen ja luon- non tasapainon tutkimisen merkitys on se väsi, joka Aholan tuotannossa vä- litetty myös meidän ajallamme. —


Kari Laitinen on kutsunut kirjallisu- den suurtautoprofessoria ja Juhani Aholan senne vuorokautena.



## **C Testdokumente**

Die Dokumentanalyse-Gruppe bei der Siemens AG, ZT IK 5, unterhält eine eigene Datenbank von Dokumenten, die aus wissenschaftlichen Zeitschriften und Veröffentlichungen mit einer Auflösung von 300 dpi gescannt worden sind. Aus dieser Datenbank wurden zu Testzwecken 10 Dokumente mit möglichst verschiedenem Layout und in drei verschiedenen Sprachen (Finnisch, Deutsch, Englisch) ausgewählt. Die Dokumente sind im folgenden verkleinert dargestellt. Neben ihrer Bezeichnung innerhalb dieser Arbeit ist jeweils ihre Bezeichnung in der Datenbank in Klammern angegeben.



|  |   |
|--|---|
| <p>T3) (maat.tif)</p>  | <p>T4) (z0107.tif)</p>  |
| <p>2. TIETÄMYKSEN MUODOSTAMINEN TIETOKANNOISTA</p> <p><b>2.2 Määritelmä</b></p> <p>Frawley et al. määrittelevät [PPM91] tietämyksen muodostamisen aiemmin tunnetutoman ja mahdollisesti hyödyllisen tiedon epätriviaaliksi uuttamiseksi joukosta faktoja. Olkoon annettu data eli joukko faktoja <math>F</math>, kieli <math>L</math> ja luotettavuusmitta <math>C</math>. Tallon <math>hahmo</math> (pattern) määritellään <math>L</math>-kieliseksi lauseeksi <math>S</math>, joka kuvaa joukon <math>F</math> osajoukon <math>F_S</math> alkioiden keskinäisiä suhteita luotettavuudella <math>c</math> siten, että <math>S</math> on jollakin tavalla yksinkertaisempi kuin kaikkien joukon <math>F_S</math> alkioiden luettelo. Hämmoa, joka on käyttäjän mielestä mielenkiintoinen ja riittävän luotettava, kutsutaan <i>tietämyksekiksi</i> (knowledge). Tietämyksenmuodostamisjärjestelmän uuttamaa tietämystä kutsutaan <i>muodostetuksi tietämyksekiksi</i> (discovered knowledge). Tietämyksen muodostaminen tietokannoista on kysessä silloin, kun data sijaitsee tietokannassa.</p> <p>Määritelmistä johdettuna ja lisäytynä Frawley et al. asettavat tietämyksenmuodostamisjärjestelmälle seuraavat vaatimukset:</p> <ul style="list-style-type: none"> <li>• Tulos on mielenkiintoinen.</li> <li>• Tulos on täsmällinen.</li> <li>• Tulos esitetään korkean tason kielellä.</li> <li>• Järjestelmä on tehokas.</li> </ul> <p>Tulos on mielenkiintoinen, kun se on uutta ja käyttökelpoista informaatiota, joka auttaa käyttäjää saavuttamaan tavoitteensa. Esimerkiksi keskiarvo voi olla aiemmin tuntematonta informaatiota, mutta sen laskeminen on triviaalia, joten keskiarvo ei ole tietämyksenmuodostusnäkökulmasta mielenkiintoinen tulos. Tulos on täsmällinen, kun se kuvaa kohdealuetta tarkasti ja luotettavasti. Luotettavuusmitta kertoo käyttäjälle tai järjestelmälle itselleen, kuinka täsmällinen tai uskottava tulos on. Korkean tason kielellä esitetty tulos on tulkittamatta ihmisen ymmärrettävissä. Tietämyksenmuodostamisalgoritmi on tehokas, kun sen aika- ja tilavaatimus on syöteen koon polynomifunktio,</p> |  <p>Informatik-Spektium (1985) N. 273-282</p> <p><b>Verfahren und Systeme der Computertomographie</b></p> <p><b>Röntgen - Magnetische Resonanz - Ultraschall</b></p> <p>E. Hündt und G. Schwierz<br/>Siemens AG, München</p> <p><b>Zusammenfassung.</b> Die bildgebenden Systeme der medizinischen Diagnostik erfahren zur Zeit eine sehr rasche und spektakuläre Entwicklung durch die Digitalisierung der Meßsignale und die rechnergestützte, numerische Verarbeitung. Röntgen, Magnetische Resonanz und Ultraschall ermöglichen drei Systeme der Computertomographie, die Schnittbilder des zu untersuchenden Objektes mit hoher Qualität, Auflösung und Aussagekraft liefern. In diesem Beitrag werden die physikalischen Prinzipien, die Algorithmen zur Rekonstruktion, die Meßsysteme sowie der datentechnische Aufbau erläutert.</p> <p><b>Schlüsselwörter:</b> Bildgebende Verfahren, Computertomographie, Bildrekonstruktion, Medizinische Diagnostik</p> <p><b>Summary.</b> Imaging systems of medical diagnostics are developing very quickly and spectacularly which is caused by the digitization of measuring signals and computerized numerical processing. X-ray, magnetic resonance and ultrasound enable three systems of computed tomography which produce sectional views of the object being of high quality, high resolution and high diagnostic power. The paper discusses the physical principles, the algorithms for reconstruction, the measuring systems and the processing performance.</p> <p><b>Key words:</b> Imaging systems, Computerized tomography, Image reconstruction, Medical diagnostics</p> <p><b>Computing Reviews Classification:</b> J.3</p> <hr/> <p><b>1. Einleitung</b></p> <p>Die bildgebenden Systeme der medizinischen Diagnostik erfahren zur Zeit eine sehr rasche und spektakuläre Entwicklung. Einerseits finden neue physikalische Prin-</p> <p>zipien, beispielsweise die Magnetische Resonanz, Eingang in die klinische Praxis, andererseits bieten tomographische Methoden neue diagnostische Aussagemöglichkeiten. Von großer Bedeutung für diese Entwicklungen ist die Digitalisierung der Meßsignale sowie die Speicherung und in zunehmendem Maße die rechnergestützte Verarbeitung von Bildern. Die Mikroelektronik mit ihrem leistungsfähigen Prozessoren und ihren hochintegrierten Schaltungen macht diese Entwicklung technisch erst möglich und trutz der großen Dimensionen auch wirtschaftlich realisierbar.</p> <p>Tomographie bedeutet Abbildung einer dünnen Schicht des dreidimensionalen Körpers. Durch geeignete Anordnung von Sender und Detektor - i. a. rings um das zu untersuchende Objekt - werden nur Signale aus der gewünschten Schicht gewonnen und durch numerische Verfahren mit Hilfe des Computers zu einem Bild kombiniert. Objekte außerhalb der abzubildenden Schicht geben keine Signale und erzeugen damit auch keine Störungen im Bild. Im Gegensatz hierzu wird mit konventioneller Röntgentechnik im allgemeinen eine Projektion erstellt, bei der Objekte aus verschiedenen Tiefen gleichzeitig im Bild erscheinen. Sie können sich gegenseitig überdecken und wesentliche Information verbergen.</p> <p>Die Röntgen-Computertomographie war das erste System der bildgebenden medizinischen Technik, bei dem konsequent alle Meßwerte digitalisiert und rein digital weiterverarbeitet werden. Durch geeignete Abtastung und numerische Rekonstruktion liefert das Verfahren Schnittbilder, die sich durch hohe Kontrastauflösung ohne Überlagerung durch Fremdstrukturen auszeichnen. Es ist insbesondere bei der Darstellung von Weichteilgeweben wie Gehirn und inneren Organen zur Differenzierung pathologischer Veränderungen der konventionellen Röntgentechnik weit überlegen. Durch seine beachtliche Bildqualität und seine hohe medizinische Aussagekraft hat dieses Verfahren in den vergangenen 10 Jahren einen unvergleichlichen Siegeszug in der medizinischen Technik erlebt. Inzwischen findet die Digitalisierung aber auch in die klassische Röntgen-</p> |

# Räumliche Fragestellungen in technischen Domänen

Manfred Kopitsch, Matthias Schick, Sabina Schmitzgen

Bei der Entwicklung von Expertensystemanwendungen für Problemstellungen aus dem technischen Bereich müssen häufig räumliche Fragestellungen berücksichtigt werden. Dabei werden besondere Anforderungen an die Repräsentations- und Verarbeitungsmechanismen gestellt, die in den Charakteristika dieser technischen Domänen begründet liegen. Neuere Arbeiten in diesem Bereich zeigen interessante Ansätze, wie dieses komplexe Feld pragmatisch angegangen werden kann. In diesem Artikel stellen wir nach einer kurzen Einordnung des Forschungsgebietes unterschiedliche Anwendungsgebiete sowie die dort relevanten Fragestellungen vor. Anhand einiger implementierter Systeme wird der derzeitige Entwicklungsstand in diesem Bereich aufgezeigt.

## 1. Einleitung

Bei der Lösung komplexer Probleme durch wissensbasierte Systeme kommt der expliziten Behandlung des räumlichen Wissens eine immer größere Bedeutung zu. Die Beschäftigung mit räumlichen Fragestellungen in der KI ist bisher überwiegend durch kognitive Ansätze geprägt, doch auch bei der Entwicklung technischer Expertensysteme ist die Integration räumlicher Beschreibungen von großer Bedeutung. Dieser Artikel stellt Lösungsansätze aus konkreten Anwendungen vor und zeigt an ihnen einen Ausschnitt aus dem aktuellen Forschungsstand auf dem Gebiet.

Ausgehend von unterschiedlichen Anwendungsgebieten mit einer Vielfalt räumlicher Fragestellungen kommen wir zu einer Einordnung der Problemlösung in die Bereiche Synthese und Analyse. Anschließend beschreiben wir die Charakteristika bei der Synthese bzw. Analyse räumlicher Szenen ebenso wie Lösungsansätze zur adäquaten Repräsentation und Verarbeitung. Aus den verschiedenen Anwendungsgebieten stellen wir exemplarisch einige Systementwicklungen vor und schließen mit einer kurzen Zusammenfassung.

## 2. Anwendungsgebiete

Technische Anwendungen, welche die Bearbeitung räumlicher Zusammenhänge erfordern, sind zumeist in den Gebieten der Konfiguration und Planung sowie der Diagnostik technischer Systeme und bei der Interpretation geometrischer Darstellungen zu finden.

Konfigurieren  
Viele Konfigurationsaufgaben sind nicht durch Erstellen eines Textfiles lösbar, sondern erfordern die Manipulation von Komponenten im Raum. Räumlich, Schichten, Ebenen, Verschneidungen, auch bei der Montage der Einzelteile, entstehen häufig an solchen Stellen, an denen räumliche Vorgänge nicht oder nur teilweise berücksichtigt werden konnten. Dies ist z.B. bei der Anordnung von Komponenten innerhalb großer Schaltschränke der Fall, wenn bei der Konfiguration nur die Anordnungs- und die Funktionalität berücksichtigt wurden.

Bei einigen Aufgabenstellungen ist die Anwendung der Objektorientierung von zentraler Bedeutung für die Generierung der Lösung, beispielsweise bei der Auslegung von Transportsystemen oder der Konfiguration von Passagierkabinen in Flugzeugen.

### Diagnose

Räumliche Information spielt ebenfalls bei der Überwachung und Diagnose technischer Systeme eine Rolle. Die Ursachen von Störungen können nur dann richtig erkannt werden, wenn die räumlichen Zusammenhänge zwischen den funktionalen Wirkungsketten, sondern auch über räumliche Wirkungs-



Manfred Kopitsch studierte Informatik in Hamburg und Edinburgh. Sein Diplom erhielt er 1991 an der Universität Hamburg. Heute arbeitet er im BAPF-Verbandprojekt PROKON mit Schwerpunkt räumliche Aspekte der Konfiguration.



Matthias Schick ist seit 1991 wissenschaftlicher Mitarbeiter der Universität Hamburg. Im Rahmen des BAPF-Verbandprojektes BEHAVIOR beschäftigt er sich mit den räumlichen Aspekten bei der Diagnose technischer Systeme.



Dipl. Wirtschaftsinformatikerin Sabina Schmitzgen ist wissenschaftliche Mitarbeiterin am Fachbereich Wirtschaftswissenschaften der Universität Hamburg. Ihre Forschungsschwerpunkte sind räumliche Fragestellungen bei der Konfiguration technischer Anlagen.

# A Knowledge Exchange Architecture for Collaborative Human-Computer Communication

David P. Hable, Joanne E. Hurd, and George M. Kasper

**Abstract**—For three decades the concept of a tightly coupled human-computer partnership that would think and process information in a manner superior to that of either human or computer alone has been advanced. However, progress toward the development of such systems has been modest. A fundamental factor limiting progress in this area is the inability of humans and computer applications to communicate effectively. The development of effective communication between such complex and diverse entities is necessary for collaborative problem solving. Based on accepted reference models of human and computer communication, a combined human-computer communication architecture is developed for exchanging knowledge. This architecture explicitly recognizes and models the unique and complementary contributions of each entity and provides for the achievement of collaborative problem solving. The functional components of this architecture are presented, and its potential for the development of more effective human-computer interaction is examined.

## 1. INTRODUCTION

WHEN FACED with a problem that cannot adequately be solved alone, one utilizes the aid of colleagues with supplemental expertise. Working together, the colleagues solve the problem more effectively than the individual would have been capable of alone. This holistic effect is the result of both the synergistic exchange of knowledge and the manner in which the problem solvers iteratively exchange, create, critique, and revise ideas [53]. The problem-solving process by which treated partners combine their expertise to better solve problems is referred to as collaboration. Collaboration requires at least two intelligent entities jointly contributing knowledge in a way that moves the whole toward a desired goal state [15].

A similar idea was first presented by Licklider [25] for human-computer (HC) partnerships when he proposed a synergistic relationship that would think and process information in a manner superior to either the human or computer alone. Licklider's partnership is based on the notion of a human interacting with a computer in the same way that one interacts with a treated colleague whose competence supplements one's own. Since Licklider's work, Posner and Easterby [35], Woods [53], [54], Henderson [18], and others have discussed the lack of progress and called for renewed efforts to develop such HC collaborative problem-solving systems. A factor fundamental

to progress in this area is the ability of humans and computer applications to communicate effectively. Communicating on the abilities of the computer, Hanson and Hix [17, p. 6] state, "... the possibilities of this amazing machine are limited not by its power to compute, but rather by its power to communicate with its human users." Inspired by the term partnership, and explicit in Licklider's notion of a HC partnership, is the exchange of knowledge among peers. However, rather than viewing them as problem-solving partners, current discussions and models of HC interaction perceive either the computer or the human as subordinate. This has resulted in the development of HC interaction in terms of systems, with little if any development of the concepts or structures needed to support an unabbreviated or unbridged exchange of knowledge.

The purpose of this paper is to develop a partnership paradigm for exchanging the knowledge needed to support collaborative HC problem solving. Specifically, accepted models of both human and computer communication are integrated to produce an architecture that identifies the essential functions and features for the design and development of protocols to implement effective dialogues between humans and computer applications. The proposed architecture recognizes the complexity and diversity of both the human and the computer, combining the two co-existing levels of knowledge exchange: syntactic and semantic. The utility of the generic architecture resulting from this combination is emphasized, and its application in collaborative systems is discussed.

## II. BACKGROUND

In a survey of the development of human-computer interaction (HCI), Gaines and Shaw [10] conclude that a fundamental purpose of HCI is to promote "symbiotic partnerships" between humans and computers. Further, they underscore that researchers must concentrate on human-computer communication when they state, "We need effective human-computer communication before we can achieve effective amplification of intelligence" [10, p. 23]. In a companion paper [17], they argue that for HCI to be successful:

- "deep theories" are needed to guide and test models of HCI;

<sup>1</sup> Consistent with the International Organization for Standardization, the term architecture and model are used synonymously throughout this paper to denote a high-level structure or framework for building and developing standards for linking teleoperation systems [17, p. 306].

Manuscript received February 13, 1990; revised October 28, 1990. The authors are with the Department of Information Systems and Decision Support Systems, College of Business Administration, Texas Tech University, Lubbock, TX 79409. IEEE Log Number 9042309.

# Object Orientation in Heterogeneous Distributed Computing Systems

John R. Nicol, C. Thomas Wilkes, and Frank A. Mamola  
GTE Laboratories Inc.

Computer usage is spreading into all functions of large organizations, as can be seen in the widespread appearance of powerful PCs for desktop use and powerful workstations for more specialized uses. Because computers are becoming commonplace in routine business functions, future information-processing environments will likely consist of vast networks of heterogeneous, autonomous, and distributed computing resources, including computers (from mainframe to personal), information-intensive applications, and data (files and databases). The telecommunications infrastructure provides a clear example of such an environment.

There is a growing need for technology to flexibly coordinate computing resources to create networkwide integrated distributed systems that address challenging new information processing requirements. In some cases, these systems might combine resources belonging to different individuals or organizations (for example, connections between customer inventory systems and supplier order-entry systems).

The first phase in creating integrated, large-scale, distributed information systems is interoperability. Until recently, most computers were stand-alone systems unable to communicate with other systems. Two or more computing resources are interconnected if they can exchange messages. However, interoperability guarantees only communication, not cooperation.

For interoperability, the client-server model is the most pervasive. This model organizes a distributed system as a number of distributed server processes that offer various services to client processes across the network. Via interprocess communication mechanisms such as remote procedure calls, servers typically provide clients with access to general system-defined services such as file-storage, printing, processing, authentication, and name services, as well as user-defined services.

**A least-common-denominator approach to object orientation is emerging as a key strategy for flexibly coordinating and integrating networked information processing resources.**

June 1993

IEEE MICROSYSTEMS, DISTANCE, & SOFTWARE

57

Pattern Recognition Letters 13 (1992) 523-528  
North-Holland

July 1992

# A modified stereo matching algorithm suitable for implementation on a convolution specialized hardware

E. Stella, A. Distanti, G. Attolico and T. D'Orazio

Istituto Elettrotelegrafico Segnali di Anagnini, C.N.R., Via Ardeatina 666/5, 00136 Roma, Italy

Received 2 July 1991

## Abstract

Stella, E., A. Distanti, G. Attolico and T. D'Orazio. A modified stereo matching algorithm suitable for implementation on a convolution specialized hardware. *Pattern Recognition Letters* 13 (1992) 523-528.

A new algorithm for stereo matching is presented. It is a modified version of the Marr-Poggio-Gotman technique for the stereo correspondence problem. Our algorithm is a coarse-to-fine strategy (like the Critchley technique) whose steps are formulated in terms of several convolution operations (including the matching phase) making the algorithm suitable for implementation on a convolution-oriented board.

**Keywords:** Image processing, stereo matching, 3-D map

## 1. Introduction

The advancements in machine vision technology have been substantial in recent years with the introduction of faster processors and the improvement in sensor technology. A depth map can be obtained with both direct and indirect methods. The first ones recover depth directly from ranging devices. The second ones recover 3-D information by cues extracted from the intensity image of the observed scene. In the latter class fall stereo techniques. The binocular stereo correspondence problem is the problem of matching two images of the

same scene from different viewing positions. Once the position of the same physical surface point is known in both images, it is possible to recover its depth value. A large number of stereo algorithms have been developed in the literature and in particular two different approaches have been identified: region-based and edge-based [2,3,4]. The region-based techniques have the advantage to be more reliable but have a coarse spatial resolution. The edge-based techniques can be more precise but generally produce a sparse depth map. However, both methods when implemented on a conventional hardware have an impracticable computing time for real-time applications. The use of advanced specialized hardware is suitable to achieve real-time response. In this paper we present an implementation of a modified version of the stereo model proposed by D. Marr and T. Poggio. The aim of this algorithm is to reformulate each step of

The work was supported by "Programma Finalizzato Robotica" of the C.N.R. (Italy).

Correspondence to: E. Stella, Istituto Elettrotelegrafico Segnali di Anagnini, C.N.R., Via Ardeatina 666/5, 00136 Roma, Italy.

0165-1685/92/50523-06 © 1992 — Elsevier Science Publishers B.V. All rights reserved

523

## Plan-based integration of natural language and graphics generation

Wolfgang Wahlster, Elisabeth André, Wolfgang Finkler, Hans-Jürgen Profittlich and Thomas Rist

German Research Centre for Artificial Intelligence (DFKI), Saarbrückenstrasse 1, D-66123 Saarbrücken 11, Germany

### Abstract

Wahlster, W. E. André, E. Finkler, H.-J. Profittlich and T. Rist. Plan-based integration of natural language and graphics generation. *Artificial Intelligence* 63 (1993) 387-427. Multimodal interfaces combining natural language and graphics take advantage of both the individual strengths of each communication mode and the fact that several modes can be employed in parallel. The central claim of this paper is that the generation of a multimodal presentation can be planned in a similar way to the generation of a text presentation to achieve a given communication goal. We describe the multimodal presentation system WIP which allows the generation of alternate presentations of the same content taking into account various contextual factors. We discuss how the plan-based approach to presentation design can be exploited to take graphics generation into account in the production of text and vice versa. We show that well-known concepts from the area of natural language processing like speech acts, arguments, and rhetorical relations take on an extended meaning in the context of multimodal communication. Finally, we discuss two detailed examples illustrating and verifying our theoretical claims.

### 1. Introduction

When explaining how to use a technical device, humans will often utilize a combination of language and graphics. It is a rare instruction manual that does not contain illustrations. Multimodal presentation systems combining natural language and graphics take advantage of both the individual strength of each communication mode and the fact that both modes can be employed

Correspondence to: W. Wahlster, German Research Centre for Artificial Intelligence (DFKI) Saarbrückenstrasse 1, D-66123 Saarbrücken 11, Germany. E-mail: wahlster@dfki.uni-sb.de.

0004-3702/93/5.00.00 © 1993 — Elsevier Science Publishers B.V. All rights reserved

# "HYTime"

## Hypermedia/Time-based Document Structuring Language

The computer and telecommunications industries have made enormous progress in communications technology, standardization in recent years. One effect of good communications technology is that people can concentrate on the information being communicated. More and more people are realizing, however, that being able to send and receive files containing information is not enough.

It is desirable that all digital documents explicitly indicate in a standard way what kind of notation is used in them. When an electronic document is created, its author should be able to incorporate active references to other on-line documents ("hyperlinks"), regardless of the heterogeneity of their notations. It should be a standard way to make "informative" interoperable. Such a standard should, among other things:

- provide a standard way to express the fact that any number of data objects in any number of different notations are related in some way or for some reason;
- provide a standard way to express the way in which any number of data objects in any number of different notations are intended to be rendered for human perception in space, in time, or in both space and time.

One method of incorporating information about the information contained in a document is to use "markup" consisting mainly of "start tags" and "end tags" that respectively precede and follow each logical portion of a document. Tags must be specially punctuated so that the markup can be identified as markup when the document is parsed, and it can be processed separately from the data which it surrounds. Obviously, the rules for specifying the punctuation that distinguishes markup from data must be powerful enough to ensure that the markup and the data cannot be confused even when a given docu-

ment contains data consisting of examples of markup (for example, an SGML document about SGML).

The proposed standard "HYTime" Hypermedia/Time-based Document Structuring Language (ISO/IEC Draft International Standard (DIS) 9774), built on the Standard Generalized Markup Language (SGML) (ISO/IEC International Standard (IS) 8879:1986), is designed to make all this possible. By using SGML/HYTime, all kinds of documents can package their information in a way that is understandable or interpretable by any application that has been provided with an appropriate data interpretation facility. The "structure" character of such documents will also make them amenable to nonsequential browsing, querying, access and version control, and maintenance over very long time spans.

### The World of Structured Documents

The majority of the information stored by human civilization is for perception by human beings. Most of that information is fully formatted for immediate and direct perception, and stored on printed pages. Some of it is fully formatted, but still requires the use of some mechanism to render it perceivable. In this category are such items as magnetically recorded plain ASCII text files, PostScript files, digital video and sound recordings, etc. The rest of the information is stored unformatted, but in such a way that it is convenient to format, perform, or otherwise render for direct perception by human beings, databases and structured documents fall into the latter category.

Structured documents are so named because the hierarchical and sequential structure of the various kinds of



Steven R. Newcomb, Neill A. Kipp, and Victoria T. Newcomb

## D Detaillierte Testergebnisse

Tabelle 8-1 zeigt die Ergebnisse des Tests der Klassifikationsverfahren für ausgewählte Layoutmerkmale, aufgeschlüsselt für die einzelnen Dokumente aus Anhang C.

| Testdokument | w∈ WORDS | Einrückungen | Einrückung ∧ Indent(w) | Keine Einrückung ∧ Indent(w) | Bullets | Bullet ∧ Bullets (w) | Kein Bullet ∧ Bullet (w) | Fettdruck | Fettdruck ∧ Bold(w) | Kein Fettdruck ∧ Bold(w) | Kursivdruck | Kursivdruck ∧ Italics(w) | Kein Kursivdruck ∧ Italics(w) | Großschreib. | Großschreib. ∧ AllCapitals(w) | Keine Großschreib. ∧ AllCapitals(w) |
|--------------|----------|--------------|------------------------|------------------------------|---------|----------------------|--------------------------|-----------|---------------------|--------------------------|-------------|--------------------------|-------------------------------|--------------|-------------------------------|-------------------------------------|
| T1)          | 617      | 5            | 5                      | 0                            | 0       | 0                    | 0                        | 8         | 8                   | 0                        | 26          | 26                       | 0                             | 2            | 2                             | 0                                   |
| T2)          | 1071     | 4            | 4                      | 0                            | 0       | 0                    | 0                        | 31        | 30                  | 1                        | 35          | 32                       | 1                             | 0            | 0                             | 0                                   |
| T3)          | 391      | 0            | 0                      | 0                            | 4       | 4                    | 0                        | 2         | 2                   | 0                        | 4           | 3                        | 0                             | 4            | 4                             | 0                                   |
| T4)          | 644      | 2            | 2                      | 0                            | 0       | 0                    | 0                        | 21        | 21                  | 0                        | 2           | 2                        | 0                             | 0            | 0                             | 0                                   |
| T5)          | 717      | 0            | 0                      | 0                            | 0       | 0                    | 0                        | 189       | 188                 | 0                        | 0           | 0                        | 0                             | 2            | 1                             | 1 <sup>*)</sup>                     |
| T6)          | 954      | 6            | 5                      | 0                            | 1       | 1                    | 0                        | 148       | 148                 | 0                        | 1           | 1                        | 0                             | 13           | 12                            | 10 <sup>*)</sup>                    |
| T7)          | 444      | 3            | 3                      | 0                            | 0       | 0                    | 0                        | 43        | 43                  | 0                        | 1           | 1                        | 0                             | 2            | 2                             | 7 <sup>*)</sup>                     |
| T8)          | 558      | 1            | 1                      | 0                            | 0       | 0                    | 0                        | 2         | 2                   | 0                        | 18          | 18                       | 1                             | 0            | 0                             | 1 <sup>*)</sup>                     |
| T9)          | 810      | 4            | 4                      | 0                            | 2       | 2                    | 1                        | 24        | 23                  | 3                        | 2           | 2                        | 2                             | 4            | 3                             | 2                                   |
| T10)         | 379      | 2            | 2                      | 0                            | 0       | 0                    | 0                        | 2         | 2                   | 0                        | 19          | 18                       | 0                             | 1            | 1                             | 4 <sup>*)</sup>                     |

Tabelle 8-1: Testergebnisse bezüglich Einrückungen, Bullets, Fettdruck, Kursivdruck und Großschreibung auf den Testdokumenten T1) bis T10) aus Anhang C

(<sup>\*)</sup> Zahlen wurden als großgeschriebener Text klassifiziert)

Tabelle 8-2 zeigt die Ergebnisse des Tests der Fontstilklassifikationsverfahren, aufgeschlüsselt für die einzelnen Dokumente S000bin.tif bis S00kbin.tif der UW-III Document Image Database ([UoW96]). Der Test wurde zur Erstellung von [Sch98a] auf Grundlage eines in- zwischen geänderten Segmentierungsverfahrens und mit zum Teil etwas anders realisierten Klassifikationsverfahren durchgeführt. Bei der Auswertung wurden Graphiken und Formeln auf den Dokumenten nicht berücksichtigt.

| Testdokument | w∈WORDS | Fettdruck | Fettdruck ∧ Bold(w)  | Kein Fettdruck ∧ Bold(w) | Kursivdruck | Kursivdruck ∧ Italics(w) | Kein Kursivdruck ∧ Italics(w) | Großschreib. | Großschreib. ∧ AllCapitals(w) | Keine Großschreib. ∧ AllCapitals(w) |
|--------------|---------|-----------|----------------------|--------------------------|-------------|--------------------------|-------------------------------|--------------|-------------------------------|-------------------------------------|
| s000bin.tif  | 1123    | 3         | 1                    | 0                        | 130         | 130                      | 4                             | 1            | 1                             | 0                                   |
| s001bin.tif  | 495     | 149       | 149                  | 0                        | 5           | 5                        | 2                             | 0            | 0                             | 1 <sup>*)</sup>                     |
| s002bin.tif  | 1615    | 211       | 211                  | 4                        | 10          | 8                        | 5                             | 0            | 0                             | 0                                   |
| S003bin.tif  | 1377    | 113       | 111                  | 0                        | 2           | 2                        | 2                             | 0            | 0                             | 0                                   |
| s004bin.tif  | 615     | 14        | 14                   | 0                        | 18          | 13                       | 2                             | 0            | 0                             | 0                                   |
| s005bin.tif  | 509     | 0         | 0                    | 0                        | 65          | 62                       | 7                             | 0            | 0                             | 13 <sup>*)</sup>                    |
| s007bin.tif  | 794     | 10        | 10                   | 0                        | 2           | 2                        | 1                             | 0            | 0                             | 1                                   |
| s008bin.tif  | 836     | 0         | 0                    | 0                        | 17          | 16                       | 0                             | 2            | 0                             | 0                                   |
| S009bin.tif  | 526     | 0         | 0                    | 0                        | 12          | 9                        | 1                             | 5            | 5                             | 10                                  |
| s00abin.tif  | 513     | 0         | 0                    | 0                        | 2           | 2                        | 1                             | 6            | 6                             | 1 <sup>*)</sup>                     |
| s00bbin.tif  | 472     | 0         | 0                    | 1                        | 14          | 14                       | 0                             | 6            | 6                             | 1 <sup>*)</sup>                     |
| s00cbin.tif  | 750     | 0         | 0                    | 0                        | 19          | 19                       | 1                             | 27           | 25                            | 9 <sup>*)</sup>                     |
| s00dbin.tif  | 629     | 0         | 0                    | 0                        | 11          | 9                        | 2                             | 6            | 6                             | 56 <sup>*)</sup>                    |
| s00ebin.tif  | 783     | 0         | 0                    | 1                        | 3           | 3                        | 0                             | 14           | 14                            | 1 <sup>*)</sup>                     |
| s00fbin.tif  | 938     | 0         | 0                    | 0                        | 0           | 0                        | 0                             | 6            | 6                             | 1 <sup>*)</sup>                     |
| s00gbin.tif  | 820     | 0         | 0                    | 0                        | 22          | 10                       | 0                             | 7            | 7                             | 7 <sup>*)</sup>                     |
| s00hbin.tif  | 581     | 36        | 24 <sup>**) 1x</sup> | 0                        | 10          | 10                       | 1                             | 5            | 4                             | 1 <sup>*)</sup>                     |
| s00ibin.tif  | 1313    | 22        | 17 <sup>**) 5x</sup> | 0                        | 16          | 16                       | 2                             | 6            | 6                             | 6 <sup>*) 4x</sup>                  |
| s00jbin.tif  | 884     | 8         | 5 <sup>**) 3x</sup>  | 0                        | 11          | 11                       | 5                             | 0            | 0                             | 19 <sup>*) 8x</sup>                 |
| s00kbin.tif  | 1143    | 2         | 0 <sup>**) 2x</sup>  | 0                        | 9           | 9                        | 0                             | 3            | 3                             | 0                                   |

Tabelle 8-2: Testergebnisse bezüglich Fettdruck, Kursivdruck und Großschreibung auf den Dokumenten S000bin.tif bis S00kbin.tif ([UoW96]).

(\*) Zahlen wurden als großgeschriebener Text klassifiziert)

(\*\*) Fehler wegen Großschreibung, wobei jeweils 1x = 1 Fehler, 2x = 2 Fehler, usw.)



## E Implementierung

Die Entwicklung der Verfahren wurde in der *Programmiersprache C* zunächst auf einer Workstation (Sparc Station IPC) mit dem *Betriebssystem UNIX* begonnen, dann auf einem PC (Siemens Nixdorf PCD-5H) mit dem Betriebssystem Windows NT weitergeführt.

Die für die Programmergänzung zur Extraktion notwendige *OCR* war die Software TEXT-BRIDGE, OCR for Windows, Version 3.0 von Xerox.

Die Testdokumente wurden mit einem *Scanner Avision AV 800* und einer *Auflösung* von 300 dpi erfaßt. Die dadurch entstandenen Bilddaten wurden in *Tiff-Format*<sup>7</sup> (unkomprimiert oder komprimiert mit CCITT Gruppe 4) abgespeichert.

Da zunächst die Machbarkeit des neuen Ansatzes demonstriert werden sollte, wurde keine zeitliche Optimierung der Programme vorgenommen. Der derzeitige *Bedarf an CPU-Zeit* (Pentium-Prozessor, 133 Mhz) für die Attraktivitätsbewertung eines Dokuments ist in Tabelle 8-3 für die Testdokumente T1) bis T10) angegeben. Offensichtlich ist das vorgeschlagene Verfahren tatsächlich in der Lage, relevante Information relativ schnell zu lokalisieren, denn die Klassifikation von Layoutmerkmalen und die Attraktivitätsbewertung benötigen nur etwa 20% der insgesamt verbrauchten Rechenzeit, die restliche Zeit wird bereits von Vorverarbeitung und Segmentierung in Anspruch genommen.

| Testdokument | Dokumentgröße:<br>Breite x Höhe<br>[pixel] | $t_{vs}$ [s] | $t_a$ [s] | $t_{ges}$ [s] |
|--------------|--|--------------|-----------|---------------|
| T1)          | 2144 x 2908                                | 15.52        | 2.35      | 17.87         |
| T2)          | 2368 x 3163                                | 19.87        | 2.05      | 21.92         |
| T3)          | 1968 x 2904                                | 3.20         | 0.84      | 4.04          |
| T4)          | 2448 x 3384                                | 6.98         | 1.57      | 8.55          |
| T5)          | 2432 x 3456                                | 8.40         | 1.94      | 10.34         |
| T6)          | 2464 x 3488                                | 18.20        | 2.02      | 20.22         |
| T7)          | 2432 x 3456                                | 4.71         | 1.04      | 5.75          |
| T8)          | 2432 x 3456                                | 4.65         | 1.07      | 5.72          |
| T9)          | 2432 x 3456                                | 8.18         | 1.83      | 10.01         |
| T10)         | 2432 x 3456                                | 3.83         | 0.84      | 4.67          |

Tabelle 8-3: Rechenzeit der texturbasierten Attraktivitätsanalyse für die Testdokumenten T1) bis T10), inklusive Einlesen der Tiff-Datei und Anzeige des Ergebnisses ( $t_{vs}$ : CPU-Zeit für Vorverarbeitung und Segmentierung,  $t_a$ : CPU-Zeit für Klassifikation bezüglich der Layoutmerkmale und für Attraktivitätsbewertung,  $t_{ges} = t_{vs} + t_a$ )

<sup>7</sup> Tiff-Format: Das „tagged image file format“ wurde 1986 von einem Firmenkonsortium entwickelt. Es dient zum Austausch von Bitmaps zwischen Anwendungen, aber auch zum Speichern optisch gescannter Daten. Das Format unterstützt verschiedene Kompressionsverfahren, darunter die hier z.T. benutzte CCITT Gruppe 4 Lauf-längen-Codierung.