

## **A.1 Zwischen den Zeilen – Ein innovatives Interfacekonzept für selbst organisierende, virtuelle Gemeinschaften**

*Ingmar S. Franke<sup>1</sup>, Severin Taranko<sup>2</sup>, Christin Henzen<sup>1</sup>*

<sup>1</sup> *Technische Universität Dresden, Institut für Software- und Multimediatechnik*

<sup>2</sup> *queo GmbH, Dresden*

### **1 Einleitung**

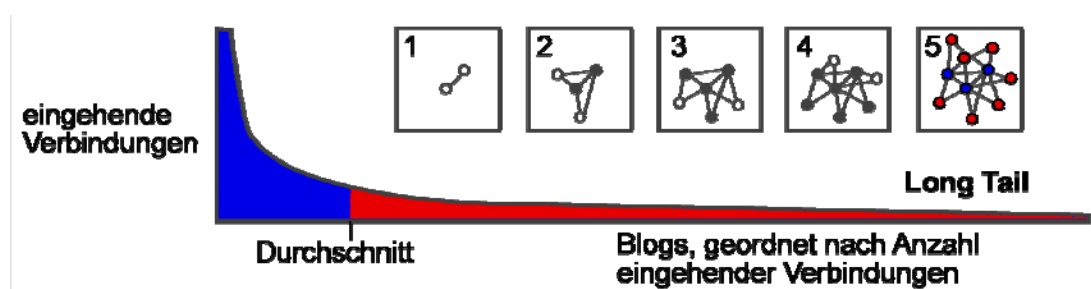
Die Blogosphäre, ein journalistisches Mikronetzwerk, wuchs in den letzten Jahren rasant an. Das Netz aus selbst verwalteten Webseiten umfasste Anfang April 2007 etwa 70 Millionen Blogs. Laut David Sifrey, Geschäftsführer des Weblogsuchmaschinenanbieters Technorati © Inc., [Sif07] betrug die Wachstumsrate von Blogs 3,6 Millionen pro Monat. Die rasch ansteigende Entwicklung kann nicht nur durch neue Technologiestrategien begründet werden, sondern auch durch eine Vielzahl von wirtschaftlichen und menschlichen Einflüssen. In seinem Buch *The Long Tail* [And06] zeigt Chris Anderson drei Einflussfaktoren auf, die den Paradigmenwechsel vorantrieben. Neben der Ungebundenheit der Programme zur Bloggerstellung und -verbreitung spielt das Zusammenwirken von Angebot und Nachfrage eine große Rolle. Der allgegenwärtige Zugang zu schnellen Internetverbindungen bietet vielen Menschen zum Beispiel die Möglichkeit miteinander zu interagieren beziehungsweise zu kommunizieren. Suchmaschinen und soziale Netzwerke fördern die Verbindung zwischen Verfasser und Leser. Durch diese werden Informationen innerhalb des Netzwerks allorts erreichbar. Blogs werden anhand endogener Metadaten oder Links im Web schnell gefunden, deren Inhalte hingegen müssen sukzessiv vom Blogleser selbst durchsucht und analysiert werden.

Heute besteht die Blogosphäre aus Millionen kleiner Leserschaften mit spezifischen Interessensfeldern, die zu einer Basis von qualitativ hochwertigen Informationen zusammengefasst werden können. Die Qualität der Informationen profitiert von Benutzerinteraktionen, wie dem Zitieren, Kommentieren oder Korrigieren von Artikeln. Die Hauptanforderung besteht hierbei in der Entwicklung effektiver Mechanismen für das schnelle Auffinden relevanter Informationen innerhalb umfangreicher Blogs. Derzeit ist die Mehrzahl der Informationen kontextlos beziehungsweise ungeordnet abgelegt und genügt dadurch nicht den Qualitätserwartungen des Lesers. Es bedarf innovativer Benutzerschnittstellen, wozu die vorliegende Entwicklung einen Beitrag leisten soll.

Das nachfolgende Konzept stellt ein blog-spezifisches Visualisierungssystem, das in Abhängigkeit von Benutzerbewertungen aufbereitet wird, vor.

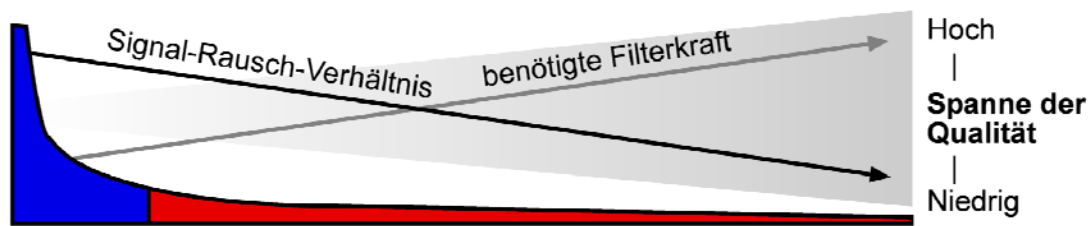
## 2 "Long Tail" der Blogosphäre

Um die Informationen der Blogosphäre analysieren zu können, muss zunächst deren Struktur betrachtet werden. Obgleich das Netzwerk selbst organisiert ist, ist es nicht unstrukturiert. Die Anordnung des Netzes ist durch die Gesetze der Netzwerktheorie geformt. Barabási beschreibt die Blogosphäre als Repräsentation eines skalenfreien Netzwerkes [Bar03]. Zwei Aspekte unterscheiden diese Netze von anderen Modellen – das Wachstum und die bevorzugten Verbindungen. Das Wachstum des Netzes entsteht durch das Hinzufügen eines Knoten. Die bevorzugten Verbindungen sind Verknüpfungen zu neuen Knoten, die nicht zufällig an Bestehende angefügt werden, sondern in Abhängigkeit eines „Popularitätsfaktors“. Dieser wird durch die Anzahl der Verbindungen zu weiteren Knoten beschrieben. In skalenfreien Netzwerken stellt die Relation zwischen Knoten und deren korrespondierenden Verbindungen eine charakteristische Potenzverteilung, wie dies in der folgenden Abbildung gezeigt wird, dar (siehe Abbildung 1).



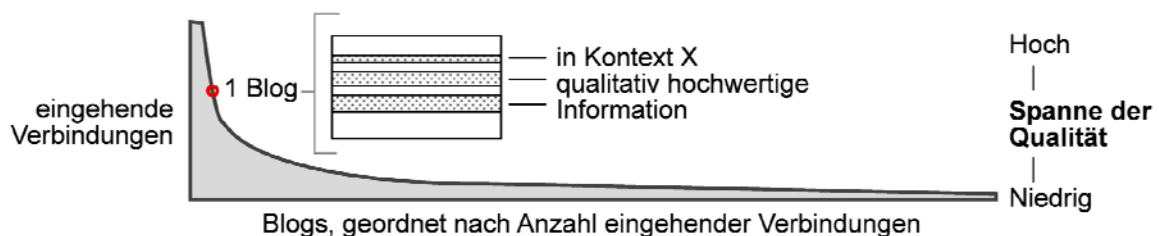
**Abbildung 1: Entwicklung eines skalenfreien Netzwerks und resultierende Potenzverteilung (vgl. [Sif07])**

Anderson untersucht die Auswirkungen solcher Potenzverteilungen für die internetbasierte Wirtschaft [And06]. Er erläutert drei Bedingungen: Vielseitigkeit, Ungleichheit und Netzwerkeinflüsse. Weitere Forschungsergebnisse ([Shi03], [Wat03], [Hub01]) deuten an, dass die Potenzverteilung ubiquitär sei und viele Aspekte auf skalenfreie Netze zutreffen, wie zum Beispiel Protokoll Daten, Benutzerstatistiken sowie Daten über eingehende Verbindungen. Das Verhältnis des Inhalts zur Qualität der Daten wird in der nächsten Abbildung dargestellt (siehe Abbildung 2).



**Abbildung 2: Qualität und Signal-Rausch-Verhältnis der "Long Tail" - Verteilung**

Die qualitativ hochwertige Informationseinheit kann entlang der gesamten Kurve gefunden werden. Diese wird jedoch von einer großen Menge irrelevanter Daten überdeckt, die den durchschnittlichen Qualitätslevel senken. Das Signal-Rausch-Verhältnis beschreibt die Beziehung zwischen Informationen, die den Anforderungen des Lesers genügen (Signal) und Informationen, welche die Erwartungen des Lesers in einem bestimmten Kontext nicht erfüllen (Rauschen). Diese Unterteilung ist sinnvoll, da hochqualitative Informationen in verschiedenen Kontexten, sowohl als Rauschen als auch als bedeutungsvolle Information, interpretiert werden können. Die stark differenzierte Informationsqualität ist nicht nur entlang der "Long Tail" – Verteilung vorzufinden. Diese kann ebenfalls, wie in der folgenden Abbildung veranschaulicht, in einem einzelnen Blog wiedergespiegelt werden.



**Abbildung 3: Qualitätsunterschiede der Informationen eines Blogs**

### 3 Informationsvisualisierung von Webloginhalten

In verschiedenen wissenschaftlichen Vorarbeiten zur Thematik „Weblogs“ wurden semantische Strukturen analysiert (z. B. [Mar04], [ALA04], [AA05]), aber nur wenige setzen sich mit innovativen endogenen Visualisierungskonzepten für Blogs auseinander. In der Datenauswertung von Blogbeiträgen gibt es laut Spoerre drei verschiedene Ansätze [Spo07a]. Die Repräsentation durch sogenannte „Tag Clouds“ zeigt die häufigsten Wörter oder Themen in einer Wolkenform. Durch das Verfahren der Gruppierung (Clustering) werden die Resultate in hierarchische Strukturen beziehungsweise nach Themen geordnet. Der dritte Ansatz befasst sich mit zwei beziehungsweise dreidimensionalen Benutzeroberflächen zur Datendarstellung. Des Weiteren schlägt Spoerre eine Unterteilung basierend auf der visuellen Integration von

verschiedenen Datenquellen vor [Spo07b]. Ein Gitternetzaufbau (grid layout) stellt im Gegensatz zum abbildungs-basierten Entwurf (map-based) jeden Datentyp in einem separaten Bereich dar.

In dem Konzept des Treemap-Ansatzes analysiert ein Visualisierungsalgorithmus zunächst die Struktur einer Hierarchie und teilt die Darstellungsgrundfläche der Informationen entsprechend horizontal und vertikal auf. Die Größe der Flächen resultiert aus der Gewichtung der dargestellten Elemente (vgl. [BDL05]). Ein stapelbasiertes Gestaltungsprinzip für die Veranschaulichung von Bloginhalten verwendet der Social-Bookmark-Anbieter Digg©. Die Blogbeiträge werden als Balken in einem Diagramm symbolisiert. In Echtzeit mit dem Erscheinen im Web, fallen deren Bewertungen als Streifen auf den Stapel des Beitrages. Zusätzlich werden die Überschriften der bewerteten Blogs in Abhängigkeit der Aktualität verschieden koloriert und unter dem Diagramm angezeigt. Die Visualisierung von Webloginhalten im Blog selbst bedarf besonderer Berücksichtigungen. Um die Darstellung in einen bestehenden Blog einzubetten, sollte diese klein sein. Sie kann nur an den Rändern des Blogs platziert werden, um Beiträge nicht zu überdecken. Weiterhin muss eine Vielzahl von Informationen auf oben beschriebenen geringen Raum leicht, schnell und verständlich veranschaulicht werden. Dies indiziert Dekomposition, Abstraktion und Klassifikation des Bloginhalts [Sii03]. Die Daten könnten in verschiedenen Sichten dargestellt werden, um eine kognitive Überlastung zu vermeiden [KBW00]. Das Interface muss für die gesamte, stark differenzierte Benutzergruppe der Blogs ansprechend gestaltet sein. Zudem sollte in den Anforderungen an die Systemperformance das asynchrone Erstellen der Blogbeiträge berücksichtigt werden.

#### **4 Evaluierung der Nische – ein implizites Bewertungssystem für Blogs**

Die Evaluierung von Informationsqualität ist ein komplexer Prozess, der weithin menschliche Fähigkeiten benötigt. Im Gegensatz zu anderen Einheiten hat die Information nur wenige quantifizierbare Attribute, die für computergestützte Evaluationen genutzt werden können. Ansätze mit Hilfe menschlichen Einwirkens, wie zum Beispiel das kollaborationsbasierte Empfehlungssystem, nutzen die soziale Gruppierung (Clustering) und Benutzerprofile zur Bewertung [OGN92]. Bereits bestehende Empfehlungssysteme wurden vorwiegend für elektronische Wirtschaftssysteme entwickelt. Diese sind somit nicht für die Bloginhaltsbewertung geeignet, da diese auf rechnerintensiven Analysen und den Bereitstellungen von Benutzerprofilen basieren. Ein verwendbarer Ansatz sollte folglich jeden Besuch beziehungsweise Besucher des Blogs, unabhängig vom Wissen des Nutzers selbst, ausschöpfen. Es ist nötig, die menschliche Interaktion zu beobachten und zu

interpretieren, um diese für die Evaluierung weiter verarbeiten zu können. Diese Methode wird impliziter Ansatz genannt, da sie keinerlei Auswirkung auf das Verhalten des Benutzers hat. Der Benutzer muss keine expliziten Daten manuell eingeben, denn dies wurde in einer Studie über die Erwartungshaltung gegenüber Blogs von den Nutzern als inakzeptabel zurückgewiesen [AZ97]. Ein weiterer Vorteil der impliziten Methode gegenüber der Expliziten ist, dass diese schwieriger zu manipulieren ist. Ein Ansatz sollte gegenüber dieser Art von Beeinflussung stets resistent sein. Die aktuellsten Forschungen zeigen außerdem, dass implizite Beurteilungen ebenso akkurat wie explizite sein können ([OM96], [MHG97], [OK98]).

Das folgende Konzept stellt einen Bewertungsalgorithmus vor, der zwischen zwei Ebenen der Informationsqualität unterscheidet. Die intrinsischen Qualitätsattribute, als erste Dimension, können direkt aus dem Blogbeitrag extrahiert werden, zum Beispiel die Textlänge. Die interaktiven Attribute können aus dem Verhalten der Benutzer gewonnen werden. Deren Auswahl ist einem Schema für beobachtbares Verhalten zur Modellierung des Informationsgehalts angepasst ([Nic98], [KO01]). Die Abbildung 4 zeigt die qualitätsbezogenen Attribute, die von dem Algorithmus benutzt werden (siehe Abbildung 4).

Qualitätsdimension	Qualitätsmerkmal	Einheit
Intrinsisch	Textlänge	Wörter
	Vorhandensein von visuellen Medien (Bilder/Videos)	Medien
	Weiterführende Links	URLs
	Quellenangabe	Ja/Nein
	Anzahl der Schlagwörter	Wörter
Interaktiv	Anzahl der Klicks	Klicks
	Durchschnittliche Lesezeit	Sekunden
	Drucken des Artikel	Klicks
	Anzahl der Social Bookmarks (Del.icio.us, Mister Wong)	Klicks
	Weiterleiten des Artikels	Klicks
	Anzahl der Kommentare	Kommentare
	Wortanzahl der Kommentare	Wörter
	Anzahl der Trackbacks	Trackbacks
	explizite Bewertungen (Digg, Technorati, eigene Techniken)	Bewertungseinheit

**Abbildung 4: qualitätsbezogene quantifizierbare Attribute eines Blogs**

Der folgende Algorithmus aggregiert die zwei Dimensionen separat, um diese in der nächsten Stufe in einem zweidimensionalen Farbraum darzustellen. Die Dimensionierung fördert zudem die Flexibilität der Visualisierung. Die intrinsischen Attribute können für eine inhaltsunabhängige Qualitätsevaluierung genutzt werden und sind ein erster (aber „schwacher“) Indikator für die Informationsqualität. Daher wird die Bewertung durch die Interaktionscharakteristiken weiter verfeinert. Der Algorithmus interpretiert die Interaktion der Benutzer. Die Mensch-Maschine-Interaktion als komplexer Verhaltensprozess bezieht das Lesen und das Diskutieren von Artikeln ein, um die inhaltsabhängige Qualität der Informationen zu ermitteln. Die Aufmerksamkeit, die ein Artikel erhält (messbar durch Lesezeit), hat eine starke Relation zu expliziten, positiven Bewertungen ([MS94], [MHG97]) innerhalb einer bestimmten Zielgruppe. Es gibt eine Korrelation zwischen verständlicher Repräsentation der Information und einer Mindestanzahl von Wörtern ([Gro02], [NM97]). Bilder und Videos können im Gegensatz zu Text ohne große kognitive Anstrengungen verstanden werden. Dies spielt besonders bei neuen Bloglesern, welche zum ersten Mal die Qualität evaluieren [DW00], eine gewichtige Rolle. Um die Aktivitäten der Leserschaft zu analysieren, kann die Anzahl der Kommentare genutzt werden. Die Anzahl der Worte pro Kommentar vermeidet es, Blogs und deren Inhalte überzubewerten, die eine große Anzahl kurzer Kommentare besitzen ([KT05], [ME05]). Ein weiterer starker Indikator für Qualität ist die persönliche Speicherung, wie das Hinzufügen von Lesezeichen (Bookmarks) oder das Drucken von Artikeln ([MHG97], [ROK00], [Ste93]). Die Speicherung zeigt, dass der Inhalt wertvoll genug ist, um diesen für das weitere Lesen aufzubewahren.

Das Zitieren eines Artikels in einem anderen Blog ist von relevanter Bedeutsamkeit. Es kann angenommen werden, dass eine Information einflussreich war, da diese einen anderen Blogbesitzer dazu veranlasst sie zu zeichnen und weiter zu kommunizieren. Diese Art Inhalte gekreuzt zu verlinken, hat eine große Bedeutung, um dem Leser neue Wissensquellen zu offerieren ([WCB99], [Mar97]). In diesem Zusammenhang werden die netzwerkrelevanten Attribute in der Gesamtgewichtung recht hoch bewertet. Zum Beispiel ist das Weiterleiten eines Artikels ein Anzeichen für den Grad der Neuheit oder die Bedeutung der Informationen für einen Leser [OGN92].

Die Qualität der Informationen eines Blogbeitrages muss im Zusammenhang mit dem Blog gewichtet sein, in dem diese präsentiert werden. Jeder Blog sollte einzeln anhand seiner Struktur beobachtet und evaluiert werden. Die Skala (min bis max) jeder spezifischen Charakteristik wird dynamisch für den jeweiligen Blog berechnet. Um die berechneten Ergebnisse entsprechend ihres Einflusses auf die Qualität zu evaluieren, wird jeder Summand mit einem Gewichtungsfaktor multipliziert. Die Werte der

Gewichtungsfaktoren werden gemäß obigen Erläuterungen gesetzt. Die initiale Konfiguration repräsentiert empirische Studien der Autoren. Die gewählten Werte sind notwendige Voraussetzung für die Implementierung des Algorithmus und bilden die Grundlage für weitere Forschungen und Verfeinerungen. Die initialen Werte der Gewichtungsfaktoren werden in einer folgenden Studie evaluiert, um den konzeptionellen Ansatz zu prüfen. Sollte der Maximalwert eines Attributs „Null“ werden, wird das Produkt aus Attribut und Gewichtungsfaktor ebenfalls „Null“ gesetzt. Der Farbwert der korrespondierenden Dimensionsqualität wird adaptiert. Die resultierenden Gleichungen werden in den Abbildungen 5 und 6 gezeigt.

Qualitätsattribute	Variable	Gewichtungsfaktor (GF)
Textlänge	TL	10%
Anzahl der Medien	AM	20%
Anzahl der Links	AL	32,5%
Anzahl der Schlagwörter	AS	5%
Quellenangabe	QA	32,5%

Wertebereich des intrinsischen Qualitätsfaktors  $INQ_A$ :  $0 \leq INQ \leq 1$

$$INQ_A = \left( \frac{TL_A}{TL_{\max}} \times GF_{TL} \right) + \left( \frac{AM_A}{AM_{\max}} \times GF_{AM} \right) + \left( \frac{AL_A}{AL_{\max}} \times GF_{AL} \right) + \left( \frac{AS_A}{AS_{\max}} \times GF_{AS} \right) + (QA \times WE_{QA})$$

**Abbildung 5: intrinsische Qualitätsfaktoren und daraus resultierende Gleichung**

Qualitätsattribute	Variable	Gewichtungsfaktor (GF)
Anzahl der Klicks	AC	5%
durchschnittliche Lesezeit	DL	10%
Drucken des Artikel	DA	17,5%
Anzahl der Social Bookmarks	ASB	17,5%
Weiterleiten des Artikel	WA	15%
Anzahl der Kommentare	AK	5%
Wortanzahl der Kommentare	WK	15%
Anzahl der Trackbacks	AT	10%
Explizite Bewertung	EB	5%

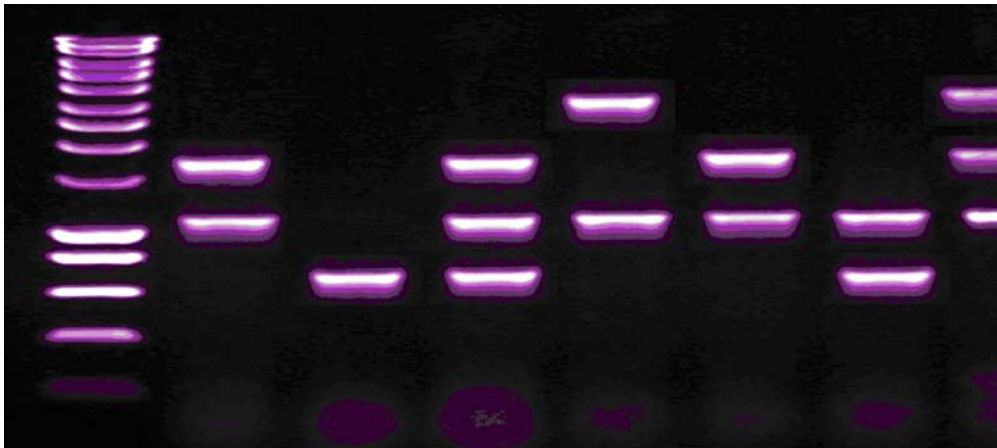
Wertebereich des interaktiven Qualitätsfaktors  $IRQ_A$ :  $0 \leq IRQ_A \leq 1$

$$\begin{aligned}
 IRQ_A = & \left( \frac{AC_A}{AC_{\max}} \times GF_{AC} \right) + \left( \frac{DL_A}{DL_{\max}} \times GF_{DL} \right) + \left( \frac{WK_A}{WK_{\max}} \times GF_{WK} \right) + \left( \frac{WA_A}{WA_{\max}} \times GF_{WA} \right) + \\
 & \left( \frac{AK_A}{AK_{\max}} \times GF_{AK} \right) + \left( \left( \frac{DA_A}{DA_{\max}} + \frac{ASB_A}{ASB_{\max}} \right) \times GF_{DA,ASB} \right) + \left( \frac{AT_A}{AT_{\max}} \times GF_{AT} \right) + \left( \frac{EB_A}{EB_{\max}} \times GF_{EB} \right)
 \end{aligned}$$

**Abbildung 6: interaktive Qualitätsfaktoren und daraus resultierende Gleichung**

## 5 Ein neues Interfacekonzept für Weblogs

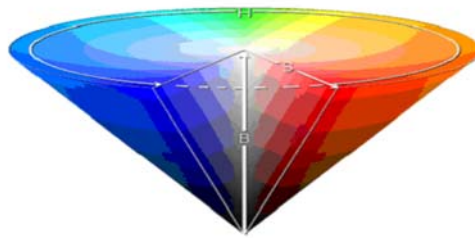
Wie können intrinsische und interaktive Qualitätsfaktoren in einem Interface dargestellt werden? Eine Inspiration für die Gestaltung bietet die naturwissenschaftliche Visualisierung von Genen in der molekularen Biologie. Gene bestehen aus Strängen von Nukleotiden, welche durch eine Methode, Elektrophorese genannt, aufgesplittet werden können. Die Genfragmente werden durch ein negativ geladenes elektrisches Feld geleitet und wandern durch ein zähes Gel. Kleinere Fragmente bewegen sich schneller als große Segmente und gestalten somit ein einzigartiges Diagramm, wie in Abbildung 7 gezeigt wird.



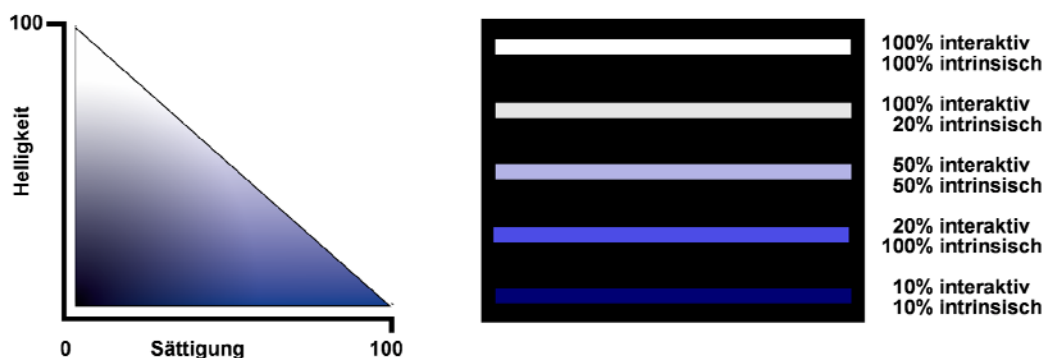
**Abbildung 7: DNA-Fragmente im UV-Licht einer Elektrophorese [NYY05]**

Mit Hilfe dieses Diagramms können komplexe molekulare Strukturen von genetischen Proben leicht erforscht werden. Die visuelle Qualität und das Erkenntnispotential dieser Techniken inspirierten das vorliegende Gestaltungskonzept. Das Interface stellt jede durch den Blogbesitzer angebotene Kategorie in einer Spalte dar. Jede dieser Spalten enthält schmale Streifen - die Artikel des jeweiligen Themas. Der thematische Fokus des Weblogs wird schnell durch eine durchgängige Profillinie sichtbar (siehe Abbildung 8). Die Artikel werden in jeder Kategorie chronologisch von unten nach oben dynamisch geordnet und in Echtzeit dargestellt. Neue Einträge fallen von oben auf den Stapel des jeweiligen Themas.





**Abbildung 10: HSB Farbmodell**

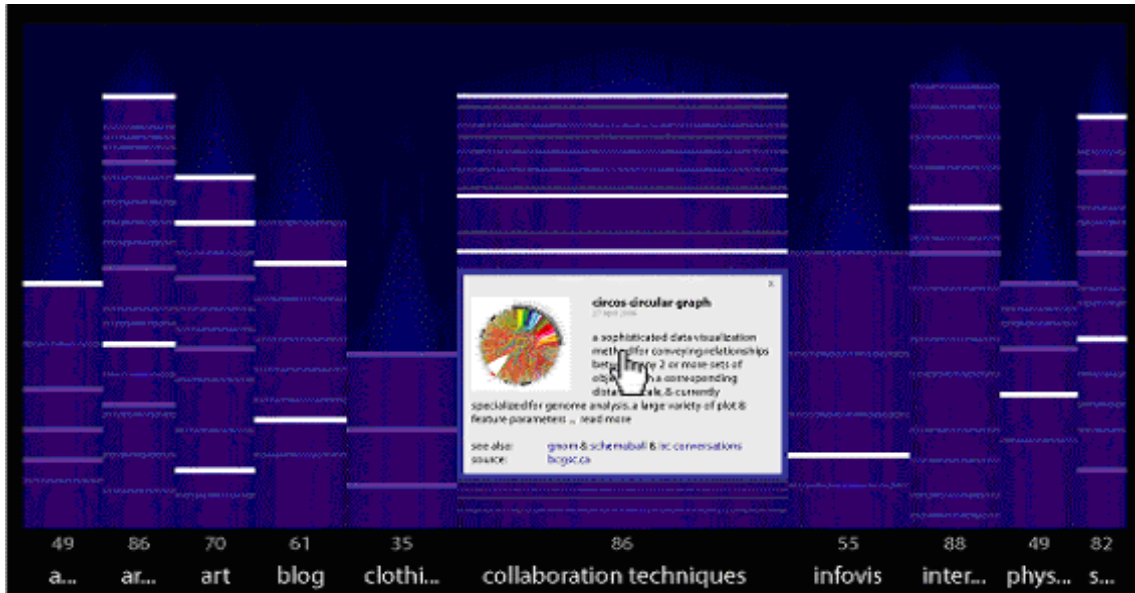


**Abbildung 11: Farbvariation in Bezug auf die Qualität**

Die Dimensionen intrinsische Qualität und Interaktionsqualität, berechnet durch den zu Grunde liegenden Algorithmus (siehe Abbildung 5 und 6), werden auf die Helligkeits- und die Sättigungswerte des Farbmodells abgebildet. Die Farbberechnung verwendet nur einen Subraum des verfügbaren HSB Farbraumes, um Schwarz- und Grauwerte zu vermeiden. Dies optimiert die Vergleichbarkeit und erhöht das ästhetische Empfinden des Interfaces durch den Menschen.

Wegen der größeren Sensitivität der menschlichen Wahrnehmung von Sättigung gegenüber Helligkeit wird die Interaktionsqualität auf die Sättigung abgebildet und die intrinsische Qualität auf die Helligkeit. Die oben gezeigte Abbildung veranschaulicht den aktuellen Farbraum und die resultierenden Farbwerte für die verschiedenen Qualitätssituationen (siehe Abbildung 11). Artikel mit maximaler Qualität werden durch weiße Streifen symbolisiert. Diese erhalten eine höhere Aufmerksamkeit durch den starken Farbkontrast. Ein Blogbeitrag, der weniger intrinsische Qualität besitzt, aber dennoch sehr populär ist und ein Maximum der Interaktionsbewertung erreicht, wird ebenfalls durch einen weißen Streifen abgebildet. Der Ansatz vermeidet falsche Qualitätsschätzungen, da die menschliche Beurteilung als wichtiger eingestuft wird als die Analyse der Struktureigenschaften. Beiträge mit hoher intrinsischer Qualität erlangen eine angemessene Aufmerksamkeit durch die Darstellung mit hohen

Helligkeitswerten. Das Ergebnis des Visualisierungskonzeptes wird in der nächsten Abbildung gezeigt (siehe Abbildung 12).



**Abbildung 12: Visualisierung des Testblogs nach dem Prinzip des neuen Ansatzes**

Als Vorbereitung auf die Gestaltung des Interfaces wurden konzeptuelle Anforderungen auf Kohärenz geprüft. Die folgende Aufzählung gibt einen Überblick über die Anforderungen:

#### Inhalt

- Darstellung einer Vielzahl von verschiedenen Kategorien und Artikeln
- schnelle und offensichtliche Erfassung von thematischen Profilen
- Hervorhebung von Artikeln mit hoher Interaktionsbewertung gegenüber denen mit hoher intrinsischer Qualität

#### Präsentation

- Hervorhebung von hoch bewerteten Beiträgen in jeder Kategorie
- Umsetzung der Hervorhebung von angeklickten Artikeln in Echtzeit
- interaktive Präsentation der Informationen mittels Fischaugenprojektion (in Abhängigkeit der Cursor-Position)
- textuelle Informationen beziehungsweise Bildvorschau im Fokus
- Flexibilität bezüglich der Darstellung neuer Daten, um die Benutzeroberfläche nicht visuell zu überlasten

Die Gestaltung des Interfaces lässt Metadaten, berechnet durch den genannten Algorithmus, in eine leicht verständliche, interaktive Benutzeroberfläche einfließen, um der stark differenzierten und großen Benutzergruppe von Blogs zu genügen. Die Informationen werden somit präsentiert, so dass der Benutzer ein intuitives Verständnis über die thematischen Schwerpunkte des Blogs und die Qualität der Artikel erlangt. Ferner unterstützt das Display die weitere Untersuchung beziehungsweise Erforschung von Daten durch die simple Möglichkeit, Artikel zu durchblättern, um dadurch zu sehen, welche Beiträge andere Benutzer konsumieren.

## **6 Die Implementierung des Interface**

Das Konzept der Benutzeroberfläche wurde als interaktive Visualisierung realisiert. Die Handlungen des Benutzers werden sofort (in Echtzeit) verarbeitet und die entsprechenden Daten durch den Algorithmus angepasst visualisiert (Adaptivität). Das Softwarestrukturmuster Model-View-Controller (MVC) bildet die Grundlage für die Implementierung. Dieses ermöglicht die Trennung des Quellcodes von den Designelementen, wodurch eine Wiederverwendung der Komponenten gewährleistet ist. Das Programm Flash ® von Adobe Systems Inc. wurde für die prototypische Umsetzung gewählt, da es eine weit verbreitete und akzeptierte Lösung für die Darstellung von interaktiven Inhalten in Webbrowsern bietet. Weiterhin stellt es eine Entwicklungsumgebung für die Gestaltung und den Entwurf von Animationen zur Verfügung, und es bietet eine objektorientierte Sprache für die Realisierung von Funktionalitäten.

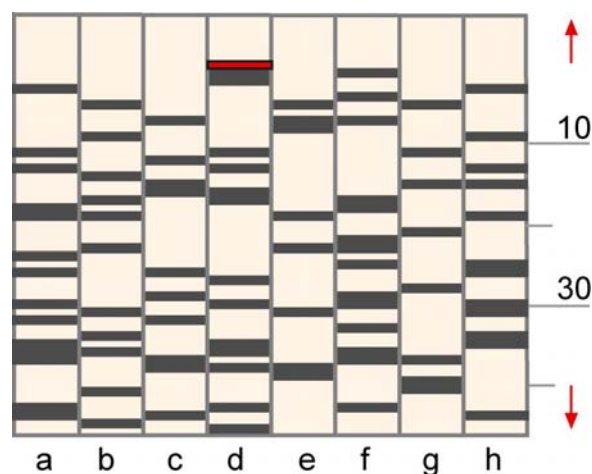
Die Flash-Dateien beinhalten lediglich die Elemente für das Interface (Ansicht/View des MVC-Muster), wie zum Beispiel die Buttons, Graphiken, Schriftzüge und Animationen. Das Content Management System Wordpress ® Inc. wurde für die Implementierung des Backends genutzt, da es eine frei verfügbare Software ist, die unter der General Public License (GPL) lizenziert wurde. Für vielschichtige Tests konnte ein bereits bestehender Blog um das neue Interface erweitert werden. Die Wordpress Plugin-Architektur stellt die Verbindung zwischen der Benutzeroberfläche und dem Wordpress Backend dar. Die Funktionalität des Plugins ist in zwei Abstraktionsebenen implementiert. Der Kern unterstützt hierbei die Funktionalität, die unabhängig von einem spezifischen Blogsystem ist. Die zweite Ebene nutzt die von Wordpress angebotenen Methoden und Erweiterungen, um die Attribute zu berechnen, die der Bewertungsalgorithmus benötigt.

## 7 Einsatz und Weiterentwicklung des Flash-Interface

Das vorgestellte Interfacekonzept kann einen bestehenden Blog mit einer informativen Visualisierung seines Inhalts aufwerten. Zum Zweck der Evaluierung ist das Interface auf die rechte Seite eines bestehenden Blogs platziert. Ein Zufallsgenerator berechnete die fehlenden initialen Werte der Interaktionsqualität. Somit konnte den Probanden ein erstes realistisches visuelles Interface angeboten werden.

Darauf aufbauende Testfälle und die Analyse von Serverlogs zeigten, dass dieses Interface die Benutzer anhält, die empfohlenen Artikel eher zu sichten. Infolge der Empfehlung verzeichneten selbst ältere (aber inhaltlich relevante) Beiträge, die nicht auf der Startseite des Blogs gelistet oder durch Suchresultate gezeigt wurden, eine erhöhte Aufmerksamkeit. Die initiale Konfiguration der Gewichtungsfaktoren wird in einer weiteren Studie zur Überprüfung des konzeptionellen Ansatzes des Algorithmus evaluiert.

Weiterhin ist eine Verbesserung der Implementierung hinsichtlich zusätzlicher Darstellungsmodifikationen, wie einer Ansicht mit einem Zeitschieberegler als Navigationsmittel (siehe Abbildung 13), denkenswert. Artikel werden in dieser Sichtweise chronologisch nach ihrem Veröffentlichungsdatum sortiert und gelistet. Durch Herauf- und Herunter-Scrollen kann ein Zeitschieber zur Navigation genutzt werden. Hierdurch besteht die Möglichkeit, große Archive von Artikeln in dem kleinen Interface darzustellen.



**Abbildung 13: Konzept für eine Ansicht mit Zeitschieberegler**

Eine zusätzliche Erweiterung könnte die Visualisierung des Anklickens der Beiträge durch gesonderte Hervorhebung, wie dem „Highlighting“ sein. Der Besucher des Blogs kann somit leicht feststellen, welche Artikel andere Benutzer aktuell lesen. Diese animierten Veränderungen sind sowohl eine ästhetische Verbesserung hinsichtlich der

menschlichen Wahrnehmung als auch eine kognitive Optimierung, da der Benutzer bewusst visuelle Veränderungen des Interfaces erkennt. Diese Eigenschaften verbessern die Benutzbarkeit der Visualisierungskomponente erheblich, erfordern allerdings auch eine fortgeschrittene technische Lösung. Die Autorengemeinschaft möchte die weitere Entwicklung des Interfaces fördern, indem sie den Quellcode der Anwendung unter der GPL Lizenz veröffentlicht und eine dokumentierte Version als Erweiterung von Wordpress zur Verfügung stellt.

## **8 Zusammenfassung**

Weblogs und ihre gegenseitige Vernetzung repräsentieren ein soziales und skalenfreies Netzwerk mit einer charakteristischen Potenzverteilung: die Blogosphäre. Die Vielfältigkeit der Qualitätsunterschiede in den Blogs erschwert es, Informationen von narrativer Qualität zu finden. Der dargestellte Ansatz verfolgt ein innovatives Interfacekonzept auf Basis eines endogenen Bewertungsalgorithmus. Letzterer teilt die implizit gegebenen Informationen in intrinsische und interaktive Qualität auf. Die Verarbeitung der Metadaten basiert auf drei Hauptansätzen zur Datenvisualisierung - der Technologie, der Biologie und dem Design. Für die Visualisierung von Genen wird in der molekularen Biologie das Verfahren der Elektrophorese verwendet. Dieses Verfahren hat sich bereits in der visuellen Datenauswertung bewährt und kann daher auf die Information der Weblogs adaptiert werden. Weiterhin werden die Eigenschaften des menschlichen Auges berücksichtigt. Im Bereich eines Farbtons des HSB-Farbmodells werden die intrinsische und die interaktive Qualität auf die Helligkeits- und die Sättigungswerte des Modells abgebildet. Eine Kombination des biologischen Ansatzes mit der Technik der Fischaugenprojektion erzielt die interaktive Repräsentation von Informationen mit einem Fokus auf das thematische Profil.

Informationen in einem Weblog sind nicht immer ungenügend, vielmehr sind diese unzuweckmäßig in Form von chronologisch sortierten Listen dargestellt. Der komplexe Prozess der Informationsevaluierung wird nur durch wenige, quantifizierbare Attribute zur rechnergestützten Messung begünstigt. Besonders implizit gegebene Daten, wie Anzahl von Klicks oder Links zu anderen Blogs, müssen analysiert werden, um einen Blogartikel gewichten zu können. Mit dem vorliegenden Ansatz wird gezeigt, dass etablierte Darstellungsprinzipien anderer Naturwissenschaften zur Verfügung stehen und zu einer visuellen Benutzerschnittstelle zur Datenauswertung weiterentwickelt werden können.

*Dem GeNeMe2008-Auditorium kann der vorliegende Beitrag in einer Echtzeit-visualisierung präsentiert und zur Diskussion gestellt werden. Das System ist mittlerweile über dem Status eines Prototyps hinaus einsetzbar.*

## 9 Literatur

- [AA05] L. A. Adamic, E. Adar, 2005, Tracking information epidemics in blogspace. Web Intelligence 2005, S. 207–214, in Proc. of the 2005 IEEE/WIC/ACM International Conference, DOI: 10.1109/WI.2005.151.
- [ALA04] L. A. Adamic, R. M. Lukose, E. Adar, L. Zhang, 2004, Implicit structure and the dynamics of blogspace.
- [And06] C. Anderson, 2006, The Long Tail - How Endless Choice is Creating Unlimited Demand. Random House Uk Ltd, ISBN-10: 1401302378.
- [AZ97] C. Avery, R. Zeckhauser, 1997, Recommender systems for evaluating computer messages, Communications of the ACM, Volume 40, Issue 3, ACM Press, S. 88–89, ISSN: 0001-0782.
- [Bar03] A. L. Barabási, 2003, Linked, Plume Books, ISBN-10: 0452284392.
- [BDL05] M. Balzer, O. Deussen, C. Lewerentz, 2005, Voronoi treemaps for the visualization of software metrics. In SoftVis '05: Proceedings of the ACM symposium on Software visualization, S. 165-172, New York, NY, USA, ACM Press.
- [DW00] D. Diaper, P. Waeland, 2000, World wide web working whilst ignoring graphics; good news for web page designer, Interacting with Computers, S. 163–181.
- [Gro02] N. Groeben, 2002, Leserpsychologie: Textverständnis, Textverständigkeit, Aschendorff Verlag, ISBN-10: 3402042983.
- [HR06] B. Hallenberger, H. Rudolf, 2006, Farben im Webdesign. Online-Dokument, eingesehen am 2007-10-29, verfügbar im Web unter:  
<http://www.metacolor.de/farben/lieblingsfarben.htm>.
- [Hub01] B. A. Huberman, 2001, The Laws of the Web: Patterns in the Ecology of Information. MIT Press, ISBN-10: 0262083035.
- [KBW00] A. Kuchinsky, M. Baldonado, A. Woodruff, 2000, Guidelines for using multiple views in information visualization, S. 110–119, in Proc. of Advanced Visual Interfaces, Palermo, Italy, ACM, ISBN:1-58113-252-2.
- [KO01] J. Kim, D. W. Oard, 2001, Modeling information content using observable behavior, Journal of the Korean Society for Library and Information Science, in Proc. of the 64 Annual Meeting of the American Society for Information Science and Technology, S. 38-45.
- [KT05] A. Keshelashvili, K. D. Trammell, 2005, Examining new influencers: a selfpresentation study of a-list blogs. Journalism and Mass Communication Quarterly, 82:S. 968–982.

- [Mar97] G. Marchionini, 1997, Information seeking in electronic environments. Cambridge University Press, ISBN: 0521586747.
- [Mar04] C. Marlow, 2004, Audience, structure and authority in weblog community.
- [ME05] A. de Moor, L. Efimova, 2005, Beyond personal webpublishing: An exploratory study of conversational, Vol. 38, S. 107a. in Journal: System Science.
- [MHG97] D. Maltz, J.L. Herlocker, L.R. Gordon, J.A. Konstan, B.N. Miller and J. Riedl, 1997, Grouplens: Applying collaborative filtering to usenet news, Vol. 40, Nr. 3, S. 77–87, ACM Press, ISSN: 0001-0782.
- [MS94] M. Morita, Y. Shinoda, 1994, Information filtering based on user behaviour analysis and best match text retrieval, in Proc. SIGIR'94, Dublin, Ireland, S. 272–281. Springer-Verlag.
- [Nic98] D. M. Nichols, 1998, Implicit ratings and filtering, in Proc. of the 5th DELOS Workshop on Filtering and Collaborative Filtering, Budapest, Hungary, ERCIM, S. 31–36.
- [NM97] J. Nielson, J. Morkes, 1997, Concise, scannable, and objective: How to write for the web. Online-Dokument, eingesehen am: 2007-10-29, verfügbar im Web unter: <http://www.useit.com/webwriting/writing.html>.
- [NYY05] T. Nakagawa, A. Yamamoto, S. Yamasaki, Y. Yamamoto et. al., K. Nishi, S. Rand, 2005, Abo blood typing from forensic materials - merits and demerits of detection methods utilized in our laboratories, and biological significance of the antigens. Abil Aggrawal's Internet Journal of Forensic Medicine and Toxicology.
- [OGN92] B. M. Oki, D. Goldberg, D. Nichols and D. Terry, 1992, Using collaborative filtering to weave an information tapestry, in Communications of ACM, Vol. 35, Issue 12, special issue on information filtering, S. 61–70, ISSN: 0001-0782.
- [OK98] D.W. Oard, J. Kim, 1998, Implicit feedback for recommender system, in Proc. of the AAAI Workshop on Recommender Systems, Madison, S. 81–83.
- [OM96] D.W. Oard, G. Marchionini, 1996, A conceptual framework for text filtering, Technical Report CS-TR3643, University of Maryland, College Park.
- [ROK00] K. Romanik, D. W. Oard, J. Kim, 2000, User modeling information filtering based on implicit feedback, Technical Report, University of Maryland, College Park.
- [Shi03] C. Shirky, 2003, Exposure: From Friction to Freedom, Aula, Meeting of Minds, Helsinki, S. 77-81.

- 
- [Sif07] D. Sifry, 2007, The state of the live web. Online-Dokument, eingesehen am: 2007-04-17, verfügbar im Web unter:  
<http://www.sifry.com/alerts/archives/000493.html>.
- [Sii03] H. Siirtola, 2003, Combining parallel coordinates with the reorderable matrix. In Coordinated and Multiple Views in Exploratory Visualization, S. 63–74, INSPEC AN: 7854017.
- [Spo07a] A. Spoerri, 2007, Coordinating linear and 2d displays to support exploratory search, in Coordinated and Multiple Views in Exploratory Visualization, S. 16–26, DOI: 10.1109/CMV.2007.5.
- [Spo07b] A. Spoerri, 2007, Visual mashup of text and media search results. S. 216–221, IEEE Computer Society, DOI: 10.1109/IV.2007.125, 2007.
- [Ste93] F. C. Stevens, 1993, Knowledge-based assistance for accessing large, poorly structured information spaces, University of Colorado, (UMI Order No: GAX93-20482).
- [Wat03] D. J. Watts, 2003, Six Degrees: The Science of a Connected Age. WW Norton & Co Inc., ISBN-10: 0099444968.
- [WCB99] N. S. Wehrle, D. C. Crow, M. D. Byrne, B. E. John, 1999, The tangled Web we wove: taskonomy of WWW use, CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems, S. 544–55,1 ACM, ISBN: 0201485591.