

Diplomarbeit

User Generated Content als Grundlage für Trend-Scouting

4. Mai 2007

Lutz Lange
Matr. Nr. 185614
<l.lange.berlin@gmx.de>

Technische Universität Berlin
Informatik und Gesellschaft
Sekretariat FR 5-10
Franklinstraße 28/29
10587 Berlin, Germany
Prof. Dr.iur Bernd Lutterbeck
Prof. Prof. Dr. Volker Trommsdorff

Eidesstattliche Erklärung

Die selbständige und eigenhändige Anfertigung versichere ich an Eides Statt.

Berlin, den / Unterschrift

Danksagung

An dieser Stelle möchte ich all jenen danken, deren Unterstützung und Rat diese Arbeit möglich gemacht hat. Ich danke Ralf Rattay, der von Beginn an meine Diplomarbeit vor praktischem Hintergrund betreute. Mit seiner Begeisterung für Web2.0 und User Generated Content Technologien hat er mich angesteckt. Seine Hinweise zu den Arbeitsmethoden von Markt-, Trend- und Zukunftsforschern waren richtungsweisend. Ich danke der Firma Trommsdorff & Drüner, deren Hardware die Bewältigung der angefallenen Datenmengen erst möglich machte. Ich danke Frank Pallas, meinem universitären Betreuer. Aus seinen Beratungsgesprächen ging ich immer mit neuer Kraft an die Arbeit, und seine Ratschläge haben mir in Zeiten des Zweifels wieder Mut gemacht. Dank gebührt auch Diana Tettling und Daniela Robel, die diese Arbeit Korrektur lasen. Ich danke meinen Eltern für ihre finanzielle Unterstützung, die mir das Schreiben dieser Arbeit ermöglicht hat.

Vor allem danke ich meiner Freundin, Daniela Robel, für ihre Geduld, Unterstützung und Liebe.

Zusammenfassung

Die mit Web2.0 und User Generated Content Technologien verfügbaren Informationen haben ein enormes Potential, die Methoden von Markt-, Trend- und Zukunftsforschern zu stützen. Zur Verarbeitung und Auswertung dieser Daten sind innovative Ansätze zu leistungsfähigen, computergestützten Systemen notwendig. Aus den Methoden und Arbeitsweisen der Markt-, Trend- und Zukunftsforscher werden Anforderungen an Daten und Systeme zu ihrer automatisierten Unterstützung extrahiert. Durch den Vergleich dieser Anforderungen mit einer Betrachtung verschiedener User Generated Content Arten werden Social Bookmarks und die Anwendung Delicious als Datenlieferant für eine beispielhafte Umsetzung eines solchen Auswertungssystems ausgewählt. Aus den Anforderungen an das Auswertungssystem und den Arbeitsweisen von Markt-, Trend- und Zukunftsforschern wird zusammen mit einer Betrachtung manueller Explorationsmöglichkeiten bei Delicious ein Systemkonzept entwickelt, mit dem eine auf Themen bezogene Auswertung von Social Bookmarks über frei wählbare Zeitintervalle möglich ist. Durch das Bilden von Usergruppen über Kriterien und den Einsatz effektiver Algorithmen zur Aggregation und Analyse der Daten lassen sich thematisch relevante Ergebnisse auch über große Datenmengen performant generieren. Die Leistungsfähigkeit dieses Systems und die Relevanz der Ergebnisse werden anhand von Beispielen demonstriert. Es zeigt sich, dass das Auswertungssystem thematisch relevante Ergebnisse produzieren kann. Gute Kriterien selektieren User, die sich mit einem beabsichtigten Zielthema aktiv beschäftigen. Dominant ist ein Zielthema, wenn sich die selektierten User stark mit diesem Thema befassen. Gute Kriterien und dominante Zielthemen sind im vorgestellten Auswertungssystem Bedingungen, um relevante Resultate zu erzeugen.

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einleitung | 1 |
| 2 | Marketing und Zukunftsforschung | 2 |
| 2.1 | Marketing | 3 |
| 2.2 | Marktforschung | 4 |
| 2.2.1 | Milieus | 4 |
| 2.2.2 | Produktdiffusion | 6 |
| 2.2.3 | Marktforschung Online | 8 |
| 2.3 | Trend- und Zukunftsforschung | 9 |
| 2.3.1 | Wissenschaftlichkeit der Trendforschung | 10 |
| 2.3.2 | Beobachtungsfelder | 12 |
| 2.3.3 | Zukunftsforschung | 13 |
| 2.4 | Informations Arbitrage | 14 |
| 2.5 | Zusammenfassung | 16 |
| 3 | Informationen und Gesellschaft | 18 |
| 3.1 | Wohlfahrt und Wachstum | 18 |
| 3.2 | Informationsökonomie | 20 |
| 3.3 | Dezentralisierung | 25 |
| 3.3.1 | Transaktionskosten | 26 |
| 3.3.2 | Kommunikationskosten | 26 |
| 3.3.3 | Internet End-to-End | 29 |
| 3.4 | Peer Production | 30 |
| 3.5 | User Generated Content | 33 |
| 3.5.1 | Foren | 34 |
| 3.5.2 | Blogs | 35 |
| 3.5.3 | Wikis | 35 |
| 3.5.4 | Social Networking | 36 |
| 3.5.5 | Spezialisierte Content Sharing Systeme | 37 |
| 3.6 | Zusammenfassung | 38 |
| 4 | Delicious und Social Bookmarking | 40 |
| 4.1 | Social Bookmarking | 41 |
| 4.2 | Tagging | 43 |
| 4.3 | Delicious | 47 |
| 4.3.1 | Zur Entstehung | 48 |
| 4.3.2 | Motivation | 48 |
| 4.3.3 | Bookmarks zentral speichern | 49 |
| 4.3.4 | Eigene Bookmarks finden | 50 |
| 4.3.5 | Fremde Bookmarks finden | 51 |
| 4.3.6 | Interessante User entdecken | 53 |
| 4.4 | Zusammenfassung | 54 |

| | | |
|----------|--|-----------|
| 5 | Systembeschreibung | 56 |
| 5.1 | Grundlagen | 57 |
| 5.2 | Umsetzung | 59 |
| 5.2.1 | Daten sammeln | 60 |
| 5.2.2 | Gruppen bilden | 63 |
| 5.2.3 | Analyse | 64 |
| 6 | Ergebnisse | 68 |
| 6.1 | Datensammlung | 68 |
| 6.2 | Usergruppen und Kriterien | 68 |
| 6.3 | Einfache Auswertung | 71 |
| 6.4 | Graphische Auswertung | 83 |
| 6.5 | Bewertung der Ergebnisse | 86 |
| 6.5.1 | Kritische Systembetrachtung | 88 |
| 6.5.2 | Markt-, Trend- und Zukunftsforschung | 89 |
| 6.5.3 | Die Kluft | 91 |
| 6.6 | Zulässigkeit der Datenverwertung | 92 |
| 6.7 | Zusammenfassung der Ergebnisse | 93 |
| 7 | Ausblick | 93 |
| | Literaturverzeichnis | 95 |

... Web 2.0 is fueled by an outpouring of creativity from the people formerly known as consumers. From YouTube auteurs to bloggers to amateur photographers competing with the paparazzi, USERGENERATED CONTENT is revolutionizing the media landscape ...

Time Magazine 17. Dezember 2006 (Howe 2006)

1 Einleitung

Diese Arbeit befasst sich mit der Auswertung von User Generated Content. Ziel dieser Arbeit ist es, ein System zur automatisierten Unterstützung von Methoden der Markt-, Trend- und Zukunftsforschung zu beschreiben und zu testen. Die vorliegende Arbeit gliedert sich in drei Teile. Im ersten Teil der Arbeit werden die Grundlagen und Anforderungen an ein solches System erarbeitet. Der zweite Teil stellt eine ausgewählte User Generated Content Art mit ihren Eigenschaften und Möglichkeiten vor. Im dritten und letzten Teil wird ein Auswertungssystem entworfen, vorgestellt und auf seine Leistungsfähigkeit untersucht.

Der erste Teil der vorliegenden Arbeit gliedert sich in den Abschnitt Marketing und Zukunftsforschung und den Abschnitt Informationen und Gesellschaft. Dort werden notwendige Grundlagen und Anforderungen für das Auswertungssystem ausgearbeitet. Dabei beleuchtet der Abschnitt Marketing und Zukunftsforschung die Methoden und Arbeitsweisen von Markt-, Trend- und Zukunftsforschern. Im Abschnitt Informationen und Gesellschaft werden die Eigenschaften von Informationen und Methoden zur Produktion von Informationsgütern betrachtet. Sie bilden die Grundlage einer Einordnung und Unterscheidung von User Generated Content Systemen.

Im Abschnitt Delicious und Social Bookmarking wird eine User Generated Content Art zur Analyse ausgewählt. Der gewählte User Generated Content wird dann anhand eines konkreten Beispiels genauer betrachtet. Bei diesem Beispiel handelt es sich um die Social Bookmarking Anwendung Delicious.

Im dritten Teil dieser Arbeit wird ein Auswertungssystem zur Unterstützung von Methoden der Markt-, Trend- und Zukunftsforschung beschrieben. Dabei wird im Abschnitt Systembeschreibung ein Konzept zur Umsetzung eines solchen Auswertungssystems entworfen und die Proof-of-Concept Implementierung des Systems vorgestellt. Im anschließenden Abschnitt zu den Ergebnissen wird die Funktionsweise und Leistungsfähigkeit des Auswertungssystems an Beispielen geprüft und vorgestellt.

2 Marketing und Zukunftsforschung

Das Wachstum des Internet scheint einzig durch die Zahl der Menschen in der Welt begrenzt zu sein. Ein Blick in die Tabelle 1 gibt Aufschluss über die Zahlen zu Beginn des Jahres 2007. Inzwischen sind mehr als eine Milliarde Menschen im Netz aktiv, das entspricht mehr als 16 % der Weltbevölkerung. Bemerkenswert ist, dass deutlich mehr Asiaten als Nordamerikaner im Netz online sind. Interessant ist auch, dass diese 389,3 Millionen Asiaten erst 10 % der asiatischen Bevölkerung ausmachen, während mit 232 Millionen schon fast 70 % der Nordamerikaner im Netz sind. Über die Bedeutung dieser Zahlen für das weitere Wachstum und die Zukunft des Internet kann man spekulieren. Hier sollen sie aber das enorme Potential des Netzes für soziale Forschungen belegen. Gerade eine Durchdringung der Bevölkerung von fast 70% ist eine gute Basis für solche Forschungsvorhaben, da fast alle Bevölkerungsschichten vertreten sind.

| World Regions | Population (2007 Est.) | Pop. % World | Internet Usage | % Pop. Penetr. | Usage % of World |
|---------------------------|-------------------------------|---------------------|-----------------------|-----------------------|-------------------------|
| Africa | 933,448,292 | 14.2 % | 32,765,700 | 3.5 % | 3.0 % |
| Asia | 3,712,527,624 | 56.5 % | 389,392,288 | 10.5 % | 35.6 % |
| Europe | 809,624,686 | 12.3 % | 312,722,892 | 38.6 % | 28.6 % |
| Middle East | 193,452,727 | 2.9 % | 19,382,400 | 10.0 % | 1.8 % |
| North America | 334,538,018 | 5.1 % | 232,057,067 | 69.4 % | 21.2 % |
| Latin America / Caribbean | 556,606,627 | 8.5 % | 88,778,986 | 16.0 % | 8.1 % |
| Oceania / Australia | 34,468,443 | 0.5 % | 18,430,359 | 53.5 % | 1.7 % |
| WORLD TOTAL | 6,574,666,417 | 100.0 % | 1,093,529,692 | 16.6 % | 100.0 % |

Tabelle 1: World Internet Usage And Population Statistics Quelle: Internetworldstats (2007)

Beigetragen zum enormen Wachstum der letzten Jahre hat die Technik des Web2.0¹. Auf den ersten Blick ist Web2.0 eine Sammlung von Techniken, die das Erstellen von Web-Anwendungen vereinfachen und beschleunigen. In der Regel werden über solche Anwendungen Daten und Dienstleistungen angeboten, die auch ohne die Techniken des Web2.0 realisierbar wären. Web2.0-Techniken verkürzen nicht nur die Entwicklungszeit, sie machen Anwendungen oft interaktiver und leichter bedienbar.

Mit den Web2.0-Techniken sind eine Reihe neuer Anwendungsarten entstanden. In einem Großteil von ihnen fällt dem Anwender eine aktivere und zentralere Rolle zu als je zuvor. Diese neuen User Generated Content Anwendungen weisen erstaunliche Nutzerzahlen auf. User Generated Content und Web2.0 waren laut Nielsen//Netratings (2006a) das Onlinethema des Jahres 2006.

Generating and sharing content – the user in charge

The three fastest growing online brands in the UK highlight how generating and sharing content is this year's online theme. YouTube, Flickr and MySpace

¹Der Begriff Web2.0 wurde von Tim O'Reilly geprägt (siehe O'Reilly 2005).

all focus on technologies that allow users to define the content that their peers are looking at online. Factor in Photobucket, the photo and video sharing website, and social network Bebo and half of the ten fastest growing brands fall under the “Web 2.0” umbrella.

Nielsen//Netratings (2006a)

In Tabelle 2 finden sich die Top 3 Onlinemarken basierend auf Daten aus Großbritannien von Anfang des Jahres 2006. Die ersten drei Plätze nach Nutzerzahlen entfallen dabei ausschließlich auf Web2.0 User Generated Content Dienste. Die Zuwächse innerhalb von sechs Monaten erreichen fast durchweg über 100%. Diese Zahlen basieren zwar auf Daten für Großbritannien, ein Bericht mit Daten aus den USA (Nielsen//Netratings 2006b) zeichnet allerdings ein sehr ähnliches Bild.

| Rank | Brand | UA % Growth Jan-Jul 2006 | UA (000's) July 2006 | Nature of Buissness |
|------|---------|-----------------------------|-------------------------|---------------------|
| 1 | YouTube | 478% | 3,585 | Video sharing |
| 2 | Flickr | 131% | 1,008 | Photo sharing |
| 3 | MySpace | 98% | 3,502 | Social networking |

Tabelle 2: Top 3 brands in 2006 by UK Unique Audience (UA) growth, Quelle : Nielsen//Netratings (2006a)

Durch die hohen Nutzerzahlen führender Web2.0 Anwendungen bieten sich Marktforschern Chancen zu Auswertungen und Analysen, die es so zuvor nicht gab. Es müsste doch möglich sein Nutzern nachzuspüren, den Versuch zu unternehmen aus ihren Spuren Schlüsse zu ziehen und Anhaltspunkte zu aktuellen Entwicklungen oder Trends zu gewinnen. Die in Web2.0 Anwendungen auftretenden Datenmengen machen eine manuelle Auswertung sehr mühsam. In dieser Arbeit wird ein System zur automatisierten Unterstützung dieser Vorhaben entwickelt.

Um die praktische Relevanz und Nützlichkeit meines Systems deutlich zu machen, möchte ich in die Bereiche Marketing / Management und Marktforschung / Futorologie² einführen. Dabei werden die für diese Arbeit und ihr Umfeld relevanten Zusammenhänge und Methoden kurz vorgestellt.

2.1 Marketing

Marketing ist der Prozess des Planens und Ausführens von Konzeption, Preisgestaltung, Werbung und Verteilung von Ideen, Gütern oder Diensten, um Handel zu treiben, der individuelle und firmenbezogene Interessen befriedigt. Klassisch werden diese Ziele als die „vier P“ des Marketing verstanden: Product, Price, Promotion und Placement (vgl. McDaniel und Gates 2006).

Aufgabe des Marketings ist es, in diesen vier Bereichen optimale Strategien zu finden, um Handel und Profit zu optimieren. Die konkreten Ausprägungen von Merkmalen in

²auch Futures Research oder Future Studies

diesen vier Bereichen bezeichnet man als den Marketing Mix. Um Handel zu stimulieren versuchen Marketing Manager den Marketing Mix zu optimieren. Um die 'richtigen' Marketing Entscheidungen treffen zu können, werden aktuelle und relevante Informationen benötigt. Sind diese Informationen nicht verfügbar, werden Marktforscher beauftragt, sie zu erheben.

2.2 Marktforschung

Marktforschung ist eine dienende Disziplin. Sie liefert Informationen an Marketing Manager, damit gut informierte Marketing-Entscheidungen getroffen werden können.

Marktforschung befasst sich mit dem Planen von Untersuchungen und dem Sammeln und Analysieren von Daten, die für Marketing-Entscheidungen wichtig sind. Auch die Kommunikation der Ergebnisse an das Management gehört zu den Aufgaben des Marktforschers (McDaniel und Gates 2006).

Zu den klassischen Methoden der Marktforschung zählen Umfragen, Beobachtungen, Gruppendiskussionen und Experimente. Es wird zwischen qualitativer und quantitativer Forschung unterschieden.

Quantitative Forschung wird verwendet, um gesicherte, reproduzierbare kausale Zusammenhänge aufzudecken. Dazu werden statistische Methoden und große repräsentative Datensätze verwendet.

Qualitative Forschung ist in der Regel nicht reproduzierbar. Intuition und Gespür des Marktforschers spielen bei qualitativen Methoden eine besondere Rolle. Mit qualitativen Methoden wird nach tiefer liegenden Gefühlen, Motivationen und Bedeutungen gesucht. Im Vergleich zu quantitativen Methoden werden weniger Probanden verwendet. Eine wichtige Methode sind Gruppendiskussionen mit geringer Teilnehmerzahl, so genannte Fokus-Gruppen.

Häufig werden qualitative Studien und quantitative Forschung kombiniert, um konkrete Fragestellungen beantworten zu können. Man könnte z.B. zuerst eine qualitative Methode verwenden, um einen ersten Eindruck zu einem Thema zu bekommen. Mit diesem explorativem Vorgehen lässt sich dann eventuell der Fragenkatalog einer nachfolgenden quantitativen Umfrage kürzen. Da kürzere Fragebögen zu besseren Antworten führen³, kann so die Qualität der quantitativen Studie verbessert werden.

2.2.1 Milieus

Ein wichtiger Aspekt im Marketing ist das Identifizieren von Zielgruppen. Als nützliches Handwerkszeug dazu haben sich die Verbraucherklassifizierungen in Form von Milieus durchgesetzt. Sie werden oft als sogenanntes Kartoffel-Diagramm dargestellt. In Abbildung 1 sind die Sinus-Milieus zu sehen. Diese Standardgruppierung der Gesellschaft ist

³ Da Probanden eine höhere Bereitschaft haben, kürzere Fragebögen zu beantworten, kann man mit kurzen Fragenkatalogen die Rücklaufquote einer Umfrage erhöhen.

länderspezifisch. Die Milieus sind keine statische und eindeutige Klassifizierung und auch die Grenzen zwischen den einzelnen Milieus verlaufen fließend. Deswegen ist eine laufende Aktualisierung der Milieus sinnvoll. Aktuelle Sinus Milieus und ihre Definitionen finden sich bei Sinus-Sociovision (2007). In Abbildung 1 sind die Sinus-Milieus 2007 für Deutschland abgebildet.

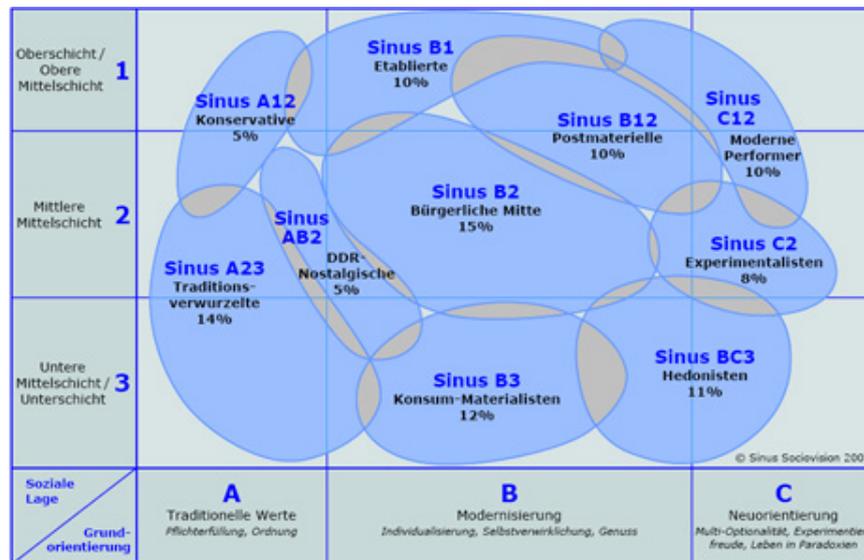


Abbildung 1: Die Sinus-Milieus, Quelle: Sinus-Sociovision (2007)

Eine neuere Entwicklung sind die Sinus-Meta-Milieus in Abbildung 2. Sie werden aus den länderspezifischen Sinus-Milieus abgeleitet und für grenzenübergreifendes Marketing genutzt. Die Sinus-Milieus sind nicht die einzige Klassifizierung von Verbrauchern. In verschiedenen Branchen oder Kontexten sind andere Klassifizierungen gebräuchlicher. In der Automobilbranche kommen Sigma-Milieus (SIGMA 2007) zum Einsatz. Für Online Marketing bietet die United Internet Media AG (2007) spezielle Web-Milieus.

Diese Verbraucherklassifizierungen nutzen das gleiche Koordinatensystem. Auf der X-Achse wird die Werteorientierung über Traditionell, Modern bis zu Postmodern oder Reorientierung aufgetragen. Die Y-Achse zeigt den Sozialen Status über Unterschicht / Untere Mittelschicht, Mittlere Mittelschicht zur Oberen Mittelschicht / Oberschicht an. Ein Vergleich der verschiedenen Verbraucherklassifizierungen deckt große Ähnlichkeiten in den identifizierten Milieus auf.

Sieben Meta-Milieus in Westeuropa

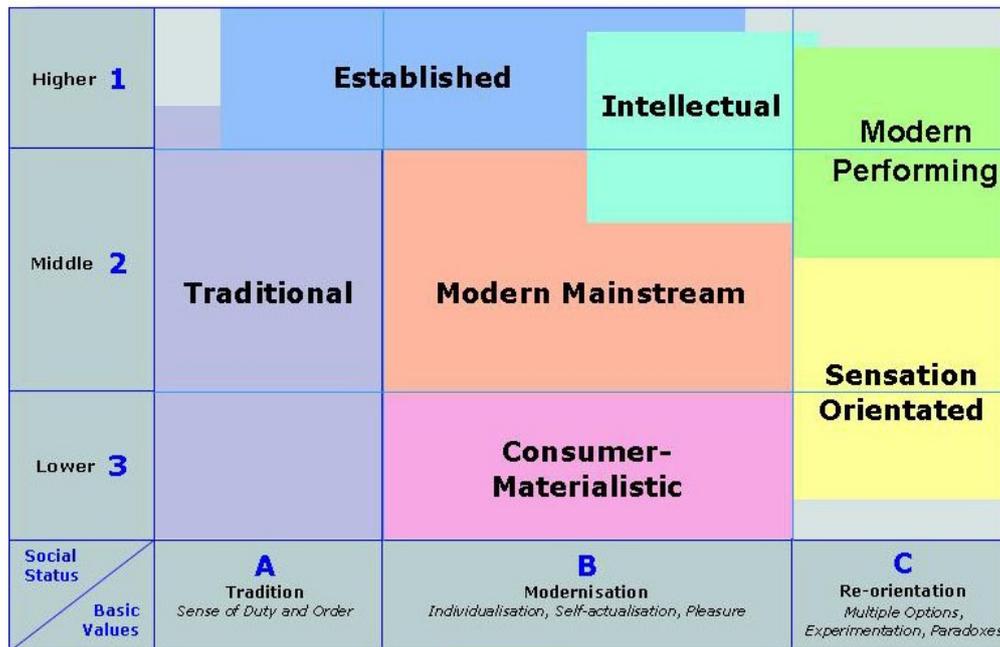


Abbildung 2: Die Sinus-Meta-Milieus, Quelle: Sinus-Sociovison (2007)

2.2.2 Produktdiffusion

Ein für Marktforscher sehr interessanter Prozess ist die Produktdiffusion. Everett M. Rogers definiert Diffusion als sozialen Prozess :

Diffusion is the process by which an innovation is communicated through certain channels over time among the members of a social system.

Rogers (1995, S.11)

Wenn ein neues Produkt oder eine neue Idee erfolgreich in den Markt gebracht wird, dann wird sie von verschiedenen Verbrauchern verschieden schnell angenommen. Die an der Produktdiffusion beteiligten Verbraucher lassen sich zeitlich klassifizieren. Produktdiffusion ist die Funktion der Annahme einer Innovation innerhalb eines sozialen Systems über die Zeit. Rogers (1995, S.252) unterteilt die Verbraucher dabei nach dem Kriterium der Innovationsfreudigkeit. Individuen unterscheiden sich in ihrer Offenheit für neue Ideen und Produkte. Innovationsfreudigkeit ist der Grad, mit dem ein Individuum relativ gesehen früher neue Ideen und Produkte annimmt, bezogen auf andere Mitglieder des gleichen sozialen Systems (Rogers 1995, S.22). Rogers (1995, S.262) definiert fünf Kategorien von Aneignern (engl. Adopters), die sich in ihrer Einstellung zu Neuem unterscheiden.

Produktdiffusion findet dabei wie in Abbildung 3 dargestellt statt. Die Diffusion ähnelt stark einer Normalverteilung. Zu Beginn greifen Innovatoren das Produkt auf. Die verschiedenen Aneignerkategorien lassen sich auch in Prozenten der letztendlichen Verbreitung quantifizieren. Risikofreudige Innovatoren stellen ca. 2,5% der Zielpopulation, auf sie folgen sozial respektierte Early Adopters mit ca. 13,5% aller Aneigner. Die Frühe Mehrheit, für die vor allem praktische Überlegungen relevant sind, schliesst sich mit 34% aller Aneigner den Early Adopters an. Auf sie folgt die skeptische späte Mehrheit, die 34% aller Aneigner stellt. Die Gruppe der Nachzügler gilt als sehr zögerlich, sie stellen die letzten 16% aller Aneigner, die eine neue Idee oder ein neues Produkt annehmen.

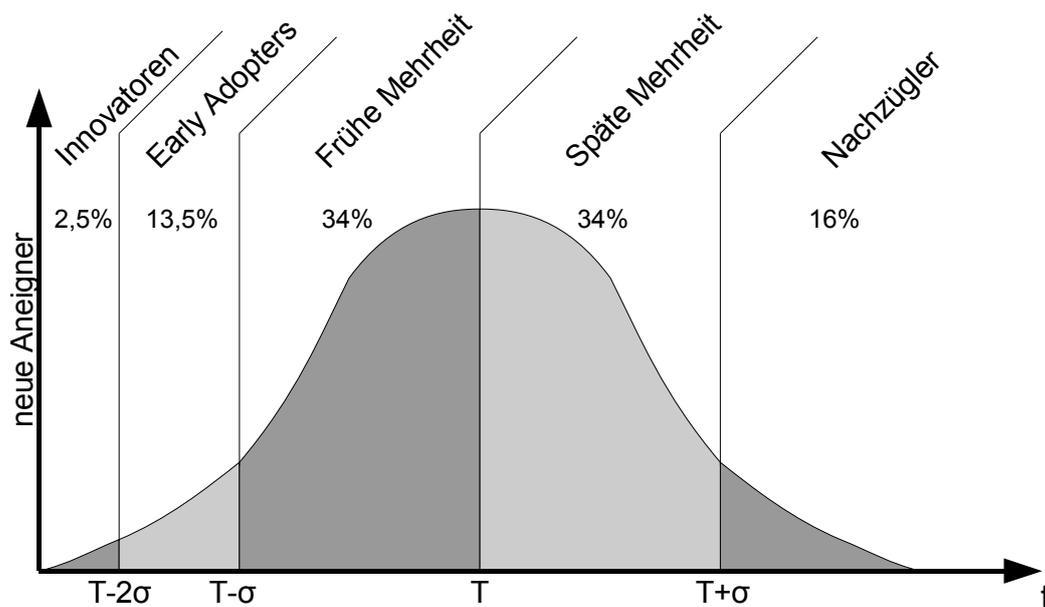


Abbildung 3: Produktdiffusion Quelle: Rogers (1995)

Die Produktdiffusion verläuft also von den Innovatoren über die Early Adopters bis hin zu den Nachzüglern. Allerdings verläuft nicht jede Innovation erfolgreich. Besonders interessant für den Erfolg einer Produktdiffusion sind die Early Adopters. Rogers definiert sie als :

Whereas innovators are cosmopolites, early adopters are localites. This adopter category, more than any other, has the greatest degree of opinion leadership in most systems. Potential adopters look to early adopters for advice and information about the innovation. The early adopter is considered by many as "the individual to check with" before using a new idea. . . .

. . . Because early adopters are not too far ahead of the average individual in innovativeness, they serve as a role model for many other members of a social system.

Rogers (1995, S.264)

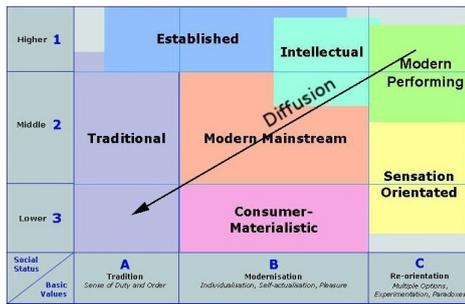


Abbildung 4: Diffusion

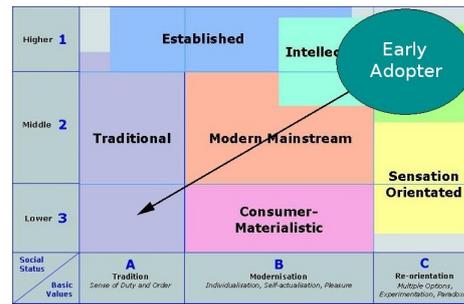


Abbildung 5: Early Adopters

Da die Akzeptanz neuer Produkte oder Ideen auch von den Werten und Einstellungen der Zielperson abhängt, und da die Milieus nach Werten klassifiziert sind, kann man den zeitlichen Verlauf der Diffusion im Milieu-Diagramm verorten. Die Diffusion neuer Produkte verläuft, wie in Abbildung 4 dargestellt, in der Regel vom oberen, rechten Abschnitt C1, zum Abschnitt A3 unten links. Die in Abbildung 5 markierten Verbraucher spielen dabei eine besondere Rolle. Aufgrund ihrer Werte und Einstellungen stehen sie Neuem sehr offen gegenüber. Sie werden Early Adopters genannt, weil sie sich als Erste auf neue Produkte und Ideen einlassen. Produkte die es schon bei Early Adopters schwer haben sich durchzusetzen, werden auch im Massenmarkt nur geringe Chancen auf Erfolg haben. Dieser Zusammenhang macht Early Adopters für Marktforscher sehr interessant. Was bei ihnen nicht ankommt, wird von ihnen nicht weiter empfohlen. Wahrscheinlicher ist sogar, dass sie dann negative Urteile an ihr Umfeld kommunizieren. Early Adopters können die Produktdiffusion also hemmen oder fördern. Durch ihre Beobachtung und Befragung kann man wertvolle Informationen zu den Stärken und Schwächen eines Produktes erlangen. Mit allgemeinen Beobachtungen von Early Adopters können frühzeitig Informationen zu möglichen Trends gewonnen werden, die erst später den Massenmarkt erreichen.

2.2.3 Marktforschung Online

Auch für Marketing und Marktforschung wird das Internet immer wichtiger. Online-Werbung und E-Commerce sind nur die ersten, offensichtlichen Bereiche. In der Marktforschung werden inzwischen viele der klassischen Methoden online durchgeführt.

A 2003 survey of marketing research suppliers found that 73 percent were currently conducting Internet survey research. One reason for the mushrooming growth of Internet research is that the online population now closely mirrors the general population.

McDaniel und Gates (2006)

Die Nutzung des Internets zur Marktforschung spart Kosten, erhöht die Zahl der potentiellen Teilnehmer einer Untersuchung und beschleunigt die gesamte Methode. Online-

Umfragen, Online-Fokusgruppen und Clickstreams⁴ sind nur der Beginn einer Entwicklung, in deren Verlauf auch die Auswertung von User Generated Content Systemen eine wichtige Rolle spielen wird.

2.3 Trend- und Zukunftsforschung

Während das Ziel der Marktforschung hauptsächlich in der Beschreibung des Ist-Zustands liegt, geht die Trendforschung einen Schritt weiter. Trendforscher versuchen vorherzusagen, wie die Welt morgen aussehen wird. In die Zukunft zu schauen ist keine exakte Wissenschaft, es ist nicht möglich, definitive Vorhersagen zu machen. Es ist aber möglich, wahrscheinliche Szenarien zu erstellen.

Dem (hier interessierenden) Verständnis von Trendforschern zufolge sind — ihre publizierten Auskünfte sozusagen auf den kleinsten Nenner gebracht — unter „Trendforschung“ diejenigen Aktivitäten zu verstehen, die auf die (Früh-) Erkennung, Benennung und Bewertung von sozialen und kulturellen Entwicklungen bzw. Veränderungen abzielen, wobei hinsichtlich der zeitlichen Ausdehnung und der räumlichen Reichweite von Trends als auch im Hinblick auf das Ausmaß bzw. die Schwere ihrer Konsequenzen von Fall zu Fall zu differenzieren ist.

Pfadenhauer (2004, Absatz 7)

Die Notwendigkeit frühzeitig Trends zu erkennen, um auf sie reagieren zu können, wird besonders deutlich an Firmen, die Produkte mit langen Entwicklungszyklen herstellen und vertreiben. Beispielsweise vergehen von der Konzeption bis zur Serienproduktion eines Autos fünf Jahre. In dieser Zeit kann sich der Zielmarkt deutlich verändert haben.

Als Verbraucher mag man sich wundern, warum selbst im neusten Auto dem Radio ein USB- oder iPod-Anschluss fehlt. Der Eindruck, Automobilhersteller wären dem Stand der Technik etwas hinterher, erklärt sich aus diesen langen Entwicklungszeiten. Hätte ein Automobilhersteller vor fünf Jahren den iPod-Trend erkannt, wäre er in der Lage gewesen darauf zu reagieren. Es wäre beispielsweise möglich gewesen, vorbereitende Maßnahmen beim Cockpit-Design einzuleiten, um eine iPod-Integration leicht nachrüstbar zu machen, sollte sich der Trend bestätigen und sich der iPod im Massenmarkt durchsetzen.

Auch Ansoff (1975) beschreibt die Notwendigkeit, als Firma oder Organisation frühzeitig auf entstehende Probleme zu reagieren. Beispiele für verpasste Gelegenheiten oder ignorierte Gefahren gibt es genug.

Hätte IBM den Trend zum Personal Computer und die Bedeutung des PC-Betriebssystems erkannt, wären sie nicht auf die Idee gekommen, das Disc Operating System (DOS) im Outsourcing an Microsoft in Auftrag zu geben. Dann hätte es vielleicht

⁴Aktivitäten, die ein Nutzer bei der Verwendung seines Web-Browsers auf Webseiten durchführt, können als Clickstream gespeichert werden. Bei den aufgezeichneten Aktivitäten handelt es sich in der Regel um eine Folge von Mausklicks und ihren Koordinaten. Die damit möglichen Analysen von Nutzerverhalten können z.B. zu Design-Verbesserungen genutzt werden.

statt einem MS-DOS ein IBM-DOS gegeben, mit dem sich IBM eine Chance auf die Marktposition hätte sichern können, die nun Microsoft inne hat.

Organisationen ruhen sich häufig auf dem Status Quo aus und planen auf Basis ihrer bisherigen Entwicklung voraus in die Zukunft. Bei einer aufwendigen periodischen Planung haben Organisationen kaum Chancen, auf schnelllebige Trends zu reagieren. Ein großes Problem für frühzeitige Reaktionen stellen unvollständig vorhandene Informationen dar. Managern fällt es leichter, auf sogenannte *Starke Signale* zu reagieren. Starke Signale sind definitive Anzeichen für eine Aktion, sind vollständige Informationen, mit denen man Marketing-Entscheidungen leicht fällt, da sie gut kalkulierbar sind.

Um frühzeitig reagieren zu können, benötigen Firmen eine Methode, die in der Lage ist, auch für unvollständige und unzuverlässige Informationen angemessene Reaktionen zu finden. Igor H. Ansoff (1975) schlägt ein Strategic Issue Management System vor. Mit diesem System können Firmen oder Organisationen auch auf noch nicht ausformulierte, voll definierte Gelegenheiten oder Gefahren frühzeitig und angemessen reagieren. Ein wichtiger Teil des Strategic Issue Managements ist das Aufspüren von *Schwachen Signalen* (Weak Signals).

Weak Signals are those ambiguous and controversial bits of information about the competitive environment that are typically hidden among the “noise” of the prevailing sensemaking paradigm and gradually coalesces to form a pattern ...

Seeing the future in Weak Signals, Saul (2006)

Diese Muster können *Trends* sein. Aus schwachen Trends können sich starke Trends entwickeln, die für die weitere Entwicklung einer Firma, Organisation oder sogar unserer Gesellschaft von Bedeutung sind. Je früher Trends und ihre Bedeutungen erkannt werden, umso besser vorbereitet kann man ihnen begegnen. Trendforscher versuchen, durch Aufspüren und Beobachten von Schwachen Signalen Trends zu identifizieren.

Trendforscher sind Experten. Sie arbeiten sich in Themenbereiche ein, um die Zusammenhänge und Muster zu lernen. Nur so sind sie in der Lage, ihre Intuition zu schulen und dann Muster in Schwachen Signalen zu sehen, darin Trends zu erkennen und zu verifizieren (vgl. Horx und Wippermann 1996, S.82).

2.3.1 Wissenschaftlichkeit der Trendforschung

Der Ruf nach der Intuition des Trendforschers scheint unwissenschaftlich. Neuere Forschung zur Intuition legt nahe, dass es sich bei Intuition um eine Art fortgeschrittener Mustererkennung handeln könnte (siehe Klein 2003). Analytische Methoden haben lange eine so große Aufmerksamkeit auf sich gezogen, das es keine verbreitete Vorstellung und Erklärung gibt, was Intuition ist und wie sie funktioniert. Gary Klein gibt in seinem Buch, „The Power of Intuition“, folgende Definition :

Intuition is the way we translate our experiences into judgements and decisions. It’s the ability to make decisions by using patterns to recognize what’s going on in a situation and to recognize the typical action script with which

to react. Once experienced intuitive decision makers see the pattern, any decision they have to make is usually obvious.

Klein (2003, S.23)

Klein (2003, ch.2) gibt ein gutes Beispiel zur funktionsweise von Intuition. Er beschreibt das Verhalten zweier Krankenschwestern auf einer Intensivstation für Neugeborene. Eine der beiden, Darleen, arbeitete schon einige Jahre auf der Station, während die zweite Krankenschwester, Linda, erst ein paar Monate dabei war. Eine der wichtigsten Aufgaben auf der Intensivstation für Neugeborene ist es, Blutvergiftungen früh zu erkennen. Da das Immunsystem zu früh geborener Babys noch sehr schwach ist, müssen Blutvergiftungen frühzeitig erkannt und schnellstens mit Antibiotika behandelt werden, um betroffenen Babys eine Überlebenschance zu geben. Die Schwierigkeit, eine Blutvergiftung bei Neugeborenen zu diagnostizieren, liegt vor allem in zuerst unspezifischen Symptomen, die nur schwache Signale für eine Blutvergiftung sind. Ein Nachweis durch Blutkultur dauert 24 Stunden. Bei eindeutigen Anzeichen, starken Signalen für eine Blutvergiftung, ist es für das betroffene Baby zu spät. Erfahrene Krankenschwestern kennen die schwachen Signale und sind somit in der Lage das Muster oder den Trend zu erkennen.

Eines der Babys unter Lindas Aufsicht war Melissa. An diesem Tag war Melissa etwas ruhiger als sonst. Linda nahm es als Zeichen, dass es Melissa besser zu gehen schien. Zur Fütterung schien Melissa ein wenig lethargisch, aber wer ist das nicht mitten in der Nacht, dachte sich Linda. Melissas Temperatur war an diesem Tag etwas niedrig, nicht außerhalb des normalen Bereichs, aber doch etwas niedrig, so dass Linda den Brutkasten wärmer stellte. Von einer Blutabnahme klebte noch ein Plaster an Melissas Verse, ein kleiner Blutfleck war auf dem Pflaster zu sehen, was vorkommen kann, wenn der Einstich nachlässig durchgeführt wird. Ein guter Einstich hätte nicht geblutet.

Melissa war Lindas Patient, und sie schlief ruhig. Gegen Ende der Schicht ging Darleen zufällig an Melissas Brutkasten vorbei und merkte sofort, dass etwas nicht in Ordnung war. Irgendwie sah Melissa komisch aus, sagte Linda später. Daraufhin schaute sich Darleen Melissa genauer an. Ihr fiel das Plaster mit dem Blutfleck auf. Irgendwie schien Melissa fehlfarben, eher grau statt eines gesunden Pink, und ihre Haut schien marmoriert. Auch war Melissas Bauch etwas gerundet. Ein Blick in Melissas Patientenblatt offenbarte Melissas niedrige Temperatur. Daraufhin erkundigte sich Darleen bei Linda, ob Melissa beim Füttern einen lethargischen Eindruck gemacht hätte. Mit Lindas Bestätigung war für Darleen klar, das sofort gehandelt werden musste, da sie in den Zeichen das Muster einer Blutvergiftungen erkannt hatte. Melissa wurden sofort Antibiotika verabreicht und nach ein paar Tagen war sie wieder wohlauf.

Gute Intuition braucht Expertise und eine Bereitschaft, Probleme erkennen zu wollen. Ohne Expertise mag man die Zeichen, die Schwachen Signale, wahrnehmen, kann sie aber meist nicht richtig interpretieren. Dann verpasst man, wie die Krankenschwester Linda im Beispiel, das Muster, weil man es nicht kennt.

Trends zu entdecken ist eine komplexe Aufgabe. Auch Tamás Kristófs Argument, „We might hypothetically argue that dealing with a high level of complexity requires the contribution of the human brain“ (vgl. Kristóf 2006), legt nahe, dass qualitative

Methoden und menschliche Intuition in der Trend- und Zukunftsforschung noch lange eine wichtige Rolle spielen werden.

2.3.2 Beobachtungsfelder

Igor H. Ansoff gibt schon 1975 einen Überblick über die zu beobachtenden Bereiche, in denen nach Schwachen Signalen Ausschau gehalten werden soll, wenn man Trends oder, wie Ansoff schreibt, „Unstetigkeiten“, entdecken will.

To broaden the awareness to include discontinuities, the firm must add special types of environmental analysis, such as environmental monitoring, technological forecasting, sociopolitical forecasting, and threat/opportunity analysis.

Ansoff (1975)

Die Beobachtungsfelder haben sich seitdem kaum verändert. Peter Saul schreibt in einem aktuellen Beitrag zur Zukunftsforschung :

...there is not yet a best practice approach to analysing new patterns that may be forming now on the edges of the social, technological, economic, environmental and political landscape ...

Saul (2006)

Obwohl es bisher keinen besten Ansatz zum Entdecken von Mustern oder Trends gibt, lassen sich zwei Aufgaben für Trendforscher definieren. Mit der Methode des Scannings werden Trends identifiziert. Die weitere Entwicklung identifizierter Trends wird durch Monitoring verfolgt.

Scanning ist ein Beobachten des geschäftlichen, technologischen und politischen Umfelds. Dadurch können Schwache Signale ausgemacht werden. Von den Eindrücken der Beobachtung lässt sich der Trendforscher inspirieren, er sucht nach Mustern, die ihm eine Trendidee geben. Im nächsten Schritt wird nach weiteren Anzeichen gesucht, die diesen Trend bestätigen. Ist ein Trend identifiziert, kann er benannt und in die Beobachtungsliste für das Monitoring aufgenommen werden.

Monitoring überwacht Trends in ihrer Entwicklung. Zu Beginn des Lebenszyklus eines Trends ist noch nicht absehbar, wie stark er werden wird und für welche Verbrauchergruppen er in welchem Umfang relevant sein wird. Auch die Lebensbereiche, in denen der Trend sich auswirkt, sind vorerst schwer abschätzbar. Damit sind auch die Konsequenzen und Entscheidungen, die eine Firma oder Organisation aufgrund des Trends treffen könnte, zu Beginn unklar. Es müssen durch Monitoring weitere Daten gewonnen werden, die helfen, bisher gesammelte Informationen zum identifizierten Trend zu vervollständigen. Die Informationen über einen Trend entwickeln sich somit von einem Wissenstand den Ansoff (1980) *sense of turbulence* nennt bis hin zu vollständigem Wissen. Vollständiges Wissen entspricht dem Wissensstand,

der nötig ist, um alle Folgen und Reaktionen der Firma oder Organisation mit Sicherheit kalkulieren zu können.

Im Monitoring werden auch Methoden aus der quantitativen Marktforschung eingesetzt. Man kann mit Gruppendiskussionen und Tiefeninterviews die Relevanz entdeckter Trends testen.

Auch gute Mathematiker verwenden Intuition, um zu entscheiden, wie sie ein bestimmtes Problem angehen. Diese Intuition wird durch Erfahrung gebildet. Durch das Lösen unzähliger mathematischer Probleme entstehen aussagekräftige mentale Modelle in denen Beziehungen und Strukturen dieser Probleme abgebildet sind (vgl. Klein 2003, S.72).

Auch Marketing Vertreter im produzierenden Gewerbe benötigen mentale Modelle der Herstellung von Bauteilen, um in der Lage zu sein, präzise Angebote abzugeben. Sie verstehen Zeit und Aufwand, der zur Anpassung von Produktionsausrüstung nötig ist. Auch die Einarbeitungsphase zur Produktion schwieriger Teile muss von ihnen berücksichtigt werden, andernfalls werden ihre Angebote auf unrealistischen und damit unwirtschaftlichen Kosten- und Zeitabschätzungen beruhen. Mentale Modelle sind unsere Vorstellungen zur Arbeits- und Funktionsweise von Prozessen. Sie leiten unsere Erklärungen und Erwartungen (siehe Klein 2003, S.27-28).

Mentale Modelle sind unsere Vorstellungen von der Welt und deren kausalen Zusammenhängen. Sie werden durch Erfahrung gebildet. Sie sind die Basis von Expertise und Intuition.

2.3.3 Zukunftsforschung

Gary Kleins Untersuchungen zur Intuition zeigen, dass entgegen der allgemeinen Annahme eines analytischen Denkens, Entscheidungen in der Mehrheit nicht durch den Vergleich von Alternativen, sondern intuitiv getroffen werden (vgl. Klein 2003, S.26). Dabei wird intuitiv eine mögliche Lösung ausgewählt und schrittweise in der Vorstellung bewertet. Der Entscheider simuliert den Ablauf der möglichen Lösung und prüft mit seinen Mentalen Modellen, ob es zu Problemen kommt. Diesen Prozess nennt Klein *mentale Simulation*. Werden Probleme gefunden, wird die Lösung angepasst. Treten zu viele Probleme auf oder sind diese in der gegebenen Zeit nicht lösbar, wird eine andere mögliche Lösung abgerufen und simuliert.

One of the best known futures approaches involves engaging employees and other stakeholders in creating different views of the possible futures (scenarios) that may confront an organisation.

Saul (2006)

Szenario-Techniken wurden zuerst von der US-Air Force während des 2. Weltkrieges eingesetzt (Hueck 2004), setzten sich dann aber allmählich auch in der Wirtschaft durch. Illustrativ für die Leistungsfähigkeit dieser Methode ist die Geschichte des Shell-Konzerns zur Zeit der Ölkrise.

Eine neue Dimension erlangten Szenarien Anfang der 1970er Jahre, als sich Pierre Wack mit einer Planungsgruppe von Royal Dutch/Shell darüber Gedanken machte, welche Ereignisse den Ölpreis beeinflussen könnten, der seit Ende des 2. Weltkrieges weitgehend stabil geblieben war. Ein Szenario beinhaltete eine plötzliche Verknappung des Öls mit entsprechend starken Preissteigerungen. Wack verpackte die damit verbundenen Konsequenzen in eine umfassende Geschichte über eine Welt mit hohen Ölpreisen. Die Manager erkannten die Brisanz dieser Entwicklung und erarbeiteten Konzepte, wie Shell auf solche Entwicklungen reagieren sollte.

Nach dem Yom Kippur-Krieg 1973 trat tatsächlich ein, was in Wacks Szenario beschrieben wurde. Der Shell-Konzern war vorbereitet und konnte im Vergleich zu den anderen Ölgesellschaften mit den Folgen der Ölkrise am besten umgehen und so seine Marktposition deutlich stärken.

Hueck (2004)

Szenario-Techniken werden eingesetzt, um mögliche Versionen der Zukunft zu erstellen und zu evaluieren. Besonders fruchtbar ist diese Methode, wenn alle Stakeholder vertreten sind und die kollektive Kraft ihrer Intuition, die sich auf die versammelten Mentalen Modelle stützt, zum Einsatz kommt.

2.4 Informations Arbitrage

Ähnlich den Produkten, die vom Startpunkt der Innovatoren über Early Adopters hin zu breiteren Schichten getragen werden können, breiten sich auch Ideen und Informationen aus. Sie werden erst produziert, dann kommuniziert und können im Laufe der Zeit ihren Weg bis in die Massenmedien finden. Eine interessante Rolle kommt hierbei User Generated Content und Web2.0 Anwendungen zu. Oft nehmen Nachrichten und Ideen von hier aus ihren Lauf und verbreiten sich wie in Abbildung 6 dargestellt. Spezialisierte Blogs mit kleinem Publikum, die dem Austausch von Informationen zwischen Experten dienen, sind häufig eine Quelle von aktuellen Informationen, die sich von dort zu weiteren Blogs mit größerem Publikum ausbreiten können. Über beliebte Top-Blogs und Internet-Nachrichten-Dienste finden sie ihren eventuellen Weg in lokale oder sogar globale Nachrichtendienste, um später auch in Zeitungen oder noch später in Zeitschriften aufgegriffen zu werden.

Um Informations Arbitrage erklären zu können, muss zuerst geklärt werden, was Arbitrage ist. Die Encyclopædia Britannica schreibt :

Arbitrage, business operation involving the purchase of foreign exchange, gold, financial securities, or commodities in one market and their almost simultaneous sale in another market in order to profit from the price differentials existing between the markets. . . . Arbitrage generally tends to eliminate price differences. . . .

Encyclopædia Britannica, Inc. (2003)

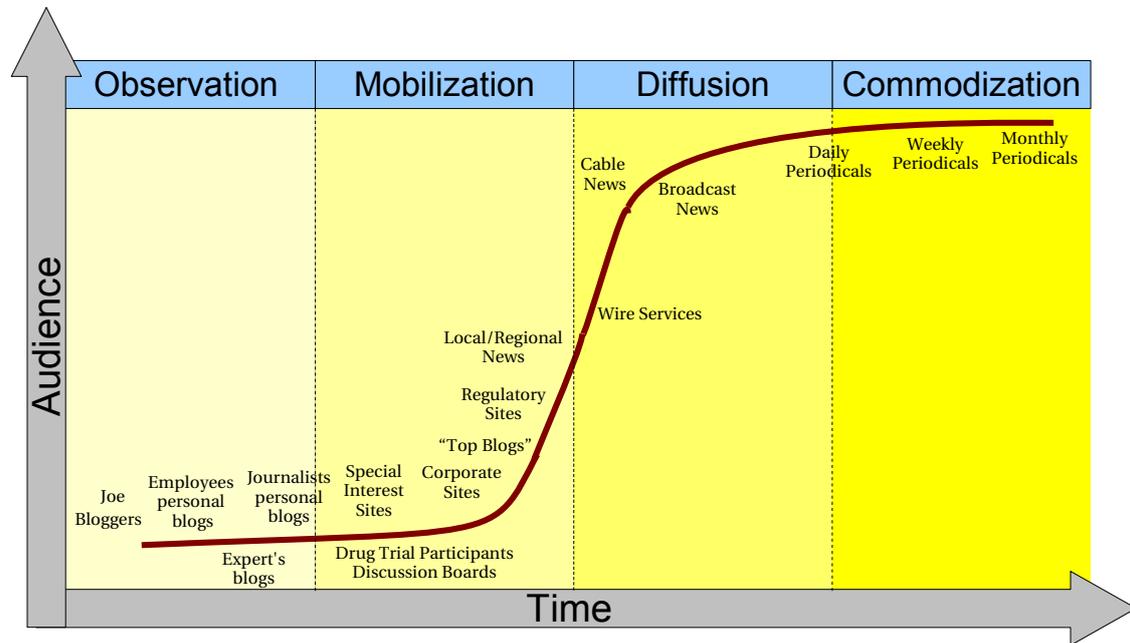


Abbildung 6: Informationsverbreitung, in Anlehnung an Monitor110 Inc. (2007)

Arbitrage ist also ein Begriff aus der Wirtschaft, der eine Situation beschreibt, in der ein Preisunterschied für ein Gut in zwei oder mehr Märkten ausgenutzt werden kann, um durch Handel Profit zu machen. Die Gelegenheit für Arbitrage Handel ist dabei zeitlich begrenzt, da sich die Preisunterschiede schnell ausgleichen.

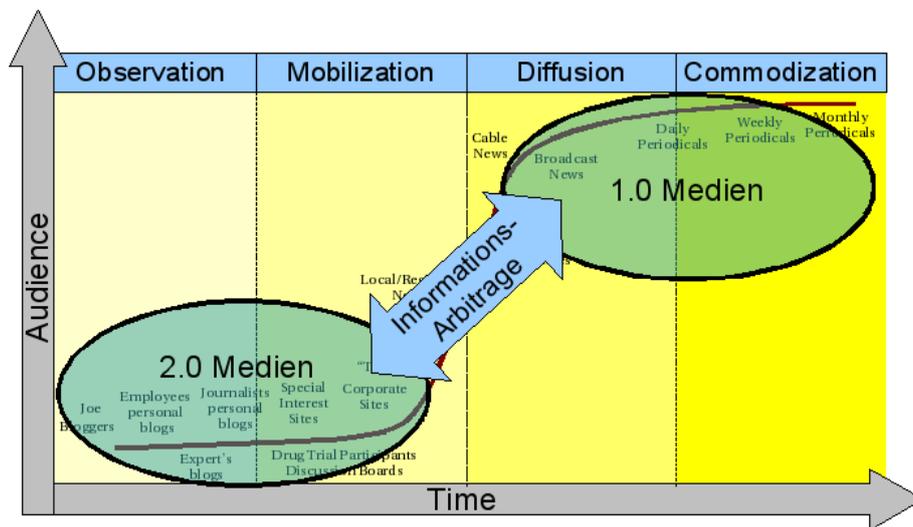


Abbildung 7: Informations Arbitrage

Informations Arbitrage nutzt Unterschiede in Informationsmärkten. Firmen wie z.B. Monitor110 Inc. nutzen dabei, wie in Abbildung 7 gekennzeichnet, die Auswertung von 2.0 Medien (Web2.0 Medien), um von wertvollen Informationen zu profitieren, bevor sie ihren Weg in die Massenmedien finden. Ein Beispiel verdeutlicht dieses Vorgehen. Man kann z.B. Angestellten-Blogs auswerten, um frühzeitig an Informationen zu gelangen, die für den Börsenkurs einer Firma relevant sein können. Findet man dabei hochaktuelle brisante Informationen, z.B. zu einem bevorstehenden Großauftrag, kann man reagieren und entsprechend mit den Aktien der Firma handeln und diese Information in Profit umsetzen.

Arbitrage kann zwischen Märkten auftreten, die schlecht miteinander verbunden sind, also eine Distanz zueinander haben. Die Entfernung zwischen Informationsmärkten kann man sich als eine Art kommunikative Distanz vorstellen. Informationsmärkte sind um so weiter voneinander entfernt, desto länger Informationen im Mittel brauchen, um vom Quellmarkt in den Zielmarkt zu gelangen. Je größer die Distanz zwischen zwei Informationsmärkten ist, um so besser stehen die Chancen für Informations Arbitrage.

Deswegen meint auch Scott (2004), dass Informations Arbitrage aufritt, wenn Leute absichtlich in fremde Territorien eindringen. Arbitrage kann nur zwischen verschiedenen Märkten auftreten. Auf dem ersten Markt müssen Informationen besorgt und auf dem zweiten angeboten werden, wenn man Gewinn erzielen will.

2.5 Zusammenfassung

Dieses Kapitel zeigte Felder auf, in denen User Generated Content verwendet werden, um angewandte Forschungsmethoden in den Bereichen Markt-, Trend- und Zukunftsforschung zu unterstützen. An dieser Stelle folgt eine Zusammenfassung der Kernaussagen dieser Ausführungen :

- Das Internet hat mehr als eine Milliarde Nutzer und wächst weiter.
- Web2.0 und User Generated Content treiben das Wachstum weiter an.
- Durch User Generated Content bieten sich neuartige Chancen zur Datengewinnung über User und deren Verhalten.
- User Generated Content ist für qualitative Methoden in Markt- Trend- und Zukunftsforschung nützlich.
- Das Internet senkt Kosten und steigert Geschwindigkeit von Methoden der Markt-Trend- und Zukunftsforschung.
- Scanning und Monitoring von Trends kann durch automatisierte Methoden unterstützt werden.
- Hoch aktuelle Informationen aus Web2.0 Anwendungen können für Informations Arbitrage genutzt werden.
- Early Adopters spielen eine sehr wichtige Rolle in der Produkt- und Ideendiffusion.

- In vielen Web2.0 und User Generated Content Anwendungen sind zur Zeit Early Adopters besonders aktiv.

Um mit User Generated Content Methoden Markt-, Trend- und Zukunftsforschung unterstützen zu können, müssen Forscher in der Lage sein, diese Informationen aufzunehmen und auszuwerten. Eine manuelle Sichtung der Daten ist zwar möglich, wird auch durchgeführt, ist aber aufgrund der Datenmengen sehr selektiv. Nur durch sinnvolle, leistungsfähige, computergestützte Methoden kann die Markt-, Trend- und Zukunftsforschung von der Menge der Daten profitieren. Ein solches System zur Auswertung von User Generated Content sollte versuchen, sich verschiedenen Idealen zu nähern, um den Nutzen der in diesem Kapitel beschriebenen Methoden zu maximieren.

Sollen die oben beschriebenen Methoden zur Markt- Trend- und Zukunftsforschung durch ein Auswertungssystem unterstützt werden, lassen sich folgende Anforderungen an das System und die zu sammelnden Daten stellen :

1. Die gesammelten Daten müssen aktuell und umfassend bzw. vollständig sein, um aktuelle Entwicklungen verfolgen zu können.
2. Die Daten sollten kontinuierlich aktualisiert werden.
3. Das System sollte das Aufspüren und Nachvollziehen aktueller Entwicklungen im Vergleich zu manuellen Methoden vereinfachen und beschleunigen.
4. Die gesammelten Daten müssen so beschaffen sein, dass sich aktuelle Entwicklungen aus ihnen ablesen lassen. Sie müssen bezüglich interessanter Themen aktuelle trendführende Daten enthalten.

3 Informationen und Gesellschaft

Now the clever remix of some political event or the latest song by your favorite band are not just something you can share with your friends. Digital technologies have made it simple to capture and share this creativity with the world. The single most important difference between the Internet ca. 1999 and the Internet circa today is the explosion of user generated creativity — from blogs, to podcasts, to videocasts, to mashups, the Internet today is a space of extraordinary creativity.

Lessig (2006, S.194)

Durch billige Computer und Internetanschlüsse haben sich die Rahmenbedingungen zur Produktion von Informationsgütern in den letzten zwei Jahrzehnten drastisch geändert. Mit dieser Entwicklung hat sich eine neue Organisationsform zur Produktion von Informationsgütern etabliert. User Generated Content Systeme sind ein Beispiel dieser Produktionsform, die Benkler (2002) Commons-based Peer Production nennt.

Digitale Informationsgüter unterscheiden sich in ihren ökonomischen Eigenschaften von physischen Gütern. Eine Betrachtung ökonomischer Eigenschaften digitaler Informationsgüter und des Zusammenhangs zu ökonomischer Wohlfahrt unterstreicht die Bedeutung der Allmende (engl. Commons) als Basis für die Peer Production von Informationsgütern.

Peer Production ist eine dezentrale Organisationsform der Produktion. Im Abschnitt Dezentralisierung werden zum Verständnis von Peer Production notwendige und in der Argumentation zu Organisationsformen gebräuchliche Kostenargumente an Beispielen vorgestellt.

Um User Generated Content besser zu verstehen, folgt ein Abschnitt zur Peer Production. Hier wird Peer Production kurz mit anderen Modi zur Produktion verglichen, und es werden Eigenschaftsdimensionen des mit Peer Production erstellten Informationsguts aufgezeigt. Sie finden im sich anschließenden Abschnitt zum User Generated Content ihre Anwendung im Vergleich verschiedener User Generated Content Arten.

Die Aufhebung der physischen und finanziellen Barrieren zur Informationsproduktion machen menschliche Kreativität zum primären Faktor in der Produktion von Informationsgütern. Neue Systeme und Organisationsformen zur Aggregation von Wissen entstehen aufgrund der veränderten Rahmenbedingungen im Internet. User Generated Content Systeme spielen dabei eine zentrale Rolle. Ein Überblick verschiedener User Generated Content Formen, ihrer Eigenschaften und ihres Potentials hinsichtlich der in Abschnitt 2 beschriebenen Methoden zur Markt-, Trend- und Zukunftsforschung bildet die Überleitung zum praktischen Teil dieser Arbeit.

3.1 Wohlfahrt und Wachstum

Wirtschaftswachstum ist, in der Art wie wir es gegenwärtig erleben, ein vergleichsweise junges Phänomen. In den Jahren zwischen 500 und 1500 wuchs die europäische Wirtschaft um ca. 0,1% pro Jahr, oder 10% pro Jahrhundert. Diese Zuwachszahlen erhöhten

sich im 18. Jahrhundert auf 0,5% pro Jahr. Seitdem hat sich das Wirtschaftswachstum weiter beschleunigt, man denke nur an das Wirtschaftswunder der 50er Jahre oder z.B. das Wirtschaftswachstum Thailands, welches innerhalb des letzten Jahrzehnts mehrheitlich zweistellige Zuwachsraten aufwies (vgl. Clarke und Islam 2004, S.31).

Es ist unbestritten, dass ein Zusammenhang zwischen Wirtschaftswachstum und Wohlfahrt besteht. Was man unter Wohlfahrt zu verstehen hat, bleibt leider oft unklar. Denn der Begriff Wohlfahrt wird häufig ohne klare Definition verwendet (vgl. Clarke und Islam 2004). Nichts desto trotz hat sich Wohlfahrt als nützliches Argument in politischen und gesellschaftlichen Diskussionen etabliert. Gesellschaftliche Wohlfahrt, deren Steigerung das Ziel jeder politischen Entscheidung sein sollte, setzt sich aus der individuellen Wohlfahrt aller betroffenen Bürger zusammen. Das Argument der gesteigerten Wohlfahrt bezeichnet somit einen gesellschaftlichen Zustand, in dem es den betroffenen Menschen nachfolgend im Schnitt besser geht.

Probleme ergeben sich sowohl beim Messen individueller Wohlfahrt, als auch bei der Aggregation zu gesellschaftlicher Wohlfahrt. Zum einen lässt sich individuelle Wohlfahrt schlecht quantifizieren, da Menschen beteiligt sind und psychologische Faktoren eine Rolle spielen. Zum anderen ist die Aggregation individueller Wohlfahrt problematisch, wie schon Überlegungen zur Verteilungsgerechtigkeit zeigen.

So würde eine politische Entscheidung, die einigen ausgewählten Bürgern die individuelle Wohlfahrt verbessert — indem sie danach z.B. mehr Geld zur Verfügung hätten — selbst für den Fall keinen Anderen schlechter zu stellen, die gesellschaftliche Wohlfahrt erhöhen. Sollte z.B. Bill Gates durch eine politische Entscheidung in die Lage versetzt werden, noch ein paar Milliarden Dollar mehr im Jahr zu verdienen, muss zwar kein Anderer etwas davon haben, die gesellschaftliche Wohlfahrt ist trotzdem gestiegen.

Trotz der mathematischen und wissenschaftlichen Probleme des Wohlfahrtsbegriffs

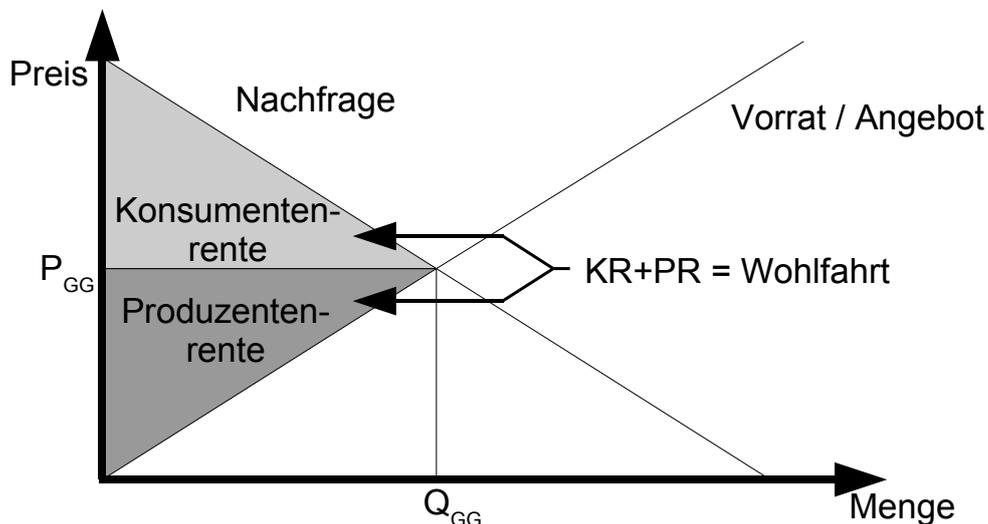


Abbildung 8: Konsumentenrente, Produzentenrente und Wohlfahrt

lässt es sich gut mit ihm argumentieren. In dieser Arbeit wird der Wohlfahrtbegriff nach der volkswirtschaftlich gängigen Definition ökonomischer Wohlfahrt benutzt.

Das Diagramm in Abbildung 8 veranschaulicht den Zusammenhang zwischen Preis und Angebot eines Produktes im idealen Konkurrenzmarkt. Preis und Menge eines nicht preisdiskriminierten Produkts pegeln sich am Gleichgewichtspunkt P_{GG}, Q_{GG} ein. Die Produzentenrente, der Mehrwert, der dem Produzenten aus dem Verkauf eines produzierten Guts bleibt, entspricht der Fläche, die von P_{GG} und der Angebotskurve begrenzt wird. Oberhalb der Produzentenrente zwischen P_{GG} und der Nachfragekurve findet sich die Verbraucherrente. Sie ist der Mehrwert, den Verbraucher durch Kauf des Gutes haben. Er entsteht bei Verbrauchern, die weniger für das Produkt zahlen müssen als es ihnen wert ist. Beide Flächen zusammengenommen werden volkswirtschaftlich als Wohlfahrt bezeichnet. Ökonomische Wohlfahrt ist die Summe aus Produzenten- und Verbraucherrente.

Im Allgemeinen wird ein positiver Zusammenhang zwischen Wirtschaftswachstum und gesellschaftlicher Wohlfahrt angenommen, wächst die Wirtschaft, wächst auch die Wohlfahrt.

Das Wirtschaftswachstum wird dabei über das Bruttonettoprodukt pro Kopf der Bevölkerung in diskreten Zeitabschnitten gemessen. Ein Steigen oder Fallen des Bruttonettoprodukts pro Kopf der Bevölkerung wird dann als Indiz für ein Steigen oder Fallen des gesellschaftlichen Wohlstands gewertet.

In den letzten Jahren wurde Kritik an dieser Vereinfachung des Zusammenhangs zwischen Wirtschaftswachstum und gesellschaftlicher Wohlfahrt geübt (vgl. Clarke und Islam 2004). Eine wachsende Wirtschaft hat nicht nur positive Auswirkungen auf individuelle und gesellschaftliche Wohlfahrt. Negative Effekte von Wirtschaftswachstum entstehen z.B. durch Umweltverschmutzung, Lärmbelästigung und Landschaftszerstörung.

3.2 Informationsökonomie

Zu Beginn dieses Abschnitts werden die Kategorien wirtschaftlicher Güter betrachtet, um die Basis zur Einordnung von Informationsgütern in dieses Schema zu legen. Mit dieser Einordnung werden die Eigenschaften von Informationsgütern erläutert. Eine Erläuterung und Darstellung zum Einfluss dieser Eigenschaften auf gesellschaftliche Wohlfahrt schließt sich an.

Wirtschaftliche Güter haben verschiedene Eigenschaften. Sie lassen sich mittels zweier Fragen klassifizieren (vgl. Mankiw und Taylor 2006, S.208).

Ausschließbarkeit: Kann verhindert werden, dass andere Menschen dieses Gut nutzen?

Rivalität: Beschneidet die Nutzung des Gutes durch einen Menschen die Nutzungsmöglichkeiten eines Anderen?

| | | ja | Rivalisierend? | nein |
|-----------|------|--|----------------|---|
| Exklusiv? | ja | Private Güter <ul style="list-style-type: none"> • Äpfel und Birnen • Kleidung | | Natürliche Monopole <ul style="list-style-type: none"> • Feuerwehr • Kabelfernsehen |
| | nein | Allmende Ressourcen <ul style="list-style-type: none"> • Fische im Meer • Umwelt | | Öffentliche Güter <ul style="list-style-type: none"> • Feuerwerke • Grundlagenforschung |

Abbildung 9: Vier Arten von wirtschaftlichen Gütern (vgl. Mankiw und Taylor 2006, S.208)

Mit diesen Fragen ergeben sich die in Abbildung 9 dargestellten Kategorien von Gütern.

Private Güter sind sowohl exklusiv als auch rivalisierend. Der Apfel, den ich gegessen habe kann von niemand Anderem gegessen werden, er ist damit exklusiv. Durch meinen Verbrauch gibt es einen Apfel weniger, damit sind Äpfel rivalisierend, da nun weniger Äpfel gegessen werden können.

Natürliche Monopole sind nicht rivalisierend aber ausschließbar. Sie entstehen häufig durch Economies of Scale Effekte⁵ die oft staatlich gestützt werden. Ein Beispiel ist die Feuerwehr. Die meiste Zeit hat die Feuerwehr in einer kleineren Stadt nichts zu tun, außerdem gibt es in der Regel nur eine. Trotzdem können Menschen vom Nutzen ausgeschlossen werden, wenn sich die Feuerwehr weigert, bestimmte Häuser zu schützen. Andere Beispiele sind das öffentlich-rechtliche Fernsehen in Deutschland, oder auch Post und Telekom vor der Öffnung der entsprechenden Märkte für Konkurrenten.

Allmende Ressourcen sind zwar rivalisierend, aber nicht ausschließbar. Fische im Meer können von Jedem gefangen werden. Wenn ich aber einen Fisch gefangen habe, dann schwimmt ein Fisch weniger im Meer. Man mag denken: „Was macht das aus?“. Die Rivalität wird spätestens klar, wenn man an die Fangflotten und das Problem der Überfischung denkt. Auch die Luft, die wir atmen, ist eine Allmende Ressource. Unausschließbar sind diese Ressourcen, weil niemand verhindern kann, dass Menschen fischen gehen oder atmen.

⁵Sinkt der Preis eines Gutes stark mit steigender produzierter Menge, kann ein Monopolist günstiger produzieren, als es mehrere Firmen könnten (vgl. Mankiw und Taylor 2006, Kap. 15).

Öffentliche Güter sind weder exklusiv noch rivalisierend. Wird ein Feuerwerk in einer Stadt abgebrannt, kann jeder in Sichtweite sich daran erfreuen. Durch das Betrachten des Feuerwerks wird keinem Anderen etwas genommen, der Konsum verbraucht das Gut nicht.

Schon Thomas Jefferson schreibt 1813 zu den Eigenschaften von Ideen :

If nature has made any one thing less susceptible than others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of every one, and the receiver cannot dispossess himself of it. Its peculiar character, too, is that no one possesses the less, because every other possesses the whole of it. He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me. That ideas should freely spread from one to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been peculiarly and benevolently designed by nature, when she made them, like fire, expansible over all space, without lessening their density in any point, and like the air in which we breathe, move, and have our physical being, incapable of confinement or exclusive appropriation.

Thomas Jefferson to Mr. Isaac M'Pherson 13. August 1813
Washington (1854, S.180-181)

Informationsgüter sind im Kern auch Ideen. Informationsgüter und Ideen ordnen sich somit in die Kategorie der öffentlichen Güter ein⁶. Sie sind nicht rivalisierend, da ich nicht weniger besitze, wenn ich meine Ideen offenbare. Und sie sind von Natur aus nicht exklusiv, denn sobald sie geäußert werden, habe ich keine Kontrolle mehr über sie.

Diese Eigenart von Informationsgütern führt unweigerlich zu einem Problem. Wenn Ideen und Informationsgüter nicht ausschließbar sind, wie können dann Produzenten motiviert werden, die hohen Kosten an Arbeit und Zeit auf sich zu nehmen, um neue Informationsgüter zu erschaffen?

Die etablierte Lösung, Erzeugern von Informationsgütern einen Arbeitsanreiz zu schaffen, liegt in der rechtlichen Konstruktion geistigen Eigentums. Das Urheberrecht in Europa und das Copyright in den Vereinigten Staaten räumen Informationserzeugern Verwertungsrechte ein. Dieses Recht, das durch Verträge und Gerichte durchgesetzt wird, schafft einen Anreiz, der Informationproduzenten dazu bewegt, weitere Informationsgüter zu erstellen.

Informationsgüter unterscheiden sich von realen physischen Waren in einem entscheidenden Punkt. Physische Waren werden durch Nutzung verbraucht. Der Apfel, den ich gegessen habe kann von niemand anderem gegessen werden. Angenommen, es gäbe keine

⁶Geistiges Eigentum, Urheberrecht und die daraus resultierenden Verwertungsrechte werden an dieser Stelle außen vor gelassen.

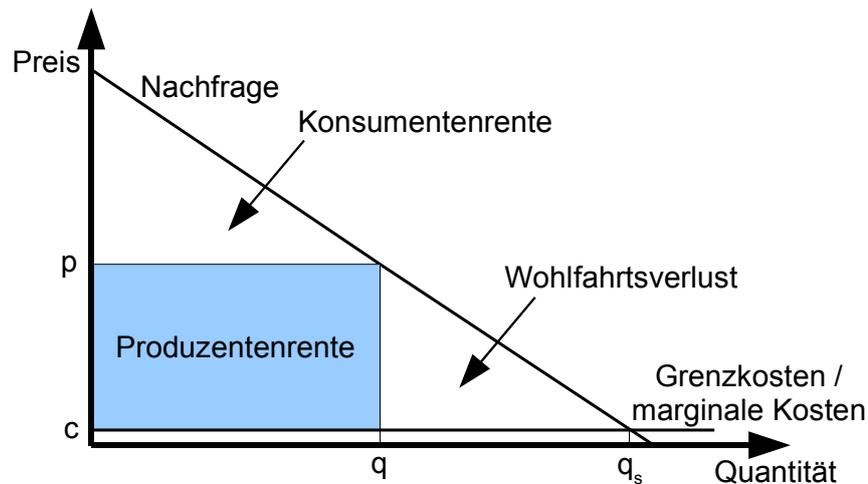


Abbildung 10: Wohlfahrtsverlust bei nahe null Grenzkosten nach Bakos und Brynjolfsson (2000)

weiteren Äpfel mehr, wollte ich einen weiteren Apfel essen, müsste er produziert werden. Diese Produktion verbraucht Ressourcen, beispielsweise Arbeit und Apfelbäume. Die nötige Arbeitskraft muss umgelenkt werden, um mehr Äpfel zu produzieren. Durch die erhöhte Apfelproduktion fehlen Ressourcen an anderer Stelle, aus dem Holz der Bäume, und mit der Arbeitskraft hätten z.B. Möbel gebaut werden können. Damit sind die sozialen Kosten der erhöhten Apfelmengen im Beispiel weniger Möbel.

Für Informationsgüter gilt dieser Zusammenhang nicht. Es fallen keine sozialen Kosten für weitere Einheiten an. Weder Zeit noch Arbeit müssen umgelenkt werden wenn ich mir Lessigs Free Culture (Lessig 2004) als PDF auf meinen Rechner lade. Wirtschaftlich argumentiert man mit den Fixkosten und den Grenzkosten.

Fixkosten sind die Produktionskosten der ersten Einheit. In der klassischen Produktion sind das hauptsächlich Entwicklungskosten. Die Kosten des Designs und des Tests von Produkten gehören zu den Entwicklungskosten.

Grenzkosten oder marginale Kosten entstehen von einer Einheit zur nächsten, nachdem die Entwicklungsphase abgeschlossen ist und sich das Produkt in der Produktion befindet.

Informationsgüter können sehr hohe Fixkosten haben, ihre Grenzkosten hingegen verlaufen nahe Null. Idealisiert kann auch von Null-Grenzkosten gesprochen werden.

Diese Null-Grenzkosten der Informationsgüter haben eine Auswirkung auf die oben definierte ökonomische Wohlfahrt. Lässt man die Fixkosten außen vor, kann aus Abbildung 8 die Abbildung 10 abgeleitet werden. Die Abbildungen weisen große Ähnlichkeit auf. Das Koordinatensystem ist das gleiche, auf der X-Achse wird wieder die Menge der Güter, auf der Y-Achse der Preis abgetragen. Auch die Nachfragekurve bleibt gleich.

Den unterschiedlichen Grenzkosten von physischen Gütern und Informationsgütern wurde durch Austausch der Angebotskurve mit der Null-Grenzkostenkurve digitaler Informationsgüter Rechnung getragen.

Basierend auf der Annahme, dass sich ökonomische Wohlfahrt maximiert, wenn Produkte im Markt zum Grenzkostenpreis gehandelt werden, ergibt sich durch die Null Grenzkosten von Informationsgütern ein Wohlfahrtsverlust (vgl. Bakos und Brynjolfsson 2000). Es ist in diesem Zusammenhang egal, zu welchem Preis Informationsgüter gehandelt werden, ihr Preis wird immer grösser sein als ihre Grenzkosten. Damit werden Nutzern deren, Wertschätzung dieses Gutes kleiner als der Preis ist, vom Informationsgut ausgeschlossen.

Brynjolfsson und Zhang haben den Zusammenhang von Grenzkosten, Anreizen und Wohlfahrt gut zusammengefasst :

On one hand, the near-zero marginal costs of reproducing digital goods suggest[ion]s that static welfare, the sum of consumer and producer surplus, would be maximized by making these goods available at zero price. In that way, all consumers with a value greater than the marginal cost, i.e. zero, would have access to them, and deadweight loss would be minimized. on the other hand, a zero price would provide no revenues to the creators of those goods, and thus no incentives for their creation in the first place, leading to potentially even larger losses of social welfare over time.

Brynjolfsson und Zhang (2007)

Der Zugang zu Informationen unterstützt den kreativen Prozess der Erstellung neuer Informationsgüter. Aus diesem Grund ist der durch den Preis beschränkte Zugang zu Informationen aus der Sicht gesellschaftlicher Wohlfahrt besonders schmerzlich.

Aber nicht nur durch die oben erwähnten Verwertungsrechte geistigen Eigentums und dem damit verbundenen Preis ergeben sich wirtschaftliche Einschränkungen bei der Informationsproduktion.

... the traditional price system not only results in significant deadweight loss, but also provides incorrect incentives to the creators of these innovations.

Brynjolfsson und Zhang (2007)

Die durch das Preisystem gesetzten Innovationsanreize geben der Weiterentwicklung von Informationsgütern eine bestimmte Stoßrichtung. Diese Anreize können dazu führen, dass Verbesserungen des Produktes nur für eine bestimmte Verbrauchergruppe vorgenommen werden.

Besonders deutlich wird diese Innovationsproblematik, wenn Produzenten in der Lage sind, mit gezielten Verbesserungen am Informationsgut Interessen ausgewählter Verbrauchergruppen zu bedienen. Zur Verdeutlichung denke man sich folgenden Fall :

Ein Softwarehersteller möchte investieren und seine Software verbessern, indem er sie erweitert mit dem Ziel

1. zufriedene Verbraucher noch zufriedener zu machen, oder

2. den Wert der Software für Verbraucher zu erhöhen, deren Wertschätzung gerade unter dem Marktpreis der Software liegt, so dass sie zu Käufern werden, oder
3. Verbesserungen einzuführen, um die Wertschätzung von Nicht-Käufern zu erhöhen, allerdings nicht genug, um sie zu Käufern zu machen.

Unter der Annahme eines begrenzten Budgets muss sich der Entwickler für einen der drei Punkte entscheiden. Obwohl Punkt eins und Punkt drei den Wert der Software erhöhen würden, wird der Entwickler den zweiten Punkt wählen, wenn er mit seiner Software und seiner Investition Geld verdienen will.

Zusammenfassung

In diesem Abschnitt zur Informationsökonomie wurden die Eigenschaften von Informationsgütern betrachtet und ihre Effekte auf die ökonomische Wohlfahrt gezeigt. Das Problem der Anreize zur Informationsproduktion wurde erläutert sowie die Ineffizienzen der Problemlösung durch geistiges Eigentum und Verwertungsrechte aufgezeigt. Die Kernaussagen der Argumentation lassen sich wie folgt zusammenfassen :

- Informationsgüter fallen ihrer Natur nach in die Kategorie der öffentlichen Güter. Sie sind weder rivalisierend noch ausschließbar. Daraus folgt das Problem der geringen Anreize zur Informationsproduktion.
- Dem Anreizproblem wird die Konstruktion des geistigen Eigentums entgegengesetzt. Die entstandenen Verwertungsrechte schaffen einen Anreiz zur fortlaufenden Produktion und Verbesserung von Informationsgütern.
- Preisbehaftete Informationsgüter führen zu Wohlfahrtsverlusten und ineffizienten Innovationsanreizen. Diese Wohlfahrtsverluste und Ineffizienzen sind zwei der Gründe, warum die Produktion von Informationsgütern durch Commons-based Peer Production erfolgreich ist.

Der folgende Abschnitt zur Dezentralisierung beschreibt und erklärt einen weiteren Faktor, der für den Erfolg und die Ermöglichung von Commons-based Peer Production Systemen von entscheidender Bedeutung ist. Dabei werden zwei Begriffe zur Kostenargumentation eingeführt.

3.3 Dezentralisierung

Zur Analyse von Organisationsformen werden Transaktionskosten und Kommunikationskosten als Argumente genutzt. Zuerst gilt es, diese Begriffe zu klären. Transaktionskosten wird als der allgemeinere der beiden Kostenbegriffe zuerst geklärt. Kommunikationskosten lassen sich als Teil von Transaktionskosten verstehen. Im Unterabschnitt Kommunikationskosten wird der Trend zur Dezentralisierung organisatorischer Strukturen mit Kommunikationskosten anhand von Beispielen argumentiert. Eine Übersicht über die Entwicklung von Kommunikationskosten innerhalb der letzten hundertfünfzig

Jahre anhand eines weiteren Beispiels schließt sich an. Einen besonderen Einfluss auf die gegenwärtige Kostenstruktur von Kommunikation hat das Medium Internet. Mit der Verknüpfung von Dezentralisierung und dem End-to-End Prinzip des Internets leitet dieser Abschnitt zur Commons-based Peer Production über.

3.3.1 Transaktionskosten

Wirtschaftswissenschaftler argumentieren gern mit Kosten. In der Argumentation zu Vor- und Nachteilen von Organisationsformen haben sich beginnend mit dem Artikel „The Nature of the Firm“ (Coase 1937) *Transaktionskosten* als Argument durchgesetzt⁷. Die Transaktion ist hier finanzell zu verstehen. Sie ist ein Handel, der den finanziellen Status von Handelspartnern verändert. Mit den Kosten der Transaktion ist aber nicht das Geld gemeint, das für die Ware oder Leistung an den Handelspartner abzutreten ist. Transaktionskosten sind die versteckten Kosten, die den Handelnden im Kontext der Transaktion entstehen.

Das folgende Beispiel verdeutlicht Transaktionskosten. Angenommen, ich will eine Banane kaufen. Als Kosten fällt nicht nur der Preis der Banane an, es wird auch Zeit und Energie verwendet um folgende Fragen zu klären :

- Welche Art Banane mag ich eigentlich?
- Wo bekomme ich die her?
- Wo ist sie wie teuer?
- Welche der vor mir liegenden Bananen kaufe ich jetzt?

Dazu kommen dann noch Kosten für Hin- und Rückfahrt zur Verkaufsstelle, eventuelle Wartezeiten, weil ich das Auto genommen habe und im Stau stehe, Gebühren am Geldautomaten, weil ich noch Geld brauchte und so weiter. Diese versteckten Kosten einer Kauf- bzw. Verkaufsbeziehung nennt man Transaktionskosten.

3.3.2 Kommunikationskosten

Ein großer Teil der Transaktionskosten sind Kosten zur Informationsbeschaffung. Dieser Teil fällt in den Bereich der *Kommunikationskosten*, mit denen Malone (2004) zu den sich verändernden Organisationsformen von Arbeit argumentiert.

Malone findet ein wiederkehrendes Muster in der Entwicklung von Organisationen. Organisationen entwickeln sich von lokalen Strukturen über zentrale Gebilde zu dezentralen Strukturen. Im Folgenden wird dieses Entwicklung, dieses Muster, anhand von zwei Beispielen dargestellt. Dabei wird zuerst auf gesellschaftliche Organisationsformen eingegangen. Im zweiten Beispiel wird das gleiche Muster an den Organisationsformen von Firmen gezeigt.

⁷Coase erwähnt die Transaktionskosten in diesem Artikel nicht mit Namen, analysiert aber bisher vernachlässigte Kosten und Aufwände der Organisationsformen Markt- und Firmenproduktion in Gegenüberstellung.

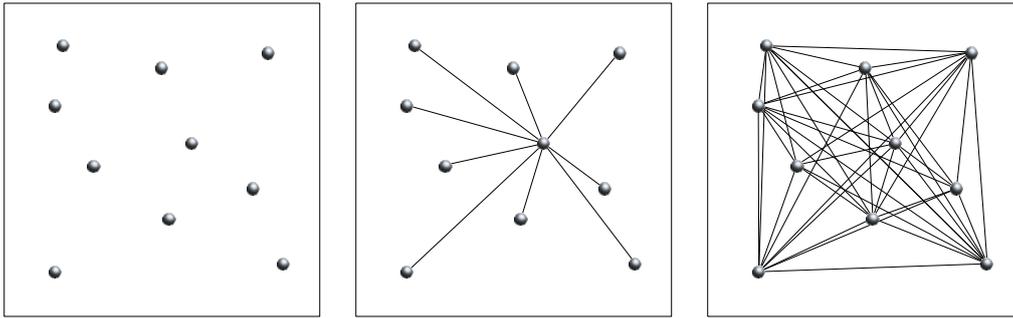


Abbildung 11: Lokal

Abbildung 12: Zentral

Abbildung 13: Dezentral

Malones erstes Beispiel deckt dieses Muster in *gesellschaftlichen Organisationsformen* auf. Als die einzige Kommunikationsmöglichkeit im Gespräch von Angesicht zu Angesicht bestand, lebten Menschen in kleineren lokalen Gruppen als Jäger und Sammler (vgl. Abbildung 11). Erst als sie lernten, über größere Entfernungen zu kommunizieren, konnten Königreiche entstehen, deren Größe enorme Vorteile, beispielsweise militärischer Art, mit sich brachten (vgl. Abbildung 12). Zentralisierung zur Nutzung von Größeneffekten hat aber ihren Preis in der Beschränkung persönlicher Freiheit.

Erst durch die Erfindung des Buchdrucks senkten sich Kommunikationskosten so weit ab, dass wieder mehr Menschen in den Genuss von für sie relevanten Informationen kamen. Besser informiert konnten sie sich auch politisch engagieren. Somit änderte sich mit der Innovation des Buchdrucks die Entwicklungsrichtung staatlicher Organisationsformen, hierarchisch zentralisierte Strukturen begannen sich zu dezentralisieren, so dass sich nun gesellschaftliche Strukturen bilden konnten, wie sie in Abbildung 13 gezeigt werden.

Aufklärung, Französische Revolution und die Gründung der ersten Demokratie der Neuzeit in Amerika sind ohne den Buchdruck kaum vorstellbar. Somit ist der Buchdruck und die mit ihm gesunkenen Kommunikationskosten einer der Faktoren, der die gesellschaftliche Organisationsform Demokratie in dieser Größenordnung⁸ überhaupt erst ermöglichte.

Das gleiche Muster findet Malone auch in der *Organisation von Firmen*. Gegen Ende des 18. Jahrhunderts waren Betriebe in der Mehrheit kleine, lokal operierende Familienunternehmen (siehe Abbildung 11). Sie ähnelten also in ihrer Struktur den Gruppen der Jägen und Sammler tausende Jahre zuvor. Erst zu Beginn des 20. Jahrhunderts veränderten sich die Kommunikationskosten durch Erfindungen wie Telegraphen, Schreib-

⁸Die erste überlieferte Demokratie im Staatstaat Athen (480-322 v.Chr.) verfügte über keinen Buchdruck und wurde durch die Arbeit von Sklaven gestützt. Damit blieb Bürgern genug Muße, an politischen Beratungen und Volksversammlungen teilzunehmen, also die hohen Kommunikationskosten einer Demokratie zu schultern. Ein weiterer Faktor, der die Athener Demokratie ermöglichte, war auch die begrenzte Bürgerzahl von bis zu 30000 Wahlberechtigten (vgl. Bleicken 1995, S.99). Die Zahl der aktiven Redner in der Volksversammlung ist allerdings im Vergleich deutlich geringer und wurde hauptsächlich durch Berufspolitiker oder Berufspredner gestellt.

maschinen und Durchschlagpapier. Firmen konnten nun zu vorher unpraktikabler Größe wachsen und zentralisierte, hierarchische Strukturen aufbauen (siehe Abbildung 12). Durch Größeneffekte konnten und können Firmen von dieser Entwicklung ökonomisch profitieren. Der Erfindung der Eisenbahn fällt dabei eine zentrale Rolle in der Reduzierung von Kosten zu, sie beschleunigte den Transport von Waren und Informationen und ermöglichte somit weiteres Wachstum.

| Medium und Jahrzehnt | Verzögerung | Preis |
|--------------------------|---------------|----------|
| ohne Eisenbahn, ca. 1840 | 260,3 Stunden | \$107,17 |
| mit Eisenbahn, ca. 1850 | 56,3 Stunden | \$85,17 |
| via Telegraf, ca. 1850 | 8,3 Stunden | \$750,00 |
| via E-Mail, ca. 2000 | ca. 0 | ca. 0 |

Tabelle 3: Entwicklung der Kommunikationskosten, Auszug aus Malone (2004, S.33)

Besonders deutlich werden die gefallenden Kommunikationskosten anhand des Beispiels vom Senden einer Textseite an einhundert Adressaten (vgl. Malone 2004, S.33). Angenommen, die Empfänger leben über die USA verstreut. Möchte man die Effekte der Kommunikationskosten in ihrer Entwicklung betrachten, so lohnt der Vergleich der Nutzung verschiedener Medien mit den entsprechenden Kosten des Versands zu verschiedenen historischen Zeitpunkten. Als Kosten werden hier zum einen die Verzögerung zwischen Versand und Empfang und zum anderen der Preis des Versendens erfasst. Die Tabelle 3 veranschaulicht diesen Vergleich⁹. Erfasst sind die Fälle kurz vor und nach der breiten Verfügbarkeit neuer Kommunikationstechnologien. Die Bandbreite der im Beispiel verwendeten Medien reicht vom Versand um ca. 1840 ohne Eisenbahn, bis zur Nutzung des E-Mail Systems in der Gegenwart.

Die verschwindend geringen Kommunikationskosten des E-Mail Systems sind Ausdruck einer drastischen Innovation, die bisherige Hemmnisse, sich dezentral zu organisieren, abbaut. Das E-Mail System ist nur eine Anwendung einer grösseren Innovation. Dabei handelt es sich um Internet basierte Anwendungen, deren Kommunikationskosten äußerst gering sind. Der Innovation Internet-basierter Kommunikation fällt auf Firmen-seite die Rolle zu, die der Buchdruck für die Organisation von Staaten spielte. Malone bringt verschiedene konkrete Beispiele für bereits existierende Ansätze dezentralisierter Organisationsstrukturen in Firmen, von denen eines folgend wiedergegeben wird, um den Bogen zur Dezentralisierung zu schließen.

Die AES Corporation ist einer der größten Stromversorger der Welt und das mit einer ungewöhnlichen Managementphilosophie. Statt auf zentrales hierarchisches Management zu setzen, befreit AES seine Mitarbeiter zu großen Teilen aus den Zwängen von Anordnungen. Bei AES wird davon ausgegangen, dass Eigenverantwortung Spaß und Engagement erzeugt. Die Überzeugung, dass selbst richtungsweisende Geschäftsentscheidungen

⁹Malone baut für dieses Beispiel auf Berechnungen einer Studie von Yates und Benjamin (1991, S.72). Die Berechnungen zum Versenden an einhundert Adressaten stammt von T. W. Malone selbst (vgl. Malone 2004, Fußnote 16 S.197)

nicht vom Top Management getroffen werden müssen, sondern Mitarbeiter in der Nähe des Problems oftmals besser in der Lage sind, konkrete Situationen einzuschätzen und nach dieser Einschätzung verantwortungsvoll zu handeln, wird bei AES gelebt. Die einzige, aber zentrale Regel, um bei AES Entscheidungen zu treffen ist: Man muss Rat einholen. Je größer die zu treffende Entscheidung, desto mehr Rat wird in der Regel eingeholt. Dieser kommunikationsintensive Prozess wird häufig über E-Mails abgewickelt. Und so wurde auch die größte wirtschaftliche Entscheidung der Firmengeschichte, der Kauf von Drax, einem Krafwerk in England, letztendlich von einer Person getroffen, die nicht im Management saß. Diese Person holte sich zwar Rat via E-Mails von verschiedenen, selbst gewählten Mitarbeitern der Firma, die Höhe des Angebots und die Kaufentscheidung wurde aber in Eigenverantwortung getroffen (vgl. Malone 2004, S.47-51). Damit ist AES ein Beispiel für dezentrale Organisationsformen von Firmen, jeder Mitarbeiter kann mit jedem kommunizieren, jeder Mitarbeiter trifft wichtige Entscheidungen selbst. Die Struktur von AES ähnelt also der Abbildung 13.

Diese extreme Delegation und Dezentralisierung ist nur ein Beispiel von vielen, mit denen Thomas W. Malone die neuen Möglichkeiten zur Strukturierung von Firmen und Produktion verdeutlicht. Im Abschnitt 3.4 wird dieses Muster wieder aufgegriffen. Dort wird der Effekt von Dezentralisierung auf die Modi zur Produktion von Informationsgütern beleuchtet.

3.3.3 Internet End-to-End

Die drastisch gesunkenen Kommunikationskosten der letzten Jahre sind auf Technologien zurückzuführen, die durch das Internet ermöglicht werden. Das zentrale Prinzip, durch welches das Netz zu seiner Größe wachsen konnte ist das End-to-End Prinzip. Lawrence Lessig schreibt :

The end-to-end argument says that rather than locating intelligence within the network, intelligence should be placed at the ends ... Thus complexity and intelligence in the network are pushed away from the network itself.

Lessig (2002, S.34)

Dieses Prinzip ist im Kern nichts anderes als Dezentralisierung. Auf den ersten Blick geht es zwar um die dezentralisierte Organisation eines Rechnernetzes, allerdings schlägt das End-to-End Prinzip und seine Freiheiten auf die Innovationsmöglichkeiten und Transaktionskosten von Entwicklern und Anwendern durch. Das End-to-End Prinzip setzt Innovationskräfte frei, indem es der Kreativität von Entwicklern freien Lauf lässt. Niemand muss um Erlaubnis gefragt werden, wenn es gilt, eine neue Anwendung auf dem Netz laufen zu lassen. Der dadurch entstandene und entstehende Anwendungsreichtum zeichnet das Internet aus und zieht immer mehr Nutzer an, deren Kreativität sich entfalten kann.

Zusammenfassung

Gerade durch User Generated Content, Web2.0 Dienste und Peer-to-Peer Anwendungen gewinnt das Prinzip der dezentralen Organisation auch für den einfachen Anwender immer mehr an Bedeutung.

Durch die gesunkenen Kommunikations- und Transaktionskosten werden dezentrale Organisationsformen möglich, mit denen bisher nicht nutzbare Kreativitätspotentiale realisiert werden können.

Die Möglichkeiten und Formen dezentraler Organisation sind vielseitig. Der nächste Abschnitt beschäftigt sich mit den Modi der Produktion von Informationsgütern. Dabei wird besonderen Wert auf Peer Production als dezentrale Organisation von Informationsproduktion gelegt. Die im nächsten Abschnitt beschriebenen Eigenschaften von Peer Production und mit ihnen produzierten Informationsgütern bilden die Grundlage des sich anschließenden Abschnitts zur Betrachtung von User Generated Content.

3.4 Peer Production

Dieser Abschnitt beginnt mit einer Übersicht zu Organisationsformen von Produktion. Besondere Aufmerksamkeit bekommt dabei der Modus der Peer Production. Eine Zusammenstellung und Beschreibung der für die Peer Production relevanten Eigenschaften von Informationsgütern schließt den Abschnitt ab und bildet den Rahmen zum Vergleich verschiedener User Generated Content Arten im darauf folgenden Abschnitt 3.5.

In der Produktion von Informationen und Waren kann man drei vorherrschende Organisationsformen oder Modi unterscheiden. Individuen arbeiten im Kontext der Firma als Mitarbeiter, im Markt als Marktteilnehmer oder staatlich organisiert z.B. als Beamte. Diese klassischen Modi der Produktion haben verschiedene Vorteile und Einschränkungen, die folgend kurz betrachtet werden.

Firma und Staat arbeiten zentral organisiert, Mitarbeiter sind in der Regel an Anweisungen gebunden. Durch Größeneffekte können Ressourcen genutzt werden, die den Individuen sonst nicht zur Verfügung stehen würden. So profitieren Entwickler z.B. vom Patentportfolio einer Firma. Die Einschränkung individueller Freiheit kann zu einem Verlust kreativen Potentials¹⁰ führen. Sowohl der Firma, als auch dem Staat stehen zur Produktion auf Basis vorhandenen Personals nur ein beschränkter Pool von Mitarbeitern und Fähigkeiten zur Verfügung.

Die *Produktion im Markt* ist dezentral organisiert und orientiert sich an Preissignalen. Dem Gewinn an unternehmerischer und individueller Freiheit stehen vielfältige Risiken gegenüber, die von zentral organisierten Produktionsstrukturen durch Größeneffekte aufgefangen werden. Der Zugang zu den zur Produktion notwendigen Ressourcen ist durch Preise geregelt und beschränkt. Im Zusammenhang mit der Produktion von Informationsgütern treten die im Unterabschnitt 3.2 beschriebenen Probleme auf. Die Nutzung des Preissystems für Informationsgüter führt zu ökonomischen Wohlfahrtsverlusten und teilweise ineffizienten Innovationsanreizen.

Neben diesen drei Modi der Produktion existiert der Modus der *Peer Production* von Informationen dessen wichtigstes Beispiel in der akademischen Idee und besonders in

¹⁰vgl. Abschnitt 3.3

wissenschaftlicher Forschung zu finden ist (vgl. Benkler 2002, S.381). Wissenschaftler wählen im akademischen Kontext Themen und Projekte zu großen Teilen selbst und produzieren Informationen, die frei verfügbar sein sollen.¹¹ Wichtige Qualitätskontrollen werden dabei über den Prozess der Peer Reviews sicher gestellt.

Neben akademischer und professioneller Produktion von Informationsgütern gab und gibt es den sehr wichtigen Bereich der nicht-professionellen Produktion von Informationen und Kultur. Durch Mund-zu-Mund Kommunikation, also persönliche Kommentare und Gespräche ohne kommerzielles Interesse, werden Informations- und Kulturgüter erzeugt, die einen großen Teil unserer Informationsumgebung bestimmen (vgl. Benkler 2002, S.282).

Die mit der Innovation Internetbasierter Kommunikation drastisch gesunkenen Kommunikations- und Transaktionskosten (vgl. Unterabschnitt 3.3) verändern die Möglichkeiten und Leistungsfähigkeit der Peer Production. Sowohl die Komplexität der produzierten Informationsgüter, als auch die Zahl der Beiträge und Akteure der Peer Production wachsen in bisher unerreichte Größenordnungen (vgl. Benkler 2002, S.383).

Bestes Beispiel für die gesteigerte Leistungsfähigkeit nicht-kommerzieller Peer Production ist die Softwareentwicklung im Open Source Modell. Zwei der erfolgreichsten Vertreter der im Open Source Modell entstandenen Produkte sind das Linux Betriebssystem und der Webserver Apache. Linux hat gerade im Servermarkt und bei Embedded Systems eine starke Marktposition. Der Webservermarkt wird von Apacheinstallationen dominiert, d.h. er liefert die meisten Webseiten an Browser aus. Bei beiden Beispielen handelt es sich um hoch komplexe und sehr leistungsfähige Softwareprodukte. Die Erstellung so komplexer und leistungsfähiger Software durch Peer Production Prozesse galt vor 20 Jahren noch als unvorstellbar.

Die Erfolge von in Peer Production entstandener Software haben in den letzten Jahren viel Aufmerksamkeit der Forschung bekommen. Von Krogh und von Hippel (2006) geben einen guten Überblick und Einstieg in relevante Artikel und Forschungsthemen zur Peer Production im Open Source Modell. Die Software Produktion im Open Source Modell ist aber nur ein, wenn auch wichtiges, Beispiel für Peer Production Systeme. In dieser Arbeit soll nicht weiter auf das Open Source Modell eingegangen werden. Stattdessen wird den User Generated Content Systemen, einer weiteren Kategorie von Peer Production, in dieser Arbeit mehr Aufmerksamkeit gewidmet.

Internetbasierte Peer Production hebt sich von den anderen Modi zur Informationsproduktion durch ihre Anreize zur Produktion und ihr besonderes Maß selbstbestimmten Handelns ihrer Akteure ab. Während in Firmen- und Staatenmodus auf Anweisung hin gehandelt wird und im Markt der Preis Ausschlag und Anreiz zur Aktion gibt, bestimmen im Modus der Peer Production gesellschaftliche Anreize und persönliche Motivationen

¹¹Ergebnisse wissenschaftlicher Forschung werden in der Regel in wissenschaftlichen Zeitschriften publiziert. Der Zugang zu diesen Zeitschriften ist mit nicht unerheblichen Kosten verbunden. Diese Art der Veröffentlichung von häufig auch mit öffentlichen Mitteln geförderter Forschung wird nicht nur wegen der Preise der Zeitschriften kritisiert. Durch die Verfügbarkeit des Internet erweisen sich auch die Publikationszyklen als äußerst lang. Nicht selten vergeht ein ganzes Jahr vom Einreichen eines Beitrags bis zur Verfügbarkeit des Abdrucks. Open Access ist eine Bewegung, die diesen Kritiken entgegentritt und wissenschaftliche Texte im Internet frei verfügbar macht (siehe auch Sietmann 2006).

die Art und den Umfang des Engagements der Akteure. Gerade Verträge und Eigentum an Informationen beschränken in firmen- und marktbasierter Informationsproduktion den Zugang von Menschen zueinander, zu Ressourcen und Projekten. Diese Beschränkungen fallen in der Peer Production. Die Akteure sind hier nicht vertraglich gebunden und auf Eigentum am Informationsgut wird in der Regel verzichtet (vgl. Benkler und Nissenbaum 2006, S.400). Die Leistungsfähigkeit von Peer Production Systemen wird also nicht nur durch die geringen Kommunikationskosten gestützt. Das Entfernen von Restriktionen, die durch Verträge und Eigentum an Informationen entstehen, senkt die Transaktionskosten noch weiter ab.

Mit Verzicht der Ausübung von Verwertungsrechten geistigen Eigentums, also dem Verzicht auf geistiges Eigentum an den durch Peer Production entstandenen Informationsgütern, schwindet auch der in Unterabschnitt 3.2 beschriebene Wohlfahrtsverlust. Dies ist einer der Gründe, weshalb sich Peer Production Systeme sehr gut zur Erstellung von Informationsgütern eignen (Benkler 2006, S.36). In Abwesenheit von Eigentum an Informationsgütern stellt sich ein kumulativer Effekt in der Peer Production von Informationsgütern ein. Neue Informationsgüter profitieren von vorhergehenden Informationsgütern, die zusammen mit menschlicher Energie und Kreativität den Input für Peer Production bilden und deren Output ein neues Informationsgut ist.

Mit Augenmerk auf diesen Verzicht an Exklusivität bezeichnet Benkler diese neue Art der Peer Production als Commons-based Peer Production. Er unterstreicht damit ihren Charakter, auf frei verfügbaren Informationsgütern aufzubauen und neue frei verfügbare Informationsgüter zu schaffen.

Die mittels Commons-based Peer Production erstellten Informationsgüter weisen drei strukturelle Eigenschaften auf (vgl. Benkler und Nissenbaum 2006, S.401f) :

Modularität bezeichnet die Zerlegbarkeit des Gutes in Module. Um mit Commons-based Peer Production Informationsgüter erstellen zu können, muss ein Weg gefunden werden, das Informationsgut in Komponenten zu zerlegen. Diese Komponenten können dann unabhängig, inkrementell und asynchron entwickelt werden.

Granularität bezeichnet die Größe der einzelnen Module. Kleine, feingranulare Module können die Zahl der Projektbeiträge erhöhen. So können auch potentielle Mitarbeiter, deren geringe Motivation nur kleine Aufwendungen an Zeit und Energie erlauben, die Möglichkeit erhalten einen Beitrag zu leisten.

Billige Integration verstärkt die Erfolgchancen von Commons-based Peer Production. Integration bezeichnet dabei den Mechanismus des Zusammensetzens der Module zum Endprodukt. Eine wichtige Eigenschaft der Integration sind die Kosten des gewählten Mechanismus. Er muss effektiv arbeiten und sollte möglichst geringe Aufwendungen von Arbeit und Zeit erfordern. Es muss dabei ein effektiver Weg der Qualitätskontrolle gefunden werden, um inkompetente oder bösartige Beiträge verwerfen zu können.

Modularität, Granularität und billige Integration sind Eigenschaften, die in verschiedenen Ausprägungen in allen Commons-based Peer Production Systemen eine Rolle

spielen. Im Abschnitt User Generated Content finden diese Begriffe ihre Anwendung im Vergleich verschiedener Arten von User Generated Content.

3.5 User Generated Content

User Generated Content Systeme sind eine Beispielgruppe von angewandter Peer Production. Der Begriff ist eng mit dem Web2.0 Umfeld verbunden. Web2.0 bezeichnet dabei ein Konglomerat verschiedener Web-Technologien, die dem Nutzer ein neues Internetgefühl vermitteln und das Web interaktiver machen. Durch diese Technologien werden Hemmnisse bei der Entwicklung und Nutzung kollaborativer System abgebaut. Die vereinfachte Entwicklung führt zu mehr Anwendungsreichtum, gesteigerter Funktionalität und einer kürzeren Entwicklungszeit für Anbieter von Web2.0 Plattformen. Der durch intuitive Gestaltung und Bedienung gesteigerte Nutzen der Anwendungen zieht Nutzerzahlen nach sich, die zuvor unerreichbar schienen. Die Technologien des Web2.0 senken die Transaktionskosten von Peer Production Systemen im User Generated Content Kontext und steigern somit ihre Leistungsfähigkeit.

Gerade in User Generated Content Systemen zeigt sich eine veränderte Konzeption des Verbrauchers. Statt einfach Konsument zu sein und Informationsgüter zu konsumieren, kann und wird der Verbraucher als User aktiv. Als User hat er die Wahl, zwischen den Rollen des Produzenten und des Konsumenten beliebig hin und her zu wechseln (vgl. Benkler 2006, S.138). Diese Wahlfreiheit eröffnet Usern Wege, sich kreativ und kritisch zu beteiligen. Die zuvor durch die Architektur¹² der Offline-Welt begrenzte Kreativität kann so ihren Weg zu einem viel größerem Publikum finden.

Professionelle Informationsproduktion und Massenmedien wird es weiterhin geben. Daneben hat sich durch User Generated Content, Web2.0 und internetbasierter Kommunikation ein neues Feld interaktiver Medien aufgetan. In diesem Bereich verändert sich nicht nur die Rolle des Verbrauchers. Auch dem Produzenten kommen neue Aufgaben zu. Statt Informationsgüter zu erstellen und zu verkaufen, können Produzenten im Web2.0 Umfeld nun Frameworks erstellen. Diese Frameworks sind der Rahmen, in dem Peer Production im User Generated Content Bereich stattfinden kann. Unter den erfolgreichen Anwendungen finden sich verschiedene Wirtschaftsmodelle. Das Spektrum reicht dabei von kostenlosen Diensten, deren Software frei verfügbar ist, bis hin zu kommerziellen Dienstleistungen gegen Gebühr, deren Code unter Verschluss gehalten wird.

Begriffsbildend für User Generated Content Systeme ist, dass in diesen Peer Production Systemen User Inhalte, die von ihnen oder anderen Usern konsumiert werden, selbst ins System bringen. User Generated Content Systeme leben von den Beiträgen der User. Dabei muss nicht einmal ein großer Teil der User in die Produzentenrolle wechseln, um solche Systeme erfolgreich zu machen. So werden z.B. im englischen Teil der Wikipedia¹³ über die Hälfte aller Änderungen und Erweiterungen von nur 0.7% der User durchgeführt. Die aktivsten 2% der User sind für ca. drei Viertel der Änderungen verantwortlich. Somit funktioniert die Wikipedia, die von Millionen Usern gelesen wird, durch aktives

¹²Architektur wird hier als eine Art der Regulierung von Verhalten nach Lessig (2006, S.123) verstanden und verwendet.

¹³siehe <http://www.wikipedia.org/>, geprüft 20.04.2007

Engagement einiger Weniger und gelegentliches Engagement Vieler (vgl. Sunstein 2006, S.152).

User Generated Content Systeme nutzen oft Größeneffekte. Durch die Zusammenarbeit und Beiträge vieler User entsteht ein Mehrwert, der den Wert der einzelnen Beiträge übersteigt. Wo läge denn der Wert einiger halbfertiger Artikel zur Zeitgeschichte? Erst die Sammlung und kontinuierliche Verbesserung schafft im Beispiel der Wikipedia eine Enzyklopädie. Ab einer bestimmten Größe können solche Systeme eine Eigendynamik entwickeln. Durch den Umfang, die Aktualität und den kostenlosen Zugang zur Wikipedia werden immer mehr Nutzer angezogen, von denen wiederum ein kleiner Teil Verbesserungen und Erweiterungen beitragen wird.

Es haben sich eine Reihe verschiedener Anwendungstypen entwickelt, die zum User Generated Content zählen. Ihre Eigenschaften weisen teilweise starke Unterschiede auf. Die im Unterabschnitt 3.4 erwähnten Dimensionen der Modularität, Granularität und der billigen Integration sollen verwendet werden um diese Systeme kurz darzustellen.

Dabei werden in dieser Reihenfolge Foren, Blogs, Wikis, Social Networking¹⁴ und spezialisierte Content Sharing Systeme betrachtet.

3.5.1 Foren

Foren sind ein Art Kommunikationmittel. In ihnen finden Diskussionen zu vielen verschiedenen Themen statt. Allerdings werden in einem Forum in der Regel Themen mit Bezug zu einem speziellem Sachgebiet diskutiert. In Foren findet man vorwiegend Texte, die in einigen Fällen um multimediale Inhalte wie Bilder, Videos oder Musik angereichert werden.

Die Modularität findet sich hier in der Teilbarkeit und Ordnung von Diskussionen und Diskussionsbeiträgen. Diskussionen werden dabei in Kategorien oder Unterforen verwaltet. Eine konkrete Diskussion besteht aus mehreren Posts oder Beiträgen, die sich aufeinander beziehen. Sie wird Thread genannt.

Granularität der Diskussion in Foren reicht von Ein-Wort-Beiträgen, bis zu Posts von mehreren Textseiten.

Billige Integration wird durch die Message Board Software erreicht, mit der Verwaltungsaufgaben in Foren vereinfacht wird. Es gibt Foren mit und auch ohne Qualitätskontrollen. Gute Ergebnisse bei der Qualitätskontrolle erreichen moderierte Foren. In ihnen werden Beiträge von Teilnehmern teilweise erst nach Prüfung durch einen Moderator für andere User sichtbar gemacht.

Die Vorteile von Foren liegen in ihrer leichten Bedienbarkeit und der inhärenten Konservierung von Informationen. In manchen Foren führt jedoch eine geringe Beteiligung der Forennutzer an der Diskussion zu Problemen. Häufig leben Foren von den Beiträgen weniger aktiver User. Gibt es keine neuen Beiträge und Diskussionen in einem Forum mehr, bezeichnet man es als ein totes Forum. Doch auch tote Foren haben einen Wert, da sie in ihnen konservierte Informationen weiter zur Verfügung stellen. Foreninformationen lassen sich z.B. über Suchmaschinen finden, so dass auch toten Foren immer wieder

¹⁴zu deutsch: soziale Kommunikationsnetzwerke

Leser bekommen.

3.5.2 Blogs

Blog ist ein Kunstwort das aus der Verkürzung des Begriffs Web-Log entstanden ist. Sie ähneln Foren und haben mit ihnen die Themenvielfalt gemein. Der Unterschied von Blogs und Foren liegt hauptsächlich in der Informationsorganisation. Während ein Forum zentral auf einem Webserver zu erreichen ist und Beiträge dort verwaltet werden, sind Blogs verteilter organisiert. Diskussionen können sich dabei durch Verlinkung und Referenzierung über mehrere Blogs erstrecken.

Module von Informationen sind wie bei Foren Textbeiträgen. Blogs sind eng mit Individuen oder Institutionen verbunden. Jedes Blog hat einen Eigentümer, der ähnlich einem Tagebuch Beiträge oder Posts schreibt. Diese Beiträge können von Lesern kommentiert werden.

Auch die Granularität von Blogs ähnelt der Granularität von Foren. Blog Posts reichen dabei von Einzeilern bis hin zu mehrseitigen Argumentationen mit Artikelcharakter.

Wieder wird billige Integration durch Software erreicht. Dabei vereinfacht die Blogsoftware vor allem das Schreiben von Einträgen und Kommentaren. Kommunikation und Diskussion wird in der dezentral organisierten Welt der Blogs durch Techniken wie Traceback oder Pingback unterstützt. Durch diese Techniken wird der Autor eines Blogbeitrags benachrichtigt, wenn sich ein anderer Blogger auf ihn bezieht. Qualitätskontrolle gibt es in der Blogwelt nicht auf direktem Weg. Für die Qualität eines Posts ist allein der Autor verantwortlich. Die Qualität eines Blogs oder Beitrags lässt aber indirekt in der Zahl der Referenzierungen und im Inhalt der Kommentare erkennen.

3.5.3 Wikis

Wikis¹⁵ unterscheiden sich stark von Foren und Blogs. In ihnen arbeiten User an Dokumenten. Dabei kann es sich z.B. um Dokumentationen oder Anleitungen handeln. Die Entwicklung dieser Dokumente findet inkrementell statt, wobei neue Beiträge auf vorhergehenden aufbauen. Wikis sind wie Foren und Blogs hauptsächlich textbasiert. Im Gegensatz zu ihnen tritt hier aber die Autorschaft der Beiträge in den Hintergrund. Sie lassen sich je nach Wikisystem schlecht oder gar nicht Autoren zuordnen.

Module in Wikis sind Webdokumente. Sie werden mittels einer einfachen Wikisprache programmiert. Teilweise lassen sie sich auch ähnlich der Erstellung eines Word-Dokuments durch What-you-see-is-what-you-get Oberflächen zusammenstellen und verändern.

Der Granularität von Beiträgen im Wiki sind kaum Grenzen gesetzt. Von der Änderung eines Satzzeichens bis zum Ändern oder Einstellen mehrseitiger Beiträge reicht das Spektrum.

¹⁵Durch Wikis werden Webseiten für User veränderbar und erstellbar. Sie erweitern die Möglichkeiten des Webs in den sie Beiträge von Usern erlauben. Der Name Wiki wurde von Ward Cunningham geprägt. Er ist eine Abkürzung des hawaiianschen Begriffs Wiki Wiki und bedeutet übersetzt schnell (Cunningham 2005). Wikis erlauben eine schnelles Erstellen und Verändern von Webdokumenten.

Billige Integration wird durch Wikisoftware erreicht. Anders als bei Blogs und Foren arbeiten Wikis nicht nachrichtenbasiert. Stattdessen werden Änderungen, Erweiterungen und auch Löschungen an Web-Dokumenten vorgenommen und von der Wikisoftware koordiniert. Qualitätskontrolle kann von beliebigen Usern durchgeführt werden. Es gibt keinen expliziten Mechanismus zur Kontrolle und Freigabe von Inhalten. Jede Änderung ist sofort sichtbar. Wenn Beiträge von schlechter Qualität oder boshafter Absicht entdeckt werden, kann sie der Entdecker umgehend verbessern. Dabei führt der schnellste Weg zur Korrektur über die in der Wikisoftware eingebaute Versionskontrolle. Alte Versionen des betroffenen Dokuments lassen sich per Klick reaktivieren.

3.5.4 Social Networking

In dieser Kategorie steht das Individuum als soziales Wesen mit persönlichen Interessen und Beziehungen zu Mitmenschen im Mittelpunkt der Anwendungen. Unter den Begriff Social Networking fallen alle Dienste, deren Mittelpunkt das Knüpfen und Pflegen von Kontakten ist. Ein paar bekannte Beispielanwendungen dieser Kategorie sind :

- Xing/OpenBC¹⁶: Social Networking im Geschäftsumfeld
- StayFriends¹⁷: Social Networking für ehemalige Schüler
- MySpace¹⁸: Allgemeine Social Networking Anwendung
- StudiVZ¹⁹: Social Networking für Studenten
- Dating-Dienste

Das hier entstehende Informationsgut ist eine Art Verzeichnis persönlicher Daten und Beziehungen. Teilweise werden sogar Vorlieben und Einstellungen ins Verzeichnis aufgenommen. Die Struktur der Zusammenarbeit ähnelt dabei den Konzepten, die auch im Domain Name System (DNS) zum Einsatz kommen. Im DNS wird die Autorität einer bestimmten Zone dem Eigentümer des entsprechenden Teiles des Internets zugewiesen. Er allein ist für seine Daten, im Fall DNS also die Zuweisung von IP-Adressen zu Rechnernamen, verantwortlich. In Fall von Social Networking Diensten wird jedem User eine eigene, persönliche Zone (oder Bereich) zugewiesen. Der User ist für seinen Beitrag zum Verzeichnis, d.h. seine persönlichen Daten selbst verantwortlich. Hier aber hören die Gemeinsamkeiten zwischen DNS und Social Networking auf. Ein bedeutender Unterschied ist die physische Organisation der Daten. DNS arbeitet dezentral, dabei werden die Zonendaten in der Regel in physischer Nähe zur Autorität der entsprechenden Zone abgelegt. Social Networking Dienste speichern Daten in der Regel zentral beim Anbieter des Dienstes.

¹⁶<http://www.xing.com/> geprüft 28.03.2007

¹⁷<http://www.stayfriends.de/> geprüft 28.03.2007

¹⁸<http://www.myspace.com/> geprüft 28.03.2007

¹⁹<http://www.studivz.net/> geprüft 28.03.2007

Die in diesen Verzeichnissen gesammelten Daten sind User Generated Content, da sie von den Nutzern selbst eingestellt werden. Der Wert dieser Daten im Bezug auf die in Abschnitt 2 beschriebenen Methoden zur Markt-, Trend- und Zukunftsforschung muss allerdings von Anwendung zu Anwendung einzeln betrachtet werden. Besonders wertvoll können diese Verzeichnisse sein, wenn sich Daten zu Vorlieben und Einstellungen ablesen lassen.

Anwendungen dieser Kategorie dienen hauptsächlich der Organisation und Pflege sozialer Beziehungen und Kontakte. Dazu werden von Betreibern solcher Plattformen dem Thema entsprechende Anwendungen zur Verfügung gestellt. Zusätzlich zum einem Bereich, den der User selbst gestalten kann werden häufig Kommunikations- und Suchdienste angeboten. Die Anwendungen werden in der Regel mit weiteren Diensten angereichert, Beispiele dafür sind eingebundene Blogs, Foren und Content Sharing Dienste.

Die Modularität der Informationen in diesen Anwendungen wird durch das verwendete Framework und die möglichen Inhalte vorgegeben. Module enthalten persönliche Daten und Kontakte.

Auf dieser allgemeinen Ebene lassen sich kaum brauchbare Aussagen zu Granularität und billiger Integration treffen. Um diese Punkte zu klären, müssten Beispielanwendungen einzeln betrachtet werden.

3.5.5 Spezialisierte Content Sharing Systeme

Es sind in den letzten Jahren eine Reihe von Plattformen entstanden, in deren Mittelpunkt jeweils ein spezieller Typ digitalen Inhalts steht. Produzenten stellen für Anwendungen dieser Kategorie spezialisierte Frameworks zur Verfügung, um den Eigenschaften des gewählten Inhalts gerecht zu werden.

Zu den Beispielanwendungen dieser Kategorie gehören :

- YouTube²⁰ als Videosharing Dienst
- Flickr²¹ als Fotosharing Dienst
- Delicious²² als Social Bookmarking Dienst

Modularität wird in diesen Diensten vom Typ des gewählten Inhalts bestimmt. Die Modulgröße ist dabei durch den Inhalt klar definiert. Inhalte dieser Systeme lassen sich Usern zuordnen. Jeder User ist dabei für seine Inhalte selbst verantwortlich. Eine echte Zusammenarbeit von Usern an gemeinsamen Inhalten findet nicht statt.

Die Granularität der Beiträge besteht aus einzelnen Einheiten des gewählten Typs. Dabei handelt es sich z.B. um ein Bild, ein Video oder ein Lesezeichen, das in der Regel öffentlich abgespeichert wird. In der Regel gibt es keine Beschränkungen wie viele Inhalte von einzelnen Usern gespeichert werden können. So können große Datenbanken oder Sammlungen entstehen.

²⁰<http://www.youtube.com/> geprüft 28.03.2007

²¹<http://www.flickr.com/> geprüft 28.03.2007

²²<http://del.icio.us/> geprüft 28.03.2007

Billige Integration kann hier nur als einfaches Hinzufügen von Inhalten zur Sammlung oder Datenbank verstanden werden. Das durch diese Art der Peer Production erzeugte Informationsgut ist die entstehende Datenbank oder Sammlung von Inhalten. Es gibt keinen offensichtlichen Mechanismus der Qualitätskontrolle in diesen Systemen.

3.6 Zusammenfassung

Damit schließt sich der Abschnitt zu Informationen und Gesellschaft. Er begann mit dem Zusammenhang von wirtschaftlichem Wachstum und Wohlfahrt. Dabei wurde die volkswirtschaftliche Definition ökonomischer Wohlfahrt vorgestellt. Im Unterabschnitt Informationsökonomie wurden die wirtschaftlichen Eigenschaften von Informationsgütern vorgestellt und der Wohlfahrtsverlust aus der Anwendung des Preissystems auf Informationsgüter aufgezeigt. Der Unterabschnitt zur Dezentralisierung stellte Kostenargumente zu Organisationsformen vor. Im Unterabschnitt Peer Production wurde die Dezentralisierung von Organisationsformen zur Produktion gezeigt. Dort wurden wichtige Eigenschaftsdimensionen der in Peer Production erstellten Informationsgüter gezeigt. Modularität, Granularität und billige Integration fanden im Unterabschnitt User Generated Content ihre Anwendung im Vergleich verschiedener Arten von User Generated Content.

Es folgen die Kernaussagen dieses Abschnitts :

- Informationen sind von Natur aus öffentliche Güter.
- Das Preissystem führt für Informationsgüter zu ineffizienten Innovationsanreizen und Wohlfahrtsverlust.
- Gesunkene Transaktions- und Kommunikationskosten ermöglichen neue dezentrale Organisationsformen mit Größenvorteilen und Freiheitengewinn.
- Mit Commons-based Peer Production von Informationsgütern steigt die Wohlfahrt.
- Modularität, Granularität, und billige Integration sind wichtige Eigenschaftsdimensionen von Informationsgütern der Peer Production.
- User Generated Content ist ein Beispiel für Peer Production.
- Blogs sind dezentralisierte Foren, in ihnen finden sich Freitextbeiträge in Nachrichtenform.
- Wikis sind dezentral entstehende veränderliche Webseiten.
- In Social Networking Diensten entstehen Verzeichnisse persönlicher Daten.
- In spezialisierten Content Sharing Diensten sind die Inhalte klar strukturiert.

Im nächsten Abschnitt werden die hier gezeigten Eigenschaften verschiedener User Generated Content Arten und die in Unterabschnitt 2.5 genannten Anforderungen an

System und Daten genutzt, um einen Kandidaten zur Unterstützung der in Abschnitt 2 beschriebenen Methoden zur Markt-, Trend- und Zukunftsforschung zu finden. Eine detaillierte Beschreibung der Eigenschaften der gewählten User Generated Content Art ist der konkreten Wahl der auszuwertenden Anwendung vorrangestellt. Im Zuge der Vorstellung der gewählten Anwendung wird besonders auf die Struktur ihres Inhalts eingegangen. Der nächste Abschnitt klärt die Grundlagen für den sich anschließenden Abschnitt, in dem Aufbau und Funktionsweise eines Auswertungssystems beschrieben wird.

4 Delicious und Social Bookmarking

Um die in Abschnitt 2 beschriebenen Methoden zur Trend-, Markt- und Zukunftsfor- schung zu unterstützen, können Daten aus User Generated Content Systemen verwendet werden. Aufgrund der Datenmenge sollen automatisierte Methoden zum Einsatz kommen. Aus den im Unterabschnitt 2.5 aufgezählten Anforderungen an das System und die Daten lassen sich folgende Anforderungen an die Daten extrahieren :

Klare Struktur vereinfacht die Entwicklung und Arbeitsweise der automatischen Auswertung.

Themenvielfalt macht Trendsuche zu vielen Themen möglich.

Aktualität ermöglicht das frühzeitige Erkennen von Trends.

Trendführende Daten sind notwendig, um aussagekräftige Ergebnisse zu erzielen.

Zusammen mit der Vorstellung der User Generated Content Arten im Abschnitt 3.5 können diese Anforderungen als Leitfaden dienen, geeignete User Generated Content Anwendungen zu identifizieren. Zuerst werde ich die User Generated Content Arten auf klare Strukturen und Themenvielfalt prüfen. Danach kann über konkrete Kandidaten nachgedacht werden, die die verbleibenden Anforderungen nach Aktualität und trend- führenden Daten erfüllen können.

In Foren und Blogs finden sich Texte, konkrete Instanzen sind häufig auf spezielle Themen ausgerichtet. Es existieren zwar inzwischen Ansätze zur Auswertung und Kategorisierung von Texten, die Textanalyse ist aber sehr aufwendig. Daten aus Blogs und Foren sind nicht besonders gut strukturiert und weisen bei Betrachtung einzelner Instanzen eine nur geringe Themenvielfalt auf.

In Wikis finden sich Webdokumente. Hier ließen sich zwar Elemente wie Überschriften, Tabellen und Freitext leicht voneinander unterscheiden, eine sinnvolle automati- sierte Verarbeitung dieses Inhalts steht aber vor ähnlichen Schwierigkeiten wie im Fall von Foren und Blogs. Es müssten wiederum Texterkennungs- und Klassifizie- rungsalgorithmen zu Einsatz kommen. Abgesehen von der Wikipedia sind konkrete Wikis in ihrer thematischen Vielfalt begrenzt. Wikis sind somit keine idealen Kan- didaten zur automatisierten Auswertung.

Im Social Networking finden sich verschiedenste Inhalte. Auch hier gibt es Freitext, oft in Nachrichten oder im Bereich der Selbstdarstellung. Social Networking Anwen- dungen können ein Kandidat für automatische Methoden sein. Die Themenvielfalt in den konkreten Anwendungen ist von der Anwendung abhängig. Durch ihre Ver- zeichnisartigkeit finden sich teilweise klare Strukturen, die sich zur automatisierten Erfassung und Verarbeitung eignen.

In spezialisierten Content Sharing Systemen finden sich durch die Spezialisierung auf eine Art des Inhalts häufig sehr klare und einfache Strukturen. Die durch die Inhalte abgedeckten Themen können, genügend User vorausgesetzt, eine große Vielfalt aufweisen. Damit sind spezialisierte Content Sharing Dienste die Art von User Generated Content Systemen, unter denen sich nach Kandidaten zur Auswertung mittels automatischer Methoden zu suchen lohnt.

Die Kandidatenmenge kann weiter eingeschränkt werden, indem konkrete Inhalte spezialisierter Content Sharing Systeme betrachtet werden. Die im Abschnitt 3.5.5 gegebenen Beispiele umfassten Filme, Bilder und Lesezeichen. Zu allen diesen Kandidaten ist eine Auswertung vorstellbar. Zwar sind Filme und Bilder ähnlich der Analyse von Freitext schlecht analysierbar, Abhilfe schafft aber die in diesen Diensten gebräuchliche Verschlagwortung, die auch Tagging genannt wird. Lesezeichen zeichnen sich durch eine besonders klare Struktur aus. Besonders interessant sind sie, da sich aus ihnen ablesen lässt, womit sich User aktiv beschäftigen. Damit können Social Bookmarking Dienste aktuelle trendführende Daten enthalten.

Die klare Struktur, Themenvielfalt und die Aussicht auf aktuelle trendführende Daten machen Social Bookmarking zum idealen Kandidaten für eine automatisierte Auswertung mit dem Ziel, Markt-, Trend- und Zukunftsforschung zu unterstützen.

Im folgenden Abschnitt wird Social Bookmarking genauer vorgestellt. Dabei werden Funktionsweise, Motivationen zur Nutzung und der Mehrwert, den dieser Dienst seinen Usern bietet, beleuchtet.

4.1 Social Bookmarking

Im vorhergehenden Abschnitt wurde Social Bookmarking als Kandidat für automatisierte Methoden zur Unterstützung von Markt-, Trend- und Zukunftsforschung erkannt und ausgewählt. Dieser Abschnitt klärt die Fragen :

- Was ist Social Bookmarking?
- Welchen Nutzen bringt Social Bookmarking?

Social Bookmarking ist ein weiteres Beispiel für einen Dienst aus dem Web2.0 Umfeld. Bookmarks oder Lesezeichen werden klassischer Weise im Browser verwaltet und auf dem Rechner des Users abgelegt. Dieses Verfahren kann Schwierigkeiten mit sich bringen. Zum einen arbeiten User immer häufiger an verschiedenen Rechnern und zum anderen gehen Rechner kaputt. Im ersten Fall fehlen dem Anwender auf anderen Rechnern seine bekannten Lesezeichen. Der zweite Fall ist tragischer, hierbei können Anwender Lesezeichen verlieren.

Aus beiden Gründen ist eine zentrale Speicherung von Lesezeichen einer peripheren Speicherung auf den Rechnern der Internetnutzer vorzuziehen. Man kann davon ausgehen, dass der Rechner eines Users fehleranfälliger als ein Server ist. Server sind durch redundante Datenspeicherung, Backup und Failover in der Regel besser abgesichert. Da

Lesezeichen ohne eine bestehende Internetverbindung wenig Sinn machen, bietet sich die zentrale Verwaltung von Lesezeichen via Internet Service an.

Klassischer Weise werden Lesezeichen in Ordnern gespeichert, die ein Wiederfinden von Lesezeichen erleichtern sollen. Steigt die Zahl der Lesezeichen über ein bestimmtes Maß, dann wird Verwaltung, Pflegen und Ordnen der Lesezeichen immer aufwendiger. Nicht nur Benennen und Sortieren dieser Ordner, auch Zuordnen von Lesezeichen zu Ordnern ist aufwendig. Dabei ist Einordnen von Lesezeichen in eine hierarchische Struktur teilweise schwierig. In einigen Fällen lässt sich schwer entscheiden, in welche Kategorie ein Lesezeichen gehört.

Angenommen, ich habe auf der Seite eines Sushi Restaurants eine Beschreibung von California Rolls und Hinweise zu deren Zubereitung gefunden. Da ich gerne Sushi esse, möchte ich den Link speichern. Noch weiß ich nicht, ob ich die California Rolls selbst zubereiten werde, oder sie doch lieber im Restaurant probieren soll. Vielleicht werde ich ja auch beides tun. Nun stellt sich die Frage : Wo speichere ich den Link ab? In welchen Ordner gehört er? Angenommen ich habe die beiden Ordner Restaurants und Rezepte schon in meiner Hierarchie. Den Link in beide Ordner zu speichern wäre redundant, ich hätte dann zwei separate Kopien, die auch separat gepflegt werden müssten. Nur in einen Ordner zu speichern ist auch ineffektiv, da es dem Inhalt der Webseite nach meiner persönlichen Bewertung des Links nicht gerecht wird.

Werden bei der Ablage der Lesezeichen Kompromisse gemacht, gibt es einen Informationsverlust. Eine Möglichkeit, Lesezeichen in mehreren Kategorien gleichzeitig abzulegen, ohne Redundanzen zu erzeugen, ist die Verschlagwortung. Durch Schlagworte oder Tags wird eine schwache Ontologie erzeugt. Damit verzichtet man auf die baumartige hierarchische Struktur der Ordner und verwendet stattdessen eine flache Struktur. Die Verschlagwortung, oder das Tagging, vereinfacht und beschleunigt Ablegen und Auffinden von Lesezeichen. Im Abschnitt 4.2 wird Tagging ausführlicher betrachtet.

Die zentrale Speicherung von Lesezeichen mit einem Internetservice ist nicht automatisch als Social Bookmarking zu bezeichnen. Sie wird sozial genannt, wenn User ihre Bookmarks öffentlich, für andere zugänglich speichern. Welchen Anreiz könnten Anwender haben, ihre Lesezeichen öffentlich zu speichern?

Die Entscheidung, seine Lesezeichen in einen Social Bookmarking System zu speichern, kann von verschiedenen Anreizen motiviert sein. Zum einen sind dabei die oben beschriebenen Vorteile einer zentralen Speicherung zu nennen. Auch der vereinfachte Austausch von Lesezeichen mit Freunden oder Kollegen kann eine Motivation zur Veröffentlichung sein. Gerade dieser Austausch durch Veröffentlichung ist der Akt, der zentralem Bookmarking die soziale Komponente einbringt. Durch die öffentliche Speicherung von Lesezeichen entsteht eine Datenbank mit aktuellen und wertvollen Informationen. Mit geeigneten Mechanismen kann diese Datenbank dann von Usern als Internet-Suchmaschine verwendet werden. Voraussetzung für Qualität und Umfang der Ergebnisse sind dabei genügend viele und aktuelle Bookmarks in der Datenbank. Oder anders gesagt, es muss eine kritische Menge aktiver User erreicht werden, um gute Suchdienste auf Basis von Social Bookmarking Daten leisten zu können. Dieser Mehrwert kann eine weitere Motivation sein, sich sozial zu verhalten und eigene Bookmarks öffentlich zu speichern.

| Rang | Social Bookmarking Dienst | 3m avg Daily Reach |
|------|---------------------------|--------------------|
| 1. | Delicious ^a | 0.4695% |
| 2. | Mister Wong ^b | 0.0435% |
| 3. | Furl ^c | 0.03345% |
| 4. | Magnolia ^d | 0.01735% |
| 5. | Spurl ^e | 0.01015% |

Tabelle 4: Social Bookmarking nach Daily Reach geordnet

^a<http://del.icio.us/> geprüft 28.03.2007

^b<http://www.mister-wong.de/> geprüft 04.04.2007

^c<http://furl.net/> geprüft 30.03.2007

^d<http://ma.gnolia.com/> geprüft 30.03.2007

^e<http://spurl.net/> geprüft 30.03.2007

Nun muss eine konkrete Social Bookmarking Anwendung gewählt werden. Die Nützlichkeit der Daten wird maximiert durch große Datenmengen und thematische Vielfalt. Die Anwendungen aus Tabelle 4 sind Beispiele für bekannte Social Bookmarking Dienste.

Sie sind nach der Anzahl der Nutzer geordnet. Daily Reach ist dabei ein Maß von Alexa.²³ Die im Ranking verwendeten Zahlen wurden am 30.03.2007 ausgelesen. Die verwendeten Werte sind Daily Reach Durchschnittswerte der vorhergehenden drei Monate.

Um von maximalen Userzahlen und Themenvielfalt zu profitieren, wird das Social Bookmarking System Delicious gewählt. Doch bevor genauer auf Delicious und die Struktur der über diesen Dienst verfügbaren Daten eingegangen wird, muss noch ein wichtiger Mechanismus erklärt werden, der in vielen Web2.0 Anwendungen zum Einsatz kommt. Der nächste Abschnitt widmet sich der Verschlagwortung, dem Tagging von Inhalten.

4.2 Tagging

Jede Art von Informationssammlung benötigt ein System der Ordnung, eine Klassifizierung. Klassische Ordnungssysteme sind häufig hierarchisch aufgebaut. Diese Ordnung hat ihre Einschränkung in der Zuordnungsfähigkeit von Objekten zu Kategorien. Für jede denkbare Kategorie gibt es sowohl Prototypen als auch Grenzfälle. Durch die in solchen Systemen häufig vorzufindende Regel, Objekte nur an eine Position einzuordnen, können Probleme entstehen. Während die Ablage oder Einordnung prototypischer Objekte trivial ist, fallen für die Handhabung von Grenzfallobjekten Mehrkosten an. Entweder werden sie an einer der für sie möglichen Positionen eingeordnet oder es muss eine neue Kategorie für sie geschaffen werden. Das Schaffen von neuen Kategorien oder die im schlimmsten Fall notwendige (zumindest partielle) Reorganisation der Klassifizierung bringen erheblichen Aufwand mit sich.

Im Fall von Grenzobjekten entstehen auch dem Suchenden zusätzliche Kosten, um an

²³<http://www.alexa.com/>, geprüft 30.03.2007.

das gewünschte Objekt zu gelangen. Sein Mehraufwand entsteht durch die Notwendigkeit, an verschiedenen Orten zu suchen zu müssen.

Durch die gesunkenen Transaktions- oder Kommunikationskosten ergeben sich neue Möglichkeiten zur Klassifizierung von Objekten und Informationsgütern. Gerade die User Generated Content Systeme erfordern aufgrund der Unmengen verwalteter Objekte einen Ansatz, der den Aufwand der klassischen Klassifizierung auf die User verteilt.

Die klassische Klassifizierung oder Katalogisierung wird in der Regel von Experten vorgenommen. Will man diese Aufgabe den Usern der User Generated Content Systeme antragen, ergeben sich zwei Probleme. Zum einen fehlt vielen Usern die nötige Expertise, den Ansprüchen der klassischen Klassifizierung zu genügen²⁴. Zum anderen sind von Experten erstellte Kategorien häufig wenig intuitiv (vgl. Golder und Huberman 2006, S.200).

Diesen Problemen kann begegnet werden, indem auf den Anspruch der akkuraten Klassifizierung verzichtet wird. Eine Klassifizierung oder Einordnung muss z.B. nicht den Anforderungen aller User an Auffindbarkeit gerecht werden. Auch die Regel, ein Objekt nur an einer Position einzuordnen, kann fallen gelassen werden. Einfache, intuitive Klassifizierung und Ordnung von Objekten kann mit dem System des Tagging, einer Verschlagwortung von Objekten oder Inhalten erreicht werden.

Wenn von Tagging gesprochen wird, handelt es sich um einen im Kern sehr einfachen Mechanismus. Tagging ist die Markierung von Objekten mit frei wählbaren Schlagwörtern. Diese Schlagwörter werden Tags genannt, ihre Wahl unterliegt keinen Beschränkungen. Tags sollten aber sinnvoll gewählt werden, so dass sie einen beliebigen, relevanten Bezug zum getaggten Inhalt aufweisen. Sinnvolle Tags sind solche, die dem Tagnutzer das Auffinden der klassifizierten Objekte erleichtern. Durch Tagging entsteht in der Regel keine Hierarchie, stattdessen wird ein flacher Namensraum gebildet, in dem alle Schlagwörter strukturell gleichwertig sind.

Obwohl der grundlegende Mechanismus des Tagging sehr einfach ist und in allen Tagging Anwendungen benutzt wird, gibt es Unterschiede in den konkreten Umsetzungen. Es lassen sich Eigenschaften feststellen, die in konkreten Tagging-Systemen verschieden ausgeprägt sind. Ziel meiner Ausführungen ist es jedoch nicht, eine Taxonomie von Tagging Systemen aufzustellen, siehe dazu Marlow et al. (2006). Um eine kurze Einführung, Übersicht und Einordnung zu vermitteln genügt hier die Unterscheidung von Taggingssystemen anhand der Frage: Wer taggt für wen?

In Abbildung 14 werden verschiedenen Taggingssysteme danach geordnet, wer für wen die Schlagwörter vergibt. In den gezeigten Beispielanwendungen unterscheiden sich also die Motive der Tagger. Folgend werden diese Motivationen genauer betrachtet.

Erzeuger taggen für sich selbst

Ein gutes Beispiel für diese Art des Tagging ist der Photosharing Dienst Flickr. Hier taggen Nutzer ausschließlich ihre eigenen Inhalte. Hauptzweck des Taggings

²⁴Man denke zur Verdeutlichung der Problematik an das Beispiel von der Taxonomie der Lebewesen wie sie von Biologen betrieben wird (vgl. Golder und Huberman 2006, S.200).

²⁵<http://www.technorati.com/>, geprüft 02.04.2007.

| | | | |
|------------|--------|--------------------------|---|
| Tag-Nutzer | Andere | Technorati ²⁵ | (Wikipedia) |
| | Selbst | Flickr | Furl Spurl Mister Wong Delicious |
| | | Selbst | Andere |
| | | Tag-Erzeuger | |

Abbildung 14: Tagging Motivation und Beispiele nach Hammond et al. (2005)

ist dabei die Organisation eigener Fotos. Das primäre Ziel dieser Ordnung ist es, später selbst in der Lage zu sein relevante Fotos schnell wiederzufinden. Die verwendeten Tags basieren zwar auf der Inhaltseinschätzung des Erzeugers, trotzdem entsteht durch Verwendung einer gemeinsamen Basis zur Wahl der Schlagwörter ein Mehrwert für andere User. Jeder User ist somit in der Lage z.B. Bilder des letzten Metallica Konzerts in Berlin zu finden. Das kann erreicht werden, indem nach Bildern mit den Tags Metallica, Waldbühne und eventuell Berlin gesucht wird.

Erzeuger taggen für Andere

In diesem Fall taggen die Erzeuger von Inhalten ihren Inhalt, um anderen Nutzern das Auffinden von Inhalten zu erleichtern. Ein wichtiges Beispiel für diese Art des Taggings zeigt sich in der Web2.0 Anwendung Technorati. Hier werden Blogs und Blogposts ausgewertet. Da die Blogposts in der Regel von ihren Autoren mit Tags versehen werden, ist Technorati nach Indizierung vieler Blogposts²⁶ in der Lage, seinen Usern einen Suchdienst zur Verfügung zu stellen. Dabei können die Schlagwörter als Suchbegriff verwendet werden.

Dieses Vorgehen ähnelt stark der Konzeption des Semantic Web (Shadbolt et al. 2006) bei dem auch die Erzeuger, in diesem Fall von Webseiten, Schlagwörter für ihre Inhalte vergeben.

Nicht-Erzeuger taggen für sich selbst

Dieser Fall des Tagging findet sich in vielen spezialisierten Content Sharing Sy-

²⁶Technorati aggregiert mehr als 71 Millionen Blogs (Stand 04.04.2007, vgl. Technorati).

stem. Hier entsteht durch die gesammelten Klassifizierungen von Objekten ein System, in dem die Wahrscheinlichkeit Objekte zu finden, verglichen mit anderen Tagging-Arten, hoch ist. Nutzer speichern Referenzen auf Objekte, die aufgrund eigener Bewertungen getaggt werden. Durch mehrfache Speicherung und Verschlagwortung von Objekten kann eine Schlagwortwolke, oder Tag Cloud, für das Objekt entstehen. Diese Tag Cloud enthält die nach den Userbewertungen wichtigen Schlagwörter für das entsprechende Objekt. Häufig verwendete Schlagwörter werden in der der Tag Cloud durch größere Schrift gekennzeichnet. Die Tag Cloud entspricht den aggregierten Bewertungen des Inhalts von Ressourcen. Ein Beispiel einer Tag Cloud ist in Abbildung 15 zu sehen, dabei handelt es sich um die Menge populärer Tags von Delicious, die über alle Delicious Bookmarks berechnet wird.

Nicht-Erzeuger taggen für Andere

Der Fall, in dem jemand, der nicht Erzeuger eines Inhalts ist, diesen zum Nutzen Anderer taggt, ähnelt dem Vorgehen von Bibliothekaren und generell der klassischen Klassifizierung. Ein Beispiel aus dem Web2.0 Umfeld ist das Kategoriensystem der Wikipedia (Voss 2006). In der Wikipedia können Artikel durch Tags zu Kategorien zugeordnet werden. Die hier verwendeten Tags sind gleichzeitig Kategorien. Eine Besonderheit des Kategoriensystems der Wikipedia ist die Möglichkeit, durch diese spezielle Art des Taggings hierarchische Strukturen zu erzeugen. Das geschieht durch die Zuordnung von Unterkategorien zu Überkategorien. Die Wikipedia Art des Taggings ist dabei viel stärker reguliert als andere Anwendungen. Es gibt eine ganze Reihe von Regeln, die beim Tagging oder Klassifizieren von Artikeln eingehalten werden müssen (vgl. Wikipedia 2007).

Tagging ist nicht als die Lösung aller Probleme zu verstehen, die in Ordnungs- oder Klassifikationssystemen auftreten. So weist Marlow et al. (2006) z.B. auf das Vokabularproblem hin. Im Kern geht es dabei darum, dass User für gleiche Dinge verschiedene Begriffe benutzen um sie zu beschreiben. Auch Golder und Huberman (2006, S.200) führen ähnliche Probleme an, die durch Synonyme, Pluralformen oder unterschiedliche Schreibweisen von Tags auftreten. Diese Probleme können zu Ineffizienzen bei der Suche nach getaggten Ressourcen führen, so dass User eventuell mehrere Suchanfragen oder eine Suche mit komplexeren Suchbegriffen durchführen müssen, um viele, oder alle relevanten Objekte zu finden. Trotz dieser Probleme können Tagging-Systeme gut funktionieren. Cameron Marlow beschreibt die Grundlagen zur Funktionsfähigkeit von Tagging Systemen so :

...social tagging systems rely on shared and emergent social structures and behaviors, as well as related conceptual and linguistic structures of the user community.

Marlow et al. (2006)

Mit allgemein bekannten Konzepten und sprachlichen Strukturen kann eine Ordnung von Ressourcen oder Objekten verteilt entstehen, von der alle User profitieren können. Diese Ordnung, die durch Tags beschrieben wird, kann man als Informationsgut verstehen, welches durch den Peer Production Prozess des Tagging erzeugt wird.



Abbildung 15: Beispiel einer Tag Cloud: Populäre Delicious Tags als vom 12.03.2007
Quelle: Delicious (2007)

Zusammenfassung

Tagging ist ein Mechanismus mit dem sich Informationen verteilt klassifizieren lassen. Die entstehende Ordnung oder Klassifikation ist das Informationsgut. Durch dezentrale Peer Produktion werden Anbieter von Content Sharing Systemen vom Aufwand, Objekte oder Inhalte selbst zu klassifizieren oder zu ordnen, befreit.

Im nächsten Abschnitt wird der Social Bookmarking Dienst Delicious betrachtet. Dabei werden sowohl die Struktur der in Delicious gespeicherten Daten, als auch Zugangsmöglichkeiten vorgestellt. Dieser folgende Abschnitt ist Grundlage für ein System zur automatisierten Unterstützung der Methoden aus dem Abschnitt 2 zur Markt-, Trend- und Zukunftsforschung.

4.3 Delicious

Im Abschnitt 4.1 wurde Delicious als idealer Kandidat zur Unterstützung der im Abschnitt 2 beschriebenen Methoden der Markt-, Trend- und Zukunftsforschung ausgewählt. In diesem Abschnitt wird die Social Bookmarking Anwendung Delicious vorgestellt. Dabei werden insbesondere die Struktur der in Delicious gespeicherten Daten und mögliche Zugangsarten zu diesen Daten beschrieben.

Im Abschnitt 4.1 wurde Social Bookmarking als eine User Generated Content Anwendung vorgestellt. Die Tatsache von inzwischen mehr als zwei Millionen registrierten Delicious Usern (vgl. Schachter 2007b) macht die Auswertung dieser Social Bookmarking Anwendung besonders interessant.



Abbildung 16: Motivationen zur Nutzung von Delicious. (<http://del.icio.us/> abgerufen am 05.04.2007)

4.3.1 Zur Entstehung

Joshua Schachter begann Anfang 2003 mit der Entwicklung eines Systems zur Verwaltung von Web-Lesezeichen. Die Idee zu diesem System entstand aus der Notwendigkeit, sehr viele Bookmarks zu verwalten (vgl. Katz 2006). Dabei entwickelte sich Schachters System von einem Werkzeug, in dem nur er Lesezeichen speichern und taggen konnte, zu einem System, das diesen Dienst jedem frei zur Verfügung stellt. Schachters System trägt den Namen der URL, unter der es erreichbar ist, del.icio.us oder Delicious. Zu Beginn war die Entwicklung von Delicious eine Freizeitbeschäftigung Schachters. Das enorme Wachstum der Userzahlen und der Erfolg seines System machte die Weiterentwicklung und Pflege des Dienstes immer aufwendiger. Anfang 2005 entschied Joshua Schachter sich ausschließlich mit der Weiterentwicklung von Delicious zu beschäftigen (vgl. upian Hot Links 2007). Er kündigte seinen regulären Job. Das Projekt Delicious entwickelte sich weiterhin so erfolgreich, dass Schachter gegen Ende des gleichen Jahres Delicious an Yahoo verkaufen konnte. Zu dieser Zeit hatte Delicious bereits über 300.000 registrierte Nutzer, der Kaufpreis wurde nicht bekannt gegeben (vgl. Green 2005).

4.3.2 Motivation

Die Motivationen für User, ihre Bookmarks zentral und öffentlich auf Delicious zu speichern, finden sich auch direkt auf der Delicious Startseite. Der entsprechende Ausschnitt ist in Abbildung 16 abgebildet.

Die drei in Abbildung 16 dargestellten Punkte sind Usermotivationen. Der dritte Punkt, das Recherchieren, welche Bookmarks andere User speichern, fällt mit der Zielsetzung dieser Arbeit zusammen.

Bevor im Abschnitt 5 ein System zur Auswertung von Delicious Daten und Unterstützung der Methoden von Markt-, Trend und Zukunftsforschung vorgestellt werden kann, soll gezeigt werden wie Bookmarks in Delicious aussehen und welche Möglichkeiten De-

```
link: http://ilari.scheinin.fidisk.fi/itunemywalkman/  
tags: itunes mac mobile music  
user: scotttag  
time: 2006-05-31 01:43:30  
description: An example how to tune a walkman on mac.  
title: iTuneMyWalkman
```

Abbildung 17: Ein Beispieldatensatz für Bookmarks in Delicious

licious bietet um diese drei gezeigten Motivationen zu bedienen.

Ein Bookmark oder Lesezeichen besteht in Delicious aus einem Link, den Tags, einem Usernamen, einem Zeitstempel und dem Titel der Webseite. In Abbildung 17 wird diese Struktur mit einem Beispiel verdeutlicht.

Der 'link' ist dabei der Uniform Ressource Locator (URL) der Webseite. Unter 'tags' finden sich die vom User gewählten Schlagworte, die einen für den User relevanten Bezug zum Inhalt der Webseite haben sollte. Es wird der 'name' des Users vermerkt, der dieses Lesezeichen gespeichert hat. Unter 'time' findet sich der Zeitstempel der Speicherung. Der Titel der Webseite wird unter 'title' gespeichert.

Folgend wird beschrieben, wie Bookmarks abgespeichert werden und welche Möglichkeiten Delicious zum Zugriff auf sie bietet. Die Motivationen aus Abbildung 16 dienen dabei als roter Faden.

4.3.3 Bookmarks zentral speichern

Um eigene Bookmarks an einem zentralen Ort speichern zu können, kann Delicious genutzt werden. Dabei bietet Delicious verschiedene Möglichkeiten zum Speichern von Bookmarks an. Entscheidet sich ein User, seine Bookmarks mit Delicious zu verwalten, so hat er die Möglichkeit vorhandene Bookmarks in seinen Delicious Account zu importieren.

Das Speichern neuer Bookmarks versucht Delicious so einfach wie möglich zu machen. Für die verbreiteten Browser sind Plugins und Erweiterungen verfügbar, mit denen die Transaktionskosten dieses Vorgangs gering ausfallen. Schon bei der Einrichtung eines User Accounts können diese Erweiterungen in den Browser des Users installiert werden. Will ein User ein Lesezeichen anlegen, braucht er dann nur den Post-to-del.icio.us Knopf in seinem Browser zu klicken. Damit wird ein Fenster ähnlich der Abbildung 18 angezeigt.

Um die zentrale Speicherung des Lesezeichens abzuschließen, werden vom User die in Abbildung 18 gezeigten Felder ausgefüllt. In das URL-Feld wird die Adresse der im Browser offenen Webseite automatisch eingetragen. Das Feld 'Name' wird mit dem Titel der Webseite initialisiert. Beschreibende Notizen können unter 'Notes' abgelegt werden. Sind in der Webseite Textbereiche vom User markiert worden, so werden sie ins 'Notes' Feld übernommen. Im Feld 'Tags' können Schlagwörter eingetragen werden. Die Tag-Auswahl wird von Delicious durch Vorschläge gestützt. Dabei schlägt Delicious Tags aus den Bereichen Recommended Tags, Network Tags und Popular Tags wie im Beispiel der

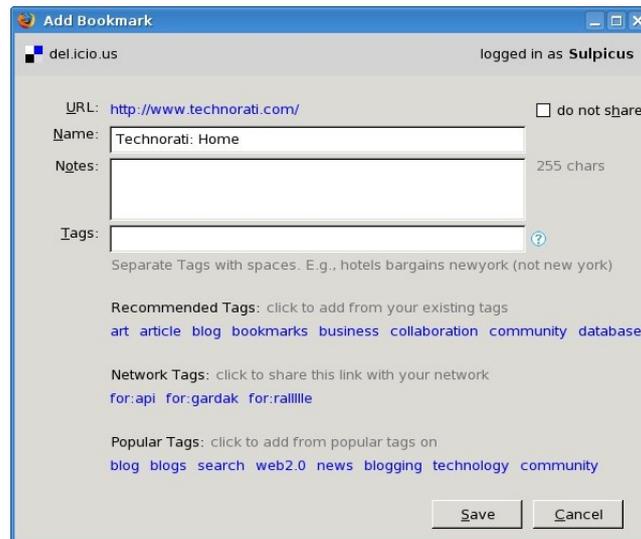


Abbildung 18: Lesezeichen mit Delicious speichern

Abbildung 18 vor. Folgend werden diese Bereiche kurz erklärt :

Recommended Tags sind Schlagwörter, die der speichernde User selbst schon einmal verwendet hat. Sie werden durch die Schnittmenge der Tags die andere User für diese URL verwendet haben, mit den Tags des speichernden Users ausgewählt.

Network Tags bieten eine Möglichkeit, Lesezeichen für andere User zu speichern. Ein Network Tag hat dabei die Form *for:Username*. Wird ein Lesezeichen mit einem Network Tag versehen, erscheint dieses Lesezeichen dann beim Zieluser im Bereich 'Links for you'.

Popular Tags sind die häufigsten für diese Ressource verwendeten Schlagwörter. Für bei Delicious unbekannte noch nicht gespeicherte URLs, also neue Lesezeichen, kann Delicious keine Vorschläge machen. In diesem Fall werden keine Tags in diesem Bereich vorgeschlagen.

Der User muss sich nicht an die vorgeschlagenen Tags halten, er kann selbst und frei Schlagwörter wählen, die seinen Ansprüchen gerecht werden. So schreibt die Delicious Hilfe zu Tags : „You make up tags as you need them, and use the tags that make the most sense to you“²⁷.

4.3.4 Eigene Bookmarks finden

Der Zugriff auf die unter Delicious gespeicherten Bookmarks ist auf verschiedenen Wegen möglich. Der einfachste Zugriff ist dabei mit Browser Erweiterungen wie z.B. dem

²⁷<http://del.icio.us/help/tags>, abgerufen 04.04.2007.

Delicious Bookmarks Add-on für den Webbrowser Firefox möglich. Eine andere Möglichkeit zum Zugriff besteht darin, durch den Aufruf einer URL die Liste der Bookmarks eines Users als Webseite anzeigen zu lassen. So würde der Aufruf der URL <http://del.icio.us/Sulpicus> z.B. die Bookmarks des Users Sulpicus im Browser auf einer Webseite zeigen. Eine weitere häufig verwendete Zugriffsmöglichkeit ist der Abruf von Bookmarks als RSS-Feed²⁸. Der Zugriff auf Delicious Bookmarks über RSS-Feeds ist so populär, dass über ihn die Hälfte aller Anfragen an Delicious laufen (Schachter 2007a). Auch der Zugriff auf die Bookmarks eines Users ist über einen RSS-Feed möglich. Für den User Sulpicus steht dieser Feed unter der URL <http://del.icio.us/rss/Sulpicus> zur Verfügung.

Der zweite Punkt der Motivationsliste (vgl. Abbildung 16), das Speichern von Bookmarks für sich selbst und auch die Möglichkeit, Bookmarks für Freunde und Bekannte zu speichern, wurde schon vorgestellt (vgl. Abschnitt 4.3.3). Durch Benutzen der beschriebenen Network-Tags können Lesezeichen explizit für andere Delicious-User zur Verfügung gestellt werden. Dieses Vorgehen ist nicht der einzige Weg, um Anderen Zugriff zu seinen Bookmarks zu geben. Durch die Voreinstellung einer öffentlichen Speicherung von Bookmarks sind Lesezeichen öffentlich verfügbar. Damit genügt es, Freunden und Bekannten einfach seinen Delicious Usernamen bekannt zu geben und ggf. ein relevantes Tag zu nennen, wenn man sie auf eine bestimmte Webseite hinweisen möchte.

4.3.5 Fremde Bookmarks finden

Zum dritten Punkt der Motivationsliste (vgl. Abbildung 16), der Möglichkeit zu sehen, welche Bookmarks von anderen Usern gespeichert werden, stehen in Delicious eine ganze Reihe von interessanten Listen und nützlichen Explorationswegen zur Verfügung. Zuerst werden die Hotlist, Recent Bookmarks und Popular Bookmarks betrachtet. Anschließend werden die Möglichkeiten des Browsing nach Tags und nach URLs gezeigt. Es folgt die

Betrachtung der von Delicious zur Verfügung gestellten Listen :

Die Hotlist ist eine Auflistung von Bookmarks, die von vielen Usern innerhalb der letzten Zeit gespeichert wurden. Hier finden sich Links, die oft erst seit wenigen Tagen im Delicious System sind und trotzdem schon zwei- oder dreistellige Userzahlen aufweisen.

Die Popular Bookmarks sind eine Art erweiterte Hotlist. Die hier gelisteten Bookmarks sind schon länger im Delicious System, als das bei den Bookmarks der Hotlist der Fall ist. Dabei werden die Bookmarks nach der Häufigkeit der Speicherung durch

²⁸RSS steht dabei für Really Simple Syndikation und ist ein XML-Protokoll zur Übertragung von Benachrichtigungen. RSS löst dabei ein Problem, das entsteht, wenn User sich für Webseiten interessieren deren Inhalte unvorhersehbar aktualisiert werden. Sobald die Zahl dieser Webseiten steigt, wird es für den User sehr mühselig, diese manuell auf Neuigkeiten zu prüfen. RSS gibt dem Webseitenbetreiber die Möglichkeit, Benachrichtigungen zu Änderungen in Listenform als so genannten RSS-Feed, seinen Usern zur Verfügung zu stellen. User können News- oder RSS-Aggregatoren nutzen, um diese Feeds regelmäßig zu prüfen und über Neuigkeiten informiert zu bleiben (vgl. Software Garden, Inc. 2004).

User absteigend gelistet, so dass die populärsten Bookmarks auf den ersten Rängen dieser Liste stehen.

Die *Recent Bookmarks* listen die gerade in Delicious gespeicherten Bookmarks. In dieser Liste finden sich Einträge, deren Aktualität im Minutenbereich liegt. Sie ist unter der URL <http://del.icio.us/rss/popular> abrufbar und verfügt über einen interessanten Parameter. Über den URL Parameter `min=X` lässt sich einstellen, ob alle neuen Bookmarks gezeigt werden sollen, oder nur jene, die schon *X* mal von Usern gespeichert wurden.²⁹

Über diese Listen lässt sich gut beobachten, was aktuell in Delicious gespeichert wird. Dieses ungezielte Vorgehen wird von Möglichkeiten zur spezifischen Suche ergänzt. Dazu gehört das Verwenden der Delicious-Suchmaschine und das gezielte Browsen über Tags und URLs nach interessanten Usern.

Die Delicious-Suchmaschine

Auf fast jeder von Delicious angezeigten Webseite findet sich die Delicious-Suchmaschine. Sie sucht den eingegebenen Begriff in Beschreibungen, Tags und Notizen aller bei Delicious gespeicherten Bookmarks. Die Menge der gefundenen Ergebnisse ist entsprechend groß. Da hier auch Beschreibungen und Notizen durchsucht werden, erhält man, verglichen mit einer auf Tags begrenzten Suche, teilweise irrelevante Resultate.

Die Suche über Tags

Um nur nach bestimmten Tags zu suchen, können speziell konstruierte URLs aufgerufen werden. Interessiere ich mich z.B. für Webseiten zum Thema Volvo, könnte ich <http://del.icio.us/tag/volvo> aufrufen. Diese Variante der Suche hat einen interessanten Vorteil. Die Ergebnisseite trägt den Verweis auf einen entsprechenden RSS-Feed, mit dem ich mich über neue Links, die das Volvo Tag tragen, benachrichtigen lassen kann.

Die Suche über URLs

In Delicious gibt es die Möglichkeit, nach gespeicherten Bookmarks über deren URL zu suchen. Mit dieser Suche kann man User und Tags zu einer URL entdecken, sie ist besonders im Zusammenhang mit der weiter unten beschriebenen Methode zum Aufspüren interessanter User relevant. Möchte ich Bookmarks finden, die auf die Heise News Webseite zeigen könnte ich <http://del.icio.us/url> aufrufen. Im Eingabefeld zu 'check url' kann eine URL eingetragen werden, beispielsweise www.heise.de. Per Knopfdruck wird die Heise-URL in einen MD5-Schlüssel umgewandelt, mit dem die Ergebnisseite abgerufen wird. Dort finden sich die aktuellsten Bookmarks, die auf <http://www.heise.de> verweisen. Es lassen sich sowohl

²⁹Dieser Parameter lässt sich also als Filter verwenden, so zeigt z.B. <http://del.icio.us/rss/popular?min=20> nur kürzlich gespeicherte Bookmarks die schon von mindestens 20 Usern hinterlegt worden sind.

die Usernamen, als auch die verwendeten Tags ablesen. Alternativ lässt sich dieser Vorgang ähnlich wie bei der Suche mittels Tags über eine konstruierte URL der Form `http://del.icio.us/url/check=?url=www.heise.de` auslösen.

Damit sind die Möglichkeiten zur Exploration der Delicious Bookmarks noch nicht ausgeschöpft. Besonderen Wert hat gerade das Aufspüren von interessanten Delicious-Usern, mit dem sich der folgende Abschnitt befasst.

4.3.6 Interessante User entdecken

Es lassen sich in Delicious interessante User entdecken. Dabei interessieren aber nicht die User an sich, sondern die von ihnen gespeicherten Bookmarks. Wenn man sich für bestimmte Themen interessiert und mit diesen genauer beschäftigt, speichert man als Delicious-User regelmässig interessante Bookmarks. Andere User zu finden, die gleiche oder ähnliche Interessen verfolgen, bringt einen Informationsvorteil. Man kann bei Delicious über diese User quasi von Informationsarbitrage profitieren.

Es gibt verschiedene Wege, interessante Delicious-User zu entdecken. Folgend sollen Möglichkeiten dazu beispielhaft verdeutlicht werden. Interessante User sind interessant weil sie sich mit Themen beschäftigen, die auch für den Informationssuchenden von Interesse sind. Diese Themen spiegeln sich sowohl in den von ihnen gespeicherten Bookmarks als auch in den verwendeten Tags.

Will man konkrete interessante User für zu einem bestimmten Thema finden, können Tags und URLs als *Kriterien* oder *Filter* dienen, um solche User aufzuspüren. Angenommen, ich interessiere mich für das OpenSource Linux-Betriebssystem Ubuntu vielleicht vor dem Hintergrund, dass ich Ubuntu administriere und weil ich eigene Erweiterungen in Form von Softwarepaketen zur Verfügung stellen möchte. Schon aus diesem letzten Satz lassen sich einfache Tag-Kriterien ableiten. D.h. ich könnte die Begriffe 'opensource', 'linux' und 'ubuntu' als Filter benutzen, um Bookmarks und über diese Bookmarks für mich interessante User zu finden. Für den Fall, dass ich diese Interessen nicht gerade erst gewählt habe, sondern mich vorher schon mit dem Thema beschäftigt habe, kann ich die Liste der Tag-Kriterien um spezifischere Schlagwörter erweitern. Ich bin dann auch in der Lage, einige für das Thema relevante URLs zu benennen, die als URL-Kriterium zur Identifikation interessanter User eingesetzt werden können.

Ein wichtiger Faktor zur erfolgreichen Suche mit diesen Kriterien ist die Unterscheidung von spezifischen und allgemeinen URLs und Schlagwörtern. Wenn ich Linux als Tag verwende, erhalte ich deutlich mehr Treffer, als wenn ich mit dem Tag 'ubuntu' suche. Geht es mir aber um spezifische Informationen zu Ubuntu wäre eine Suche mit dem Tag Linux ineffizient, da ich damit eine Menge irrelevanter Ergebnisse bekommen würde.

Tags oder URLs können eingesetzt werden, um interessante User zu indentifizieren. Im Folgenden ist eine Auswahl von Tag- und URL-Kriterien, die dem Beispiel gerecht werden können, aufgelistet. Sie sind dabei von allgemeinen zu speziellen Kriterien geordnet :

Tag-Kriterien :

- opensource
- linux
- ubuntu
- apt

URL-Kriterien :

- <http://www.ubuntu.com>
- <http://ubuntuforums.org>
- <http://cutlersoftware.com/ubuntuinstall>
- <http://www.debian.org/doc/manuals/apt-howto>

Mit diesen Kriterien und den beschriebenen Verfahren zur Suche lassen sich User entdecken. Konkrete Suchanfragen startet man am besten von speziellen Kriterien ausgehend. In der Regel wird die Suche mit Tag-Kriterien einfacher sein, da zur Auswahl sinnvoller URL-Kriterien mehr Vorwissen nötig ist.

Werden über diese Kriterien User gefunden, müssen diese evaluiert werden. Einen guten Überblick zu den Themen, für die sich ein User interessiert, gibt ein Blick auf seine Tag Cloud, die rechts auf der Delicious Webseite des Users angezeigt wird. Auch die angezeigten Bookmarks des Users können Aufschluss geben, ob eine weitere Beobachtung lohnt.

Hat man einen interessanten User identifiziert, kann man ihn in sein Netzwerk aufnehmen. Diese Delicious-Komponente erleichtert die Beobachtung interessanter User. Die Seite <http://del.icio.us/network/MyUserName> zeigt die aktuellsten Bookmarks der beobachteten User.

4.4 Zusammenfassung

In diesem Abschnitt wurde Social Bookmarking im allgemeinen und Delicious im speziellen als Basis für automatisierte Methoden zur Unterstützung von Markt-, Trend- und Zukunftsforschung vorgestellt.

Es folgen die Kernaussagen in Stichpunktform :

- In Social Bookmarking Systemen werden Web-Lesezeichen gespeichert.
- Social Bookmarking ist ein Peer Production System, mit dem als Informationsgut eine Sammlung aktueller Web-Lesezeichen entsteht.
- Klare Struktur der Daten und Themenvielfalt begründen die Wahl von Social Bookmarking Systemen als Kandidat für automatisierte Auswertung und Verarbeitung.
- Delicious ist der Marktführer unter den Social Bookmarking Systemen mit mehr als zwei Millionen registrierten Usern, die Themenvielfalt und Aktualität der Bookmarks sicherstellen.
- Delicious senkt mit Browser-Erweiterungen, RSS-Feeds und klar strukturierten URLs die Transaktionskosten zur Nutzung seines Dienstes.

- Mit Hilfe der Hotlist, Popular- und Recentlisten lassen sich Userverhalten und aktuelle Delicious Bookmarks beobachten.
- Es lassen sich mit der Delicious-Suchemaschine, den Tag- und den URL-Kriterien interessante User entdecken.
- Die Beobachtung von interessanten Usern wird durch die Delicious Netzwerk Komponente ermöglicht.

Die Möglichkeiten zur Beobachtung von Userverhalten auf Delicious ist eine Voraussetzung für die Einsetzbarkeit von automatischen Methoden zur Auswertung von Daten. Interessante Bookmarks und interessante User zu finden ist, abgesehen von den zur Verfügung gestellten Listen, ein für den Informationssuchenden aufwendiges Unternehmen.

Im folgenden Abschnitt wird ein System vorgestellt, mit dem die in Abschnitt 2 beschriebenen Methoden zur Markt-, Trend- und Zukunftsforschung unterstützt werden können. Dieses System wertet Social Bookmarking Daten am Beispiel von Delicious Bookmarks aus.

5 Systembeschreibung

Dieser Abschnitt stellt ein System zur automatisierten Auswertung von Lesezeichen am Beispiel der Social Bookmarking Anwendung Delicious vor. Damit können die in Abschnitt 2 beschriebenen Methoden zur Markt-, Trend- und Zukunftsforschung gestützt werden.

Im Abschnitt 5.1 werden die Grundlagen des Systems und die Lösungsidee vorgestellt. Der Abschnitt 5.2 beschreibt die Funktionsweise und Umsetzung des Systems detailliert.

Die Auswertung von Social Bookmarking Systemen und überhaupt von User Generated Content ist für viele Forschungszwecke sinnvoll. Bevor näher auf das hier vorzustellende System eingegangen wird, lohnt ein Blick auf bereits existierende Ansätze zur automatisierten Auswertung solcher Daten.

Literatursichtung

Einen guten allgemeinen Überblick zu Forschungsmethoden, die auf Internet-Daten aufbauen, geben Wood und Kauffman (2006). Sie ordnen neue Vorgehensweisen anerkannten theoretischen Forschungsmethoden zu, neue Methoden und Erweiterungen existierender Forschungsmethoden werden von ihnen vorgestellt. Ihre Untersuchung basiert dabei auf einer umfangreichen Literaturrecherche, deren Ergebnisse durch ihre Ordnung klassifiziert werden (vgl. Wood und Kauffman 2006, S.16).

Tagging

Besondere Aufmerksamkeit in der Forschung hat der Mechanismus des Tagging auf sich gezogen. Einige relevante Artikel zu diesem Thema wurden bereits im Abschnitt 4.2 erwähnt. Zwei weiterhin erwähnenswerte Untersuchungen, die am Social Bookmarking Beispiel Delicious durchgeführt wurden, sind in diesem Zusammenhang Macgregor und McCulloch (2006) und Cattuto et al. (2006).

Schmitz (2006) und Begelman et al. (2006) suchen nach Strukturen in Tagging Daten. Während Schmitz (2006) mit dem Beispiel des Fotosharing Dienstes Flickr arbeitet, wird von Begelman et al. (2006) Delicious als Basis der Tagging Daten verwendet. Auch Shen und Wu (2005) untersuchten Delicious Tagging Daten und stellen Skalierungsfreiheit (scale free). Scharffe und Ding (2006) stellen einen Mechanismus zur Strukturierung von Webdaten vor dessen Basis getaggte Dokumente sind.

Trends und Social Bookmarking

Besondere Aufmerksamkeit wird in dieser Arbeit auf Mechanismen zur Erkennung von Trends in Social Bookmarking Daten gelegt. Die für diese Aufgabe interessanten Artikel müssen sich nicht mit Social Bookmarking befassen, um wertvoll zu sein. Social Bookmarking benutzt, wie viele andere User Generated Content Arten, auch den Mechanismus des Tagging. Mit Trend-Erkennung in Tagging Daten beschäftigen sich Glance et al. (2004) am Beispiel von Blogs, Dubinko et al. (2006) am Beispiel Flickr. Ein wichtiger Punkt in der Erkennung von Trends ist die zu beachtende zeitliche Komponente von

Informationen. Amitay et al. (2004) zeigen, wie wertvoll die bisher im Internet häufig vernachlässigte zeitliche Komponente von Informationen zur Erkennung von bedeutsamen Ereignissen und Trends ist.

Eine ganze Reihe interessanter Artikel zu Social Bookmarking, Auswertungsmethoden und zur Trend Erkennung in Social Bookmarks sind von der Arbeitsgruppe Wissensverarbeitung des Fachbereichs 17 an der Universität Kassel³⁰ publiziert worden. Jäschke et al. (2006) suchen nach Wegen, Nutzergruppen in Folksonomies³¹ zu entdecken. Hotho et al. (2006a) befassen sich mit der Struktur von Folksonomies und stellt Bibsonomy³² vor. Bibsonomy ist ein von der Arbeitsgruppe Wissensverarbeitung implementiertes Social Bookmarking System, mit dem nicht nur Web-Lesezeichen, sondern auch Informationen zu wissenschaftlichen Artikeln verwaltet werden können. In diesem Artikel wird kurz der FolkRank-Mechanismus vorgestellt, der in Hotho et al. (2006b) genauere Betrachtung findet. Der FolkRank-Mechanismus ist eine Adaption des PageRank Algorithmus auf die dreigeteilte Hypergraph-Struktur einer Folksonomy. Mit ihm können themenspezifische Ranglisten von Usern, Tags und Ressourcen berechnet werden. Hotho et al. (2006c) vergleichen den FolkRank Mechanismus mit der von Dubinko et al. (2006) beschriebenen Methodik.

Im folgenden Abschnitt wird aus Überlegungen zu den Anforderungen an ein System zur automatisierten Auswertung von Social Bookmarking Daten und aus bestehenden Forschungsansätzen eine Lösungsidee für ein solches System entwickelt.

5.1 Grundlagen

Um ein System zur Unterstützung der Methoden zur Markt-, Trend- und Zukunftsforschung mit automatisierten Auswertungen von User Generated Content erstellen zu können, sind Vorüberlegungen nötig. Diese Überlegungen lassen sich in drei Bereiche aufteilen :

Anforderungen an die Daten

Als Grundlage für die zur Analyse verwendeten Daten ist der Social Bookmarking Dienst Delicious (vgl. Abschnitt 4.1) ausgewählt worden. Diese Datengrundlage verspricht, den Ansprüchen der Themenvielfalt und Aktualität zu genügen. Ein konkretes System, das von dieser Datenbasis profitieren möchte, muss diese Delicious Daten möglichst vollständig erfassen und aktuell halten können.

Anforderungen an das System

Ein System zur Auswertung von Social Bookmarking Daten muss gegenüber manuellen Methoden Vorteile bringen. Im Abschnitt 4.3 wurden manuelle, von Delicious zur Verfügung gestellte Methoden beschrieben, mit denen interessante User

³⁰Fachbereich 17 umfasst an der Universität Kassel die Disziplinen Mathematik und Informatik.

³¹Der Begriff Folksonomy ist eine Zusammensetzung aus Folk und Taxonomy. Er wurde von Vander Wal (2004) geprägt.

³²<http://www.bibsonomy.org> geprüft 12.04.2007

aufspürbar und beobachtbar sind. Um die Vorteile automatischer Auswertungsmethoden sicher zu stellen, muss, gerade aufgrund der notwendigen Datenmenge, besonders auf effiziente Algorithmen und Implementierung geachtet werden.

Anforderungen aus der Arbeitsweise von Markt-, Trend- und Zukunftsforschern

Ein konkretes System soll Markt-, Trend- und Zukunftsforscher bei ihrer Arbeit unterstützen. Im Abschnitt 2 sind ihre Arbeitsmethoden vorgestellt worden. Horx und Wippermann (1996, S.82) schreiben im Zusammenhang mit dem Scanning von Medien (vgl. Abschnitt 2.3.2) : „Jeder hat dabei bestimmte Schwerpunkte und auch bestimmte Mediensegmente zu beobachten. Jeder muß Expertenwissen auf diesem Sektor ansammeln. “Das ist der Ansatzpunkt für einen möglichen Algorithmus. Experten beschäftigen sich mit bestimmten Themen auf lange Zeit, nur so können sie Mentale Modelle (vgl. Abschnitt 2.3.2) aufbauen, die sie zu Experten werden lassen.

Die Anforderungen an Daten und System werden im nachfolgenden Abschnitt der Umsetzung ein Rolle spielen. Hier wird mit den Anforderungen aus der Arbeitsweise von Markt-, Trend- und Zukunftsforschern, manuellen Explorationsmöglichkeiten von Delicious Bookmarks und Ansätzen aus der wissenschaftlichen Literatur das Konzept eines Systems zur automatisierten Auswertung von Social Bookmarks erstellt.

Das im Abschnitt 4.3 beschriebene manuelle Verfahren zum Aufspüren und Beobachten interessanter User hat besonderen Wert für Markt-, Trend- und Zukunftsforscher. Dabei entspricht gerade das Identifizieren von Usern, die sich mit bestimmten Themen beschäftigen, dem Vorgehen von Trendexperten. Allerdings ist dieses Vorgehen aufwendig und gerade wenn viele User indentifiziert und beobachtet werden sollen, ineffizient.

Das Schlüsselement der Methode zur Indentifikation interessanter User ist die oben beschriebene Auswahl von Tag- und URL-Kriterien. Sie können auch in einem System zur automatisierten Auswertung verwendet werden. Statt aber einzelne User zu identifizieren, können mit diesen Kriterien User Gruppen gebildet werden. Verfahren zur automatisierten Clusterbildung oder Beziehungsanalyse, wie z.B. der FolkRank, können auch zur Gruppenbildung verwendet werden. Das automatische Bilden von Gruppen erscheint hier allerdings wenig sinnvoll, da es mit entsprechendem Aufwand verbunden ist. Durch die manuelle Auswahl der Kriterien zur Gruppenbildung wird das System effizient gehalten, es werden nur Gruppen angelegt und analysiert, die für den User des Auswertungssystems von Wert sind.

Zur effizienten Beobachtung der Usergruppen bietet sich der von Dubinko et al. (2006) beschriebene Mechanismus an. Dieser Algorithmus lässt sich auf beliebige mit Zeitstempel versehene Ressourcen anwenden. Mit ihm kann ein Ranking von Ressourcen über beliebige temporale Intervalle effizient implementiert werden. Mit diesem Verfahren, der zeitlichen Aggregation von Bookmarkingereignissen, wird es möglich Trends im Bookmarkingverhalten der Delicious-User zu entdecken.

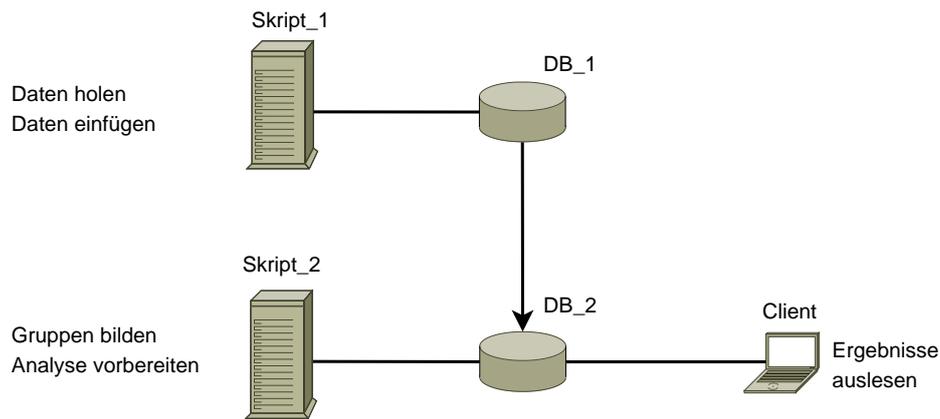


Abbildung 19: Server und Arbeitsaufteilung

Für User des Systems ergeben sich damit folgende Aufgaben oder Möglichkeiten :

1. Mit Tag- oder URL- Kriterien kann er User zu Gruppen zusammenfassen.
2. Mit frei wählbaren Intervallen kann er die in den Gruppen gespeicherten Bookmarks nach Häufigkeit der Speicherung beobachten.

Der folgendende Abschnitt befasst sich mit der Umsetzung dieses Systems. Dabei werden die verwendeten Algorithmen und Methoden detaillierter beschrieben. Zuerst wird auf das Sammeln der Daten eingegangen. Die Umsetzung der Gruppenbildung schließt sich an. Die Erläuterung der Verarbeitung und Analyse von in den Gruppen gespeicherten Bookmarks beschließt den nächsten Abschnitt.

5.2 Umsetzung

In diesem Abschnitt wird die Umsetzung des in Abschnitt 5.1 beschriebenen Konzepts für ein System zur automatisierten Unterstützung der in Abschnitt 2 beschriebenen Methoden zur Markt-, Trend- und Zukunftsforschung vorgestellt.

Um diese Aufgabe zu bewältigen müssen folgende Punkte bearbeitet werden :

1. Daten sammeln
2. Gruppen bilden
3. Analyse vorbereiten
4. Ergebnisse auslesen

Die Auswahl der zu verwendenden Software folgt Erwägungen zu Verfügbarkeit und Performance. Zur Umsetzung des Systems wurden Perl-Skripte und MySQL-Datenbanken gewählt. In Perl geschriebene Programme werden beim Programmstart in Maschinencode

übersetzt. Abgesehen von der dadurch erzeugten Startverzögerung, ist die Ausführungsgeschwindigkeit von Perl Code mit in C geschriebenen Programmen vergleichbar. Für Perl steht mit dem CPAN³³ ein zentrales Repository von Perl Erweiterungen in Form von Modulen zur Verfügung. Sie helfen, den Entwicklungsaufwand für Perl Programme gering zu halten. MySQL ist wie Perl unter der Gnu Public Licence lizenziert und damit frei verfügbar.

Die Anforderungen an die Leistungsfähigkeit der Datenbank sind, bedingt durch die zu verarbeitenden Datenmengen, hoch. Das MySQL diesen Anforderungen gerecht werden kann, zeigt sich an Beispielen großer Internet-Anwendungen wie z.B. Google, die auf MySQL-Datenbanken setzen um ihre Suchmaschine zu betreiben. Delicious selbst setzt auch MySQL als Datenbank für seine Bookmarks ein, wie aus Einträgen im Delicious-Blog (vgl. Schachter 2005) deutlich wird.

In Abbildung 19 ist die Verteilung der Aufgaben auf vier Rechner dargestellt. Perl und MySQL wurde dort auf Basis des Linux Betriebssystems Ubuntu installiert. Der Rechner Skript_1 holt Bookmarks von Delicious und speichert sie in der MySQL-Datenbank auf DB_1. Von dort werden die gespeicherten Bookmarks auf eine MySQL-Datenbank auf DB_2 repliziert. Der Rechner Skript_2 verwaltet die Gruppen und bereitet die Bookmarks auf, so dass Ergebnisse vom Client ausgelesen werden können.

5.2.1 Daten sammeln

Die in Abschnitt 4.3 beschriebenen Zugangsmethoden zu Delicious können genutzt werden, um Bookmarks für das Auswertungssystem einzusammeln. RSS-Feeds eignen sich aufgrund ihrer klaren Struktur besonders gut, um die Bookmarks der Delicious-User auszulesen. Das Abrufen der Bookmarks von User-RSS-Feeds der Form `http://del.icio.us/rss/Username` kann nur ausgeführt werden, wenn Delicious-Usernamen bekannt sind. Eine Möglichkeit, Usernamen zu finden besteht im regelmäßigen Abruf der Recent-Liste. Dort sind die gerade neu gespeicherten Delicious Bookmarks mit ihren Usernamen gelistet.

Delicious Bookmarks werden durch das Auslesen der Recent-Liste, mit der neue Usernamen entdeckt werden, und einer regelmässigen Überprüfung von User-RSS-Feeds von Delicious abgeholt. In Abbildung 20 ist die Funktionsweise des Perl-Skripts zum Abholen und Einfügen der Daten schematisch dargestellt. Aus Performancegründen wurde das Abholen und Parsen der Daten durch Worker-Threads parallelisiert. Damit wird der bremsende Effekt, der durch das Warten auf die Delicious Daten entsteht, klein gehalten.

Die Worker bauen ihre Verbindung zu Delicious über Proxyserver auf. Diese wird zufällig aus einer Proxyliste ausgewählt. Die Verwendung eines Proxysystems ist notwendig, da viele Daten abgeholt werden sollen. Delicious, wie auch andere Yahoo Dienste, überwacht den Netzwerkverkehr auf IP-Basis. Übersteigt der von einer IP erzeugte Traffic bestimmte Grenzwerte, verweigert Delicious für eine Weile jeden weiteren Zugriff von dieser IP. Durch das Nutzen von Proxies kann dieser Mechanismus teilweise umgangen werden. Somit beschleunigen Proxies das Abholen der Bookmarks.

³³<http://www.cpan.org>, geprüft 13.04.07.

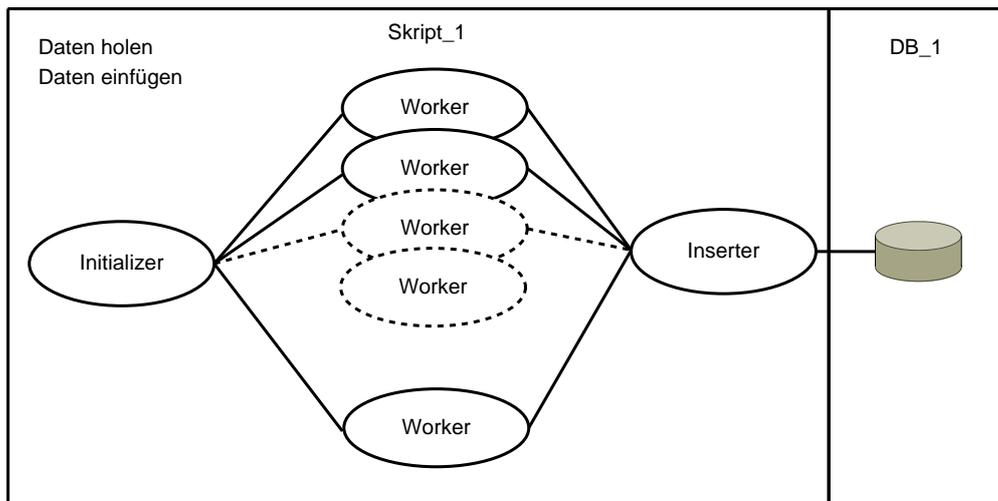


Abbildung 20: Thread Struktur für Holen und Einfügen der Bookmarks

Ist ein Nutzernamen einmal im System bekannt, kann er regelmäßig geprüft werden. Um unnötiges Abrufen zu verhindern, wurden eine Timeout-Funktion eingebaut. Enthält ein User-RSS-Feed nur wenig neue Daten, wird das Timeout für diesen Feed erhöht. Sind hauptsächlich neue, im System noch unbekannte Bookmarks, im Feed enthalten wird die Wartezeit zum nächsten Abruf des Feed reduziert.

Das Einfügen der Daten in die Datenbank wird vom Inserter durchgeführt. Um Kollisionen beim Einfügen der Daten in die Datenbank zu verhindern, wird nur ein Inserter für alle Worker verwendet. Um die Geschwindigkeit der Inserts zu erhöhen, wurde MySQL 5 verwendet. In dieser Version sind sowohl *Datenbank-Transaktionen* als auch *Stored Procedures* implementiert. Der Inserter verwendet beide Features, um Bookmarks möglichst schnell einfügen zu können.

Transaktionen stellen hauptsächlich Datenintegrität sicher, indem im Fall von Fehlern ein Rollback von zusammengehörenden Änderungen durchgeführt werden kann. Ein Nebeneffekt von Transaktionen kann zur Beschleunigung der Inserts genutzt werden. Durch Bündelung mehrerer Inserts in eine Transaktion kann der MySQL-Server Datensätze effizienter schreiben. Dadurch werden die Einfügeoperationen beschleunigt.

Stored Procedures sind eine Programmiersprache, mit der Funktionen im Datenbank Server definiert werden können. Sie helfen den Kommunikationsoverhead zwischen dem Inserter und der Datenbank zu minimieren und beschleunigen die Einfügeoperationen weiter.

Beim Einfügen der Bookmarks in die MySQL-Datenbank auf DB_1 werden vom Inserter die Tabellen der Abbildung 21 befüllt. Jedes Lesezeichen wird dabei doppelt gespeichert. In der Tabelle 'bookmarks' werden Lesezeichen komplett je als ein Datensatz

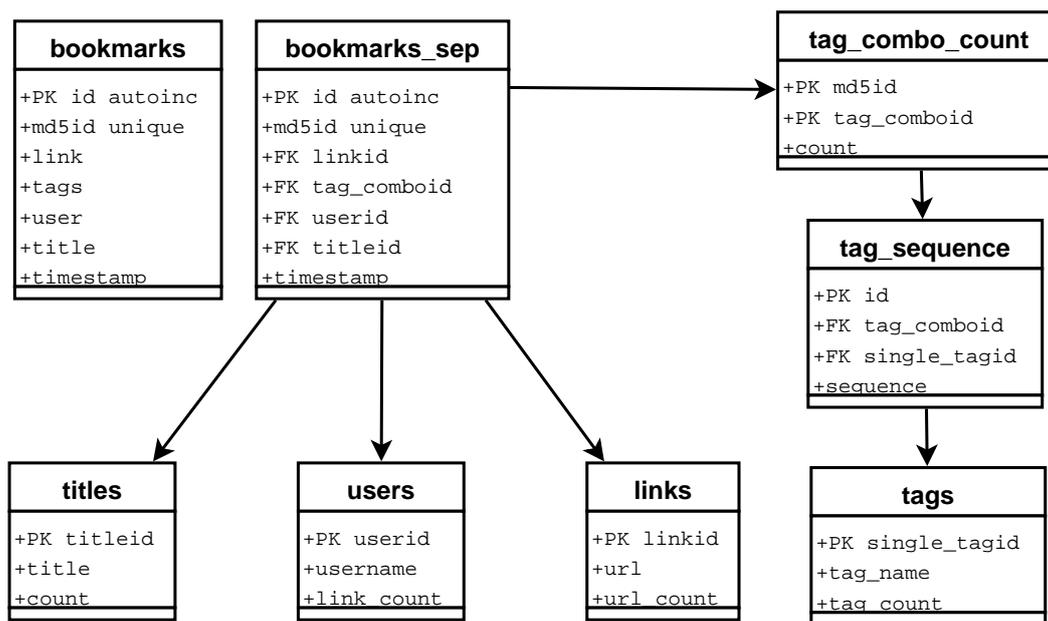


Abbildung 21: Tabellenstruktur zur Speicherung von Bookmarks

gespeichert. Dort werden keine weiteren Tabellen referenziert. Die 'bookmarks_sep' Tabelle ist die Haupttabelle zur Speicherung von Lesezeichen mit separaten Tabellen. Durch dieses Vorgehen können für die Elemente eines Lesezeichens Schlüssel erzeugt werden. Diese Schlüssel sind in 'bookmarks_sep' mit 'FK' markiert. Die separaten Tabellen 'links', 'tags', 'users' und 'titles' ermöglichen die Pflege von Datensatzzählern. Somit lässt sich z.B. sehr schnell herausfinden, wie viele Links ein User gespeichert hat oder wie häufig eine bestimmte URL gespeichert wurde.

Der in Abbildung 21 abgebildete Umweg zur 'tags' Tabelle ist durch die Struktur des 'Tags'-Felds und den Versuch, Datenverlust zu vermeiden, begründet. Das verwendete Skript liest Tags als String aus. Dabei haben Tags eine Reihenfolge. Sie ist mit der Reihenfolge der Tagauswahl des Bookmarkbesitzers identisch. Mit der Vermutung, dass die Reihenfolge der Tags bedeutungstragend sein kann, wurde nach einer Tabellenstruktur gesucht, die diese Information erhält und trotzdem Keys für die 'bookmarks_sep' Tabelle liefert. Die gefundene Lösung erfüllt diesen Anspruch. Mit ihr sind zudem performante Abfragen nach z.B. allen Bookmarks, die an der Position zwei ein bestimmtes Tag besitzen, durchführbar. Für die Funktionsweise des Systems ist diese Aufteilung aber nicht weiter relevant, sie kann als Basis für weiter Forschung zu Tagstrukturen dienen.

Besonders wichtig für eine performante Umsetzung der Analyse ist die Definition von Link-Schlüsseln, die in der Tabelle 'links' erzeugt werden. Im Abschnitt 5.2.3 wird beschrieben, wie diese 'linkid' benutzt wird, um die Auswertung der Bookmarks vorzubereiten.

Die Datensammlung ist eine kontinuierlich durchzuführende Aufgabe, weil Daten umfangreich und aktuell vorhanden sein müssen, wenn man relevante Trends frühzeitig

erkennen möchte.

Mit der sichergestellten Datensammlung endet dieser Abschnitt und leitet zum nächsten Aufgabenbereich über. Im nächsten Abschnitt wird das Bilden von Usergruppen durch Kriterien beschrieben. Das Bilden von Usergruppen ist notwendig, um den von Dubinko beschriebenen Algorithmus auf themenspezifische Ergebnisse zu erweitern.

5.2.2 Gruppen bilden

Dieser Abschnitt beschreibt Umsetzung und Möglichkeiten zum Bilden von Usergruppen. Dabei werden User über Tag- oder URL-Kriterien zu Gruppen zusammengefasst. Diese Gruppen sind eine Vorbedingung zur Auswertung von Social Bookmarks, deren Arbeitsweise in Abschnitt 5.2.3 beschrieben wird.

Momentan ist im System nur die Verwendung eines einzelnen Tag- oder URL-Kriteriums zur Bildung einer Usergruppe umgesetzt. Die Abbildung 22 zeigt die zur Gruppenverwaltung relevanten Strukturen der Datenbank. Zur Lastverteilung werden die Bookmarks kontinuierlich vom Rechner DB_1 auf DB_2 repliziert. Die Tabelle 'criteria' nimmt die vom User definierten Kriterien zur Gruppenbildung auf. Dabei wird zwischen Tag- und Domain-Kriterien unterschieden. Nach der Definition von Kriterien und damit von Gruppen können diese genutzt werden, um mit den in der DB_2 vorhandenen Bookmarks Usergruppen zu erzeugen. Das 'lastid' Feld der Tabelle 'criteria' dient dabei als Markierung dafür, bis zu welchem primären Schlüssel die Bookmarks auf dieses Kriterium untersucht worden sind.

Aus Gründen der Performance wird vom Kriterien-Scanner die 'bookmarks' Tabelle als Quelle für Bookmarks verwendet. Durch die Erweiterung des URL-Kriteriums zu einem Domain-Kriterium wird eine Verbesserung im Vergleich zu den von Delicious selbst zur Verfügung gestellten Möglichkeiten des Filterns nach Webadressen erreicht. Delicious selbst bietet bei der Suche nach URLs nur die Möglichkeit, exakte Treffer oder Bookmarks zu finden³⁴. Die von diesem System verwendete Methode erlaubt durch Zeichenkettenvergleiche Domain-Muster zu erkennen. Damit kann statt einem URL-Kriterium, das nur Lesezeichen auswählt, die auf eine konkrete Webseite zeigen, ein Domain-Kriterium realisiert werden. Mit diesem Domain-Kriterium können ganze Domains ausgewählt werden, z.B. würde das Domain-Kriterium mit dem Wert „http://www.heise.de/“ alle Lesezeichen auswählen, die auf eine Seite in der Heise-Domain verweisen.

Der Kriterien-Scanner entdeckt über diese Kriterien Usernamen. Für diese in den Bookmarking-Daten gefunden Treffer werden entsprechende Datensätze in der Tabelle 'criteria_users' gepflegt. Jede Kombination aus User und Kriterium bekommt dort einen Eintrag. Damit lassen sich aus der 'criteria_users' Tabelle die Usernamen der Gruppen

³⁴Delicious identifiziert URLs intern mit MD5-Schlüsseln. Diese lassen sich effizient berechnen. Offensichtlich beschränkt Delicious die Suchmöglichkeiten nach Bookmarks aus Performancegründen. Wird nur eine exakte Suchmöglichkeit angeboten, reicht die Umwandlung der Such-URL in einen MD5-Schlüssel aus, um über einen entsprechenden Index äusserst effizient Datensätze auszuliefern, die diesem Schlüssel entsprechen.

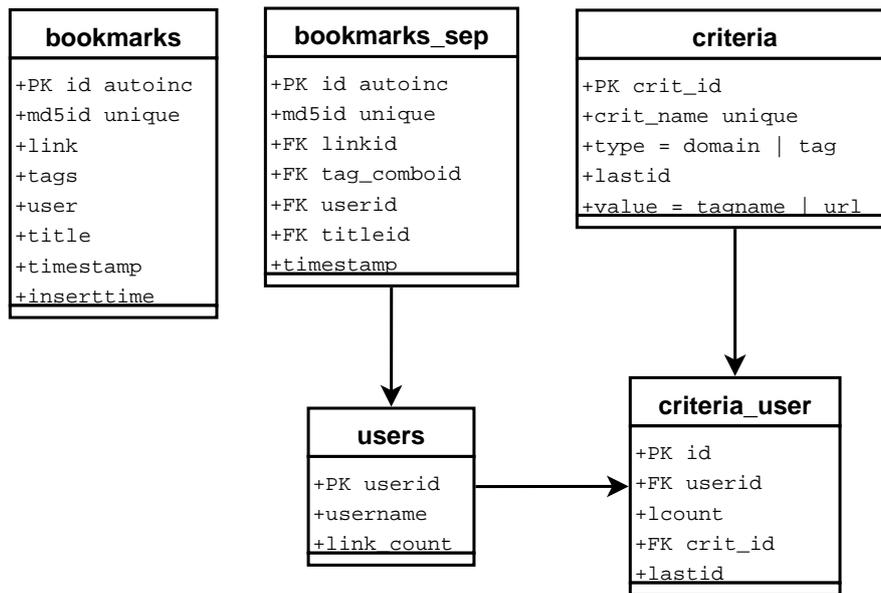


Abbildung 22: Tabellenstruktur zur Gruppenbildung nach Kriterien

auslesen. Diese Usernamen werden im nächsten Abschnitt verwendet, um die URL-Daten auf die Auswertung durch System-User vorzubereiten.

5.2.3 Analyse

Im letzten Abschnitt wurde das Bilden von Gruppen durch Tag- und Domain-Kriterien beschrieben. Folgend wird ein effizienter Mechanismus zum Ranking von Bookmarkingereignissen über beliebige Zeitintervalle beschrieben.

Der von Dubinko et al. (2006) beschriebene Algorithmus zur Auswertung von Web2.0 User Generated Content Daten ist ein effizienter Weg, um Trends in mit Zeitstempeln versehenen Daten zu entdecken. Ziel des Mechanismus ist es, eine effiziente Implementierung zum Ranking von Objekten über frei wählbare Auswertungsintervalle zu ermöglichen. Erreicht wird dieses Ziel u.a. durch Aufteilung des notwendigen Rechenaufwandes in einen *Vorbereitungsschritt* und einen *Echtzeitschritt*. Diese Aufteilung des Rechenaufwandes beschleunigt die Antwortzeiten des Systems.

Durch Datensätze mit Zeitstempeln werden Ereignisse, die sich hinter den Zeitstempeln verbergen, zählbar. Diese Zählbarkeit wird im System genutzt, um interessante Bookmarkingereignisse zu entdecken. Das Speichern eines Bookmarks ist auch eine Wertung des Inhalts dieser Webseite. Ein User legt ein Lesezeichen an, wenn der Inhalt dieser Seite einen Wert für ihn hat. Durch das Zählen dieser Ereignisse und die Betrachtung der Häufigkeiten über Zeitabschnitte können interessante Bookmarks entdeckt werden.

Zuerst wird der Vorbereitungsschritt dieses Mechanismus vorgestellt. Um ein Ranking von Bookmarkingereignissen zu erreichen, müssen diese in Zeitintervallen aggregiert werden. Um die Antwortzeit des Systems zu beschleunigen, werden Aggregationen vor-

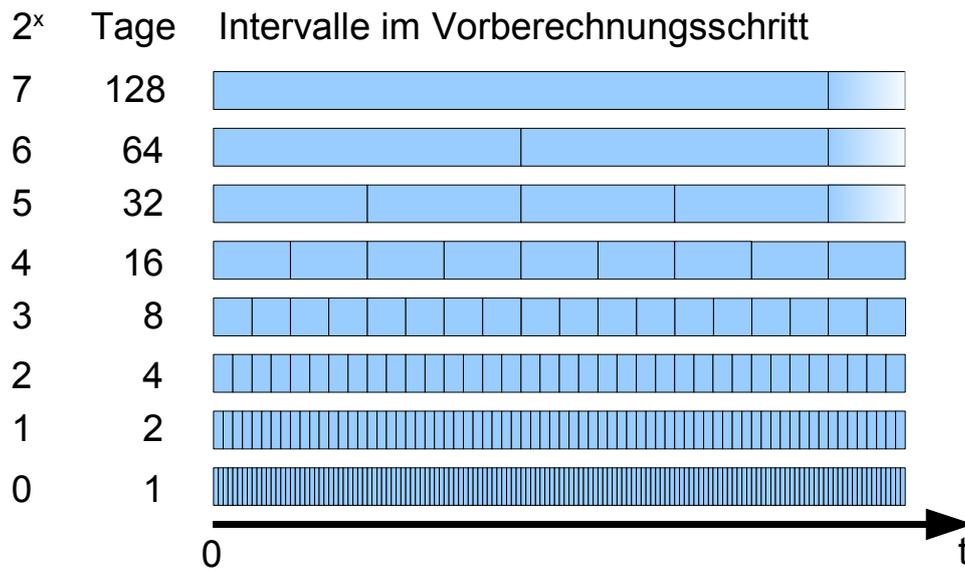


Abbildung 23: Vorberechnungsintervalle zur Analyse von Bookmarkingereignissen

verlagert. Ihre Ergebnisse werden in der Datenbank abgelegt. Hier wird der klassische Trade Off zwischen Daten und Rechenzeit ausgenutzt. Durch die Vorberechnung kann zwar Zeit bei der Abfrage eingespart werden, dieser Vorteil kann aber nur durch entsprechende Vergrößerung der vorgehaltenen Daten erreicht werden.

Um einen Kompromiss zwischen erforderlichem Datenvolumen und Abfragegeschwindigkeit zu erreichen, werden dabei nicht alle theoretisch möglichen Intervalllängen vorberechnet. Stattdessen werden die vorberechneten Intervalle auf Zweier-Potenz-Intervalle reduziert. In Abbildung 23 sind die vorzuberechnenden Intervalle graphisch dargestellt. Als Basiseinheit für die Aggregation von Ereignissen wurde ein Intervall von einem Tag gewählt. Für jede vorberechnete Intervalllänge ist in der Datenbank eine Tabelle vorgesehen. Für jedes Ereignis gibt es in jeder dieser Tabellen einen relevanten Datensatz, der einen Ereigniszähler enthält. Diese Ereigniszähler erhöhen sich mit Ereignishäufigkeit im entsprechenden Intervall.

Nach dem Befüllen der Analyse-Tabellen ist die Identifikation von Top-Ereignis-Objekten in den vorberechneten Intervallen trivial. Über einen entsprechenden Index der Datenbanktabellen können diese Top-Ereignis-Objekte effizient abgerufen werden.

Etwas schwieriger gestaltet sich die Berechnung von Top-Ereignissen über beliebige Intervalle. Dubinko et al. (2006) schlagen verschiedene Methoden vor, wie sich beliebige Intervalle aus den vorberechneten Intervallen zusammensetzen lassen. Im ersten Schritt wird der frei gewählte Intervall in vorberechnete Teilintervalle zerlegt. Im zweiten Schritt werden die Häufigkeiten der Top- k Objekte berechnet.

Zur Zerlegung des frei gewählten Intervalls schlagen Dubinko et al. (2006) zwei Methoden vor. Die erste Methode zerlegt den Intervall in additive Teile, so dass der frei gewählte Intervall aus den Summen von vorberechneten Intervallen zusammengesetzt

werden kann. Die zweite Methode verwendet sowohl Additionen als auch Subtraktionen zur Zerlegung des frei gewählten Intervalls in vorberechnete Einheiten. Die Methode der Addition und Subtraktion ist effektiver als die additive Methode, da sie einen frei gewählten Intervall in eine geringere Zahl von Teilintervalle zerlegt. Aus Gründen der einfacheren Implementierung wurde für diese System die additive Methode verwendet, um frei gewählte Intervalle in Teilintervalle zu zerlegen.

Der additive Algorithmus arbeitet rekursiv. Er sucht in einem frei gewählten Intervall nach den größten passenden vorberechneten Teilintervall. Ist dieser identifiziert, werden die verbleibenden Teilstücke auf der linken und rechten Seite des Ursprungsintervalls in einen rekursiven Aufruf dem gleichen Verfahren unterzogen.

Wenn die Zerlegung des frei gewählten Intervalls in vorberechnete Teilintervalle bekannt ist, lassen sich nun die Häufigkeiten berechnen. Dabei wird typischerweise nach den Top- k Objekten im gesamt Intervall gesucht. Die offensichtliche Methode, um die Häufigkeiten der Objekte des gesamten Intervalls zu berechnen, ist die Aggregation aller Einzelwerte der Teilinteralle. Dieses Vorgehen ist zwar korrekt, aber unnötig aufwendig. Dubinko et al. (2006) verwenden einen von Fagin et al. (2003) beschriebenen *Threshold-Algorithmus*, der die Top- k Objekte mit deutlich weniger Aufwand korrekt ermittelt. Auch dieses System verwendet den Threshold-Algorithmus, um die Top k URLs im Gesamtintervall zu identifizieren.

Der Threshold-Algorithmus greift auf die einzelnen Elemente, die Teilintervalle, sowohl sortiert als auch frei zu. Die vorberechneten Teilintervalle enthalten Objektschlüssel und Ereignishäufigkeit. Im Fall dieses Systems sind das URL-Schlüssel und Zähler, die die Speicherhäufigkeit der URL im Intervall zählen. Im Folgenden wird die Arbeitsweise des Threshold Algorithmus beschrieben (vgl. Fagin et al. 2003, S.620) :

1. Greife parallel auf die nach Häufigkeit sortierten Daten der Teilintervalle zu. Wenn ein Objekt R in einer Liste unter sortiertem Zugriff auftaucht, greife auf die Daten der anderen Teilintervalle zu, um die Gesamthäufigkeit $f(R)$ zu berechnen. Ist die Gesamthäufigkeit $f(R)$ eine der k höchsten bisher gesehenen Häufigkeiten, dann merke R und seine Häufigkeit $f(R)$.
2. Für jeden Teilintervall T_i sei x_i die Häufigkeit des letzten unter sortierten Zugriff gesehenen Objekts. Definiere den Threshold τ als $f(x_1, x_2, \dots, x_m)$. Beende den Algorithmus, wenn mindestens k Objekte gesehen wurden, deren Häufigkeit mindestens gleich τ ist.
3. Sei Y die Menge mit den k gesehenen Objekten größter Häufigkeit. Damit ist die Ausgabe des Threshold-Algorithmus die Menge $\{(R, f(R)) | R \in Y\}$.

Fagin et al. (2003) zeigen, dass dieser Algorithmus korrekt und in einem starken Sinn optimal zur Aggregation ist.

Das hier beschriebene System verwendet eine serialisierte Implementierung des Threshold-Algorithmus. Zuerst wird τ berechnet, danach werden die relevanten Häufigkeiten zu den Elementen unter sortiertem Zugriff berechnet und nach Punkt eins gemerkt. Die Abbruchbedingung aus Punkt zwei wird geprüft, bevor die nächste Runde beginnt, also

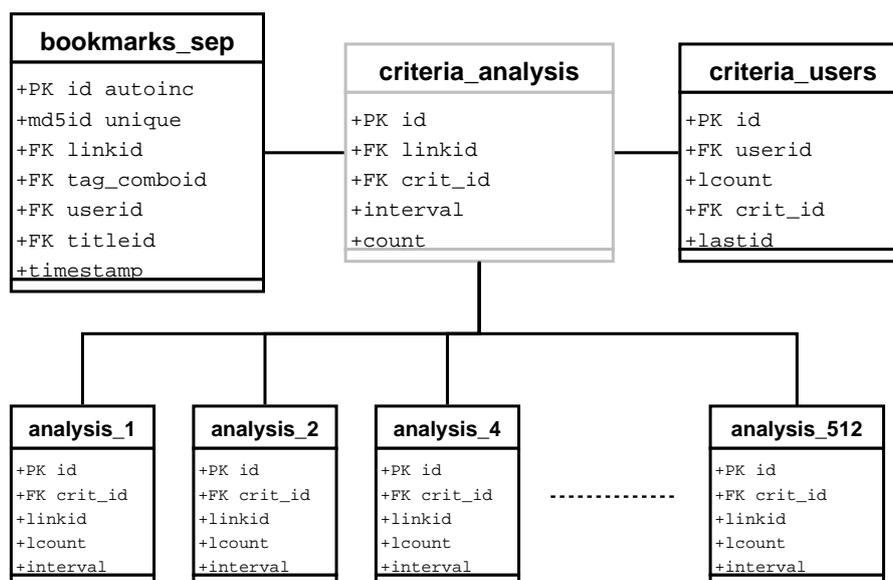


Abbildung 24: Datenbank-Tabellen zur Vorberechnung der Bookmarkingrankings

bevor auf das nächste niedrigere Element der Teilintervalldaten zugegriffen wird. Eine kleine eigene Optimierung ist das Auslassen der Prüfung von Elementen, die unter sortierten Zugriff in den Teilintervalldaten auftauchen, sich aber schon in der Top- k Arbeitsliste finden.

Die an der Umsetzung von Vorbereitungsschritt und Aggregationsschritt beteiligten Datenstrukturen der Datenbank sind in Abbildung 24 abgebildet. Die Tabellen 'analysis_1' bis 'analysis_512' werden im Vorbereitungsschritt befüllt. Die in die Analysetabellen eingetragenen Daten sind dabei mit der 'crit_id' gruppengebunden. Es werden dort nur Bookmarks von Usern gezählt, die in der Tabelle 'criteria_users' auftauchen. Dabei kann ein User und seine Bookmarks mehrfach, in verschiedenen Gruppen, gezählt werden, wenn er die definierten Kriterien dieser Gruppen erfüllt. Die grau markierte Tabelle 'criteria_analysis' ist ein prototypischer Stellvertreter der kleiner dargestellten 'analysis_X' Tabellen. Implementationsbedingt findet sich in criteria_users zu jedem User-Kriterien Datensatz eine 'lastid', in der der letzte für diesen Datensatz geprüfte Bookmarkschlüssel gespeichert wird.

Für den Echtzeitschritt der Analyse werden keine weiteren Datenbankstrukturen benötigt. Der Threshold-Algorithmus wurde wie alle Abfragemöglichkeiten via Perl-Skript realisiert.

Hier wurde die Umsetzung eines Systems zur automatisierten Auswertung von Social Bookmarks beschrieben. Im nächsten Abschnitt werden gesammelte Daten beispielhaft ausgewertet. Um die Leistungsfähigkeit des Systems zu verdeutlichen, werden die Abfragemöglichkeiten erweitert, so dass die Entwicklung von Bookmarkingereignissen graphisch dargestellt werden können.

6 Ergebnisse

In diesem Abschnitt werden einige Ergebnisse dieses Systems anhand der von Delicious gesammelten Social Bookmarking Daten beispielhaft betrachtet. Es werden erweiterte Auswertungsmöglichkeiten zur graphischen Darstellung von Bookmarkingereignissen vorgestellt und angewendet. Mit einer kritischen Betrachtung des Systems hinsichtlich der Leistungsfähigkeit zur Unterstützung von in Abschnitt 2 beschriebenen Methoden der Markt-, Trend- und Zukunftsforschung schliesst diesen Abschnitt ab.

6.1 Datensammlung

Mit dem Sammeln von Daten, also den Abruf von Delicious Bookmarks, wurde im April 2006 begonnen. Idealerweise sollte die Datensammlung kontinuierlich erfolgen. Bedingt durch die Weiterentwicklung des Systems und Schwierigkeiten beim Abruf der Daten sind aber Unterbrechungen im Einsammeln der Bookmarks aufgetreten. Die kleineren Unterbrechungen sind dabei durch eine zu geringe Zahl aktiver Proxies aufgetreten.³⁵ Dabei handelte es sich um Unterbrechungen von wenigen Tagen. Eine größere Unterbrechung von ca. einem Monat³⁶ gab es im Zeitraum September/Okttober 2006. Die daraus folgenden Datenlücken konnten teilweise durch Daten aus den Delicious-User RSS-Feeds geschlossen werden.

Bis zum 16.04.2007 wurden im System über 46 Millionen Delicious Bookmarks gespeichert. Damit sind 575.260 Delicious-User bekannt, deren Bookmarks ca. 14,8 Millionen eindeutige URLs enthalten. Zur Verschlagwortung nutzten sie ca. 2,3 Millionen eindeutige Schlagwörter oder Tags.

Bedingt durch die Tabellen zur Speicherung der Ergebnisse des Vorberechnungsschrittes und die an sich schon gewaltigen Datenmengen, ergibt sich ein recht großer Platzbedarf für die Datenbank. Am 17.04.2007 verbrauchten die Datenbanktabellen inklusive der Logfiles auf DB_2 (vgl. Abbildung 19) etwas mehr als 73 GB Speicherplatz.

6.2 Usergruppen und Kriterien

Im Abschnitt 5 wurden Tag- und Domain-Kriterien als Möglichkeit beschrieben, um User zu Gruppen zusammenzufassen. Im verwendeten System ist derzeit nur die Verwendung von einem Kriterium zum Bilden von Usergruppen implementiert. Folgend werden Beispiele solcher Kriterien gegeben, die als Basis einer Beispielauswertung dienen, deren Ergebnisse in den folgenden Abschnitten behandelt werden.

³⁵Durch das Nutzen von Proxies wird das Abholen der Daten beschleunigt. Damit kann das Sammeln der Daten allerdings nur fehlerfrei und performant laufen, wenn genügend aktive Proxies im System bekannt sind. Neue Proxies zu finden ist derzeit eine manuelle Aufgabe. Sind keine oder nur wenig aktive Proxies bekannt, können keine Daten abgeholt werden.

³⁶Zu Beginn der Implementierung lief das System, d.h. alle Skripte und Datenbanken, auf einem Rechner. Mit dem Anwachsen der gesammelten Daten wurde eine Aufteilung der Einzelaufgaben des Systems auf mehrere Rechner notwendig. Aus diesem Grund wurden im Zeitraum September/Oktober 2006 keine Daten von Delicious abgeholt.

| ID | Name | Type | Users | Value |
|----|-----------------|--------|--------|---|
| 1 | heise | domain | 6201 | http://www.heise.de/ |
| 2 | pageflakes | domain | 3238 | http://www.pageflakes.com/ |
| 3 | spiegel | domain | 6022 | http://www.spiegel.de/ |
| 4 | dict.leo | domain | 2373 | http://dict.leo.org/ |
| 5 | lwn | domain | 1665 | http://lwn.net/ |
| 6 | farecast | domain | 5047 | http://www.farecast.com/ |
| 7 | usability | tag | 30877 | Usability |
| 8 | web2.0 | tag | 102774 | web2.0 |
| 9 | personalization | tag | 896 | personalization |
| 10 | spotback | domain | 653 | http://spotback.com/ |

Tabelle 5: Gewählte Beispielkriterien zur Gruppenbildung inklusive Userzahlen

Die in Tabelle 5 dargestellten Kriterien wurden in Zusammenarbeit mit Ralf Rattay³⁷ gewählt.

Folgend wird beschreibend auf die verschiedenen gewählten Kriterien und ihre Zielthemen eingegangen.

1. Heise Domain-Kriterium

Die Heise News³⁸ sind eine Dienstleistung des Heise-Verlags. Dort erscheinen auch die Zeitschriften CT und IX. Ziel dieses Domain-Kriteriums ist es, deutschsprachige User zu selektieren, die sich für IT-Belange interessieren. Bis zum 17.04.2006 wurden 6201 Delicious-User in dieser Gruppe identifiziert.

2. Pageflakes Domain-Kriterium

Pageflakes³⁹ ist ein Web2.0-Dienst, der seinen Usern personalisierte Startseiten anbietet. Die eigene Seite kann dabei aus so genannten Flakes zusammengestellt werden. Ein Flake ist ein Informations- und Interaktionsmodul, mit dem Informationen und Funktionalität anderer Web2.0-Dienste angeboten werden. Ziel dieses Domain-Kriteriums ist es, User zu selektieren, die Personalisierung aktiv nutzen. Bis zum 17.04.2006 wurden 3238 Delicious-User in dieser Gruppe identifiziert.

3. Spiegel Domain-Kriterium

Der Spiegel⁴⁰, das bekannte deutsche Print-Magazin, unterhält auch einen News-Service im Internet. Mit diesem Domain-Kriteriums werden deutschsprachige User selektiert. Bis zum 17.04.2006 wurden 6022 Delicious-User in dieser Gruppe identifiziert.

³⁷Ralf Rattay ist als Mitarbeiter von Trommsdorff & Drüner, einer Marketing Firma aus Berlin, mit Projekten im Web2.0-Umfeld betraut. Markt-, Trend- und Zukunftsforschung wird bei Trommsdorff & Drüner praktisch angewendet.

³⁸<http://www.heise.de/>, geprüft 17.04.2007.

³⁹<http://www.pageflakes.com/>, geprüft 17.04.2007.

⁴⁰<http://www.spiegel.de/>, geprüft 17.04.2007.

4. *dict.leo Domain-Kriterium*,

Unter Dict.Leo⁴¹ finden sich umfangreiche Fremdsprache-Deutsch-Wörterbücher. Die verfügbaren Fremdsprachen umfassen Englisch, Französisch und Spanisch. Dabei führt das Englisch-Deutsch-Wörterbuch mit 448601 Einträgen⁴² die Liste an. Als frei nutzbarer Service ist der Dienst so populär, dass März 2007 teilweise mehr als 9 Millionen Anfragen pro Tag an das Englisch-Deutsch Wörterbuch von dict.leo gestellt wurden. Ziel dieses Kriteriums ist die Auswahl von Usern, die sich zwischen den Sprachwelten bewegen. Bis zum 17.04.2006 wurden 2373 Delicious-User in dieser Gruppe identifiziert.

5. *LWN Domain-Kriterium*

LWN⁴³ ist ein Nachrichtendienst in englischer Sprache. Er beschäftigt sich mit den Themen Linux und freier Software. Dabei wendet er sich mit seinen Artikeln direkt an Linux und OpenSource-Spezialisten. Ziel dieses Domain-Kriteriums ist die Auswahl von Linux und OpenSource Experten. Bis zum 17.04.2006 wurden 1665 Delicious-User in dieser Gruppe identifiziert.

6. *Farecast Domain-Kriterium*

Farecast⁴⁴ bietet seinen Usern Vorhersagen zu Preisen von Flugtickets innerhalb der USA. Durch die Beobachtung von Preisschwankungen ist Farecast in der Lage, gute Vorhersagen zu treffen. Ziel dieses Domain-Kriteriums ist es, sehr mobile User auszuwählen. Bis zum 17.04.2006 wurden 5047 Delicious-User in dieser Gruppe identifiziert.

7. *Usability Tag-Kriterium*

Usability fasst Überlegungen zur Verwendbarkeit und auch zur Nützlichkeit von Dingen zusammen. Diese Überlegungen werden in der Regel von Designern und Entwicklern angestellt, aber auch Erfahrungsberichte zur Usability konkreter Produkte fallen in diesen Bereich. Dieses Tag-Kriterium soll User selektieren, die sich mit Usability beschäftigen. Bis zum 17.04.2006 wurden 30877 Delicious-User in dieser Gruppe identifiziert.

8. *Web2.0 Tag-Kriterium*

Das Web2.0 Tag-Kriterium selektiert Nutzer, die wenigstens eines ihrer Lesezeichen mit dem Schlagwort 'web2.0' versehen haben. Ziel dieses Tag-Kriteriums ist es, User zu selektieren, die sich mit Themen aus dem Web2.0-Bereich beschäftigen. Bis zum 17.04.2006 wurden 102774 Delicious-User in dieser Gruppe identifiziert.

⁴¹<http://dict.leo.org/>, geprüft 17.04.2007.

⁴²Stand 17.04.2007.

⁴³<http://lwn.net> geprüft, 17.04.2007.

⁴⁴<http://www.farecast.com/> geprüft 17.04.2007

9. *Personalization Tag-Kriterium*

Personalization als Tag-Kriterium wählt Nutzer aus, die sich mit Personalisierung befassen. Personalisierung ist dabei als ein Vorgang zu verstehen bei dem ein Produkt von einem Nutzer auf seine speziellen Bedürfnisse angepasst wird. Durch Marktsättigung und die neuen technischen Möglichkeiten funktioniert das klassische Schema der Produktion, in dem eine Firma ein Produkt herstellt und es in einer Konfiguration anbietet, immer schlechter. Verbraucher können und wollen zwischen Alternativen wählen. Eine Möglichkeit, die Firmen gehen, um den Verbraucher entgegen zu kommen, ist Produkte personalisierbar zu machen. Bis zum 17.04.2006 wurden nur 896 Delicious-User in dieser Gruppe identifiziert.

10. *Spotback Domain-Kriterium*

Spotback⁴⁵ ist ein News-Service mit Personalisierung. Dabei wird durch Userbewertungen von Nachrichten ein persönliches Profil erzeugt. Durch dieses Profil ist Spotback in der Lage, seinen Usern personalisierte Nachrichten zusammenzustellen. Ziel dieses Domain-Kriteriums ist, es User zu selektieren, die Personalisierungen nutzen. Bis zum 17.04.2006 wurden 653 Delicious-User in dieser Gruppe identifiziert.

Alle diese Kriterien erzeugen entsprechende Gruppen von Delicious-Usern. Die hier beschriebenen Gruppen wurden durch Scannen der Bookmarks gebildet. Das letzte untersuchte Lesezeichen trug dabei die Nummer 45.582.500. Die Bookmarks der Gruppenmitglieder wurden im Vorverarbeitungsschritt (vgl. Abschnitt 5.2.3, S.64) aufgearbeitet und in die Auswertungstabellen eingetragen. Im nächsten Abschnitt werden Ergebnisse betrachtet. Dabei werden beispielhaft ein Zeitabschnitt gewählt und die in den Gruppen am häufigsten gespeicherten URLs extrahiert.

6.3 Einfache Auswertung

Unter einer einfachen Auswertung verstehe ich hier das Abrufen von Top- k Listen der Bookmarkingereignisse über frei gewählte Zeitintervalle. Ziel der einfachen Auswertung ist es, an dieser Stelle zum einen die in 6.2 definierten Kriterien zu testen und zu vergleichen und zum anderen die Funktionsweise des Systems zu demonstrieren. Wichtig bei der Betrachtung dieser Auswertung ist es, sich darüber im Klaren zu sein, dass es nicht um User-Bookmarkbeziehungen geht, sondern dass nur neue Bookmarks betrachtet werden. Die dargestellten Top-Listen enthalten Rangfolgen von Bookmarkingereignissen. Jede Useraktion, die ein neues Lesezeichen im Bereich des Users ablegt ist ein solches Bookmarkingereignis und Basis der folgenden Ranglisten.

Folgend werden zu allen definierten Beispielkriterien Top-10 Listen dargestellt. Als Auswertungszeitraum wurde für alle Gruppen oder Kriterien ein identischer 60 Tage-Intervall gewählt, der mit dem 08.01.2007 beginnt. Durch die Wahl eines identischen Intervalls für die Top-10 Listen aller Gruppen werden die Ergebnisse vergleichbar. Damit

⁴⁵<http://spotback.com> geprüft, 17.04.2007.

lassen sich Aussagen zur Leistungsfähigkeit der gewählten Kriterien und des Systems machen. Die in den nachfolgenden Auswertungstabellen gezeigten URLs wurden am 18.04.2007 geprüft.

Zehn Top-10 Listen zu den zehn definierten Gruppen ergeben 100 URLs. Eine komplette Diskussion jeder einzelnen URL ist nicht sinnvoll. Um trotzdem in der Lage zu sein, die Ergebnisse zu bewerten und einzuschätzen, werden zu jeder Top-10 Liste kurz interessante URLs und ihr Zusammenhang zu den Zielen des Kriteriums betrachtet.

| Nr. | Gruppe | Ziele |
|-----|-----------------|---|
| 1 | Spiegel | deutschsprachige User |
| 2 | Heise | deutschsprachige User mit IT-Interesse |
| 3 | dict.leo | User zwischen Sprachenwelten |
| 4 | Farecast | mobile User, z.B. Vielflieger |
| 5 | Pageflakes | Personalisierung |
| 6 | Spotback | Personalisierung |
| 7 | Personalization | Personalisierung |
| 8 | Usability | Verwendbarkeit und Nutzlichkeit von Dingen |
| 9 | Web2.0 | User mit Web2.0-Interessen |
| 10 | LWN | Linux und OpenSource-Enthusiasten und -Experten |

Tabelle 6: Beispielgruppen und Ziele

Die Tabelle 6 fasst die in 6.2 gewählten Beispielgruppen noch einmal nach Themen geordnet zusammen. Folgend werden die Ergebnisse der Auswertung, also die Top-10 Listen in der gleichen Reihenfolge betrachtet.

Wichtiger Hinweis! (20.04.2007)

Die Daten der Top-10 Listen der Web2.0- und Usability-Gruppen können Fehler beinhalten. Kurz vor Ende der Arbeit habe ich Fehler in den Top-Listen festgestellt. Sie rühren aus einer fehlerhaften Implementation des Vorverarbeitungsschritts. Durch einen Datentypfehler in der Stored Procedure zum Einfügen der Daten in die Analysetabellen wurden Intervallnummern falsch berechnet. Alle Vorberechnungsintervalle ab zwei Tagen Intervalllänge enthalten fehlerhafte Ereigniszähler. Nach Korrektur des Fehlers konnten für acht von zehn Gruppen die Vorberechnungen erneut durchgeführt werden. Aufgrund der Datenmengen war der Vorberechnungsschritt für die Web2.0- und Usability-Gruppen in verbleibender Zeit nicht wiederholbar.

An der Funktionsweise des Systems ergeben sich dadurch keine Änderungen. Die Daten in den Ein-Tages-Intervallen sind für alle Gruppen korrekt. Die fehlerhaften Daten aus dem Vorberechnungsschritt führen nur zu einer Verfälschung der Rangfolgen.

Die Spiegel-Gruppe

| | Σ | URL |
|----|----------|---|
| 1 | 468 | http://www.spiegel.de/ |
| 2 | 191 | http://www.smashingmagazine.com/2007/01/19/53-css-techniques-you-couldnt-live-without/ |
| 3 | 190 | http://pipes.yahoo.com/ |
| 4 | 168 | http://www.visual-literacy.org/periodic_table/periodic_table.html |
| 5 | 125 | http://www.wisegeek.com/what-does-200-calories-look-like.htm |
| 6 | 119 | http://www.apple.com/hotnews/thoughtsonmusic/ |
| 7 | 102 | http://www.spiegel.de/netzwelt/tech/0,1518,457359,00.html |
| 8 | 93 | http://www.youtube.com/watch?v=6gmP4nk0E0E |
| 9 | 93 | http://www.eliteskills.com/free_education/?foo=x |
| 10 | 92 | http://www.getafirstlife.com/ |

Tabelle 7: Gruppe 1: **Spiegel** Domain-Kriterium Top-10 über 60 Tage ab dem 02.01.2007

Die Spiegel-Gruppe sollte deutschsprachige User und ihre Bookmarks erfassen. Die entsprechenden URLs sind als Top-10 Liste in Tabelle 7 dargestellt.

Bis auf die Verweise auf Spiegel-Online Seiten auf den Positionen eins und sieben finden sich keine weiteren deutschsprachigen Seiten in der Top-10 Liste. Bemerkenswert ist aber ein Trend zum Visuellen. Position vier gibt eine gute Übersicht zu Visualisierungsmethoden, an Position zwei werden 200 Kalorien-Mahlzeiten mit vielen Fotos dargestellt. Weiterhin zum visuellen Thema passt der Spiegelartikel zur HDR-Photographie auf Position sieben. An Position sechs findet sich ein Verweis auf Apple News, in dem Apple ankündigt, Musik auch ohne DRM-Beschränkungen zu verkaufen. GetaFirstLife auf Position sieben ist eine Parodie auf den SecondLife-Hype. Auf Position acht ist ein sehr interessantes Video zu den Möglichkeiten des Web2.0 verlinkt.

Betrachtet man die Ergebnisse im Auswertungszeitraum, kann man schlussfolgern, dass das Spiegel-Kriterium nicht besonders gut geeignet ist, um deutschsprachige User auszuwählen. Es fanden sich zwar auch deutsche Webseiten, es handelte sich bei ihnen aber nur um Spiegel Artikel. Aus den Ergebnissen der Top-10 Liste lässt sich folgern, dass die ausgewählten User über gute Englischkenntnisse verfügen.

Die Heise-Gruppe

Die Heise-Gruppe sollte deutschsprachige User zusammenfassen die sich für IT-Themen interessieren. In Tabelle 8 findet sich die Top-10 Liste der URLs dieser Gruppen im Auswertungszeitraum.

Der Hauptteil der gelisteten URLs beschäftigt sich mit Themen, die für Administratoren relevant sind. Die Positionen eins, sieben und zehn beschäftigen sich mit Webdesign-techniken. Auf Position fünf findet sich wieder der Verweis auf die Visualisierungsmethoden, der auch schon in den Ergebnissen der Spiegel-Gruppe zu finden war. Interessant ist auch die URL an Position drei. Es handelt sich dabei um einen englischsprachigen Artikel aus dem Jahr 1999, der sich mit dem in Windowssysteme eingebauten Zugriffscode der

| | Σ | URL |
|----|----------|---|
| 1 | 265 | http://www.smashingmagazine.com/2007/01/19/53-css-techniques-you-couldnt-live-without/ |
| 2 | 248 | http://pipes.yahoo.com/ |
| 3 | 215 | http://www.heise.de/tp/r4/artikel/5/5263/1.html |
| 4 | 197 | http://www.heise.de/ |
| 5 | 188 | http://www.virtualbox.org/ |
| 6 | 136 | http://www.visual-literacy.org/periodic_table/periodic_table.html |
| 7 | 110 | http://www.miniajax.com/ |
| 8 | 106 | http://www.spiegel.de/ |
| 9 | 103 | http://www.kroah.com/lkn/ |
| 10 | 101 | http://www.smashingmagazine.com/2007/01/19/53-css-techniques-you-couldnt-live-without// |

Tabelle 8: Gruppe 2: **Heise** Domain-Kriterium Top-10 über 60 Tage ab dem 02.01.2007

NSA⁴⁶ beschäftigt. Das Gruppenkriterium Heise findet sich auf der Position vier. Die Yahoo-Pipes auf Position zwei sind eine graphische Umsetzung den Pipes-Konzeptes, wie es von Unix und Linux Systemen bekannt ist. Mit diesem Yahoo-Dienst lassen sich kleine Programme aus graphischen Bausteinen zusammensetzen. Auf Position neun findet sich der Verweis auf die Webseite eines Buchs zum Linux Kernel. Auffällig ist die Unterscheidung zwischen den URLs auf Position eins und Position zehn. Ein Vergleich der URLs zeigt, dass auf Position zehn die URL keinen abschließenden Schrägstrich trägt. Diese Verweise zeigen ein Problem des vorgestellten Auswertungssystems. Eine Webseite kann über verschiedene URLs erreichbar sein, diese URLs können sich im Domainnamen, aber auch nur in Parametern oder abschließendem Schrägstrich unterscheiden. Jede dieser Seiten taucht dann einzeln in der Datenbank und dann evtl. auch in den Top-Listen auf.

Die Ergebnisse der Heise-Gruppen hinterlassen einen für das Heise-Kriterium erfolgreichen Eindruck. Die User dieser Gruppe haben im Auswertungszeitraum die Heise- und die Spiegel-Webseite gespeichert und eine ganze Reihe von Verweisen auf Webseiten, die sich mit IT-Themen beschäftigen, angelegt

Die dict.leo-Gruppe

Mit der dict.leo-Gruppe sollten User ausgewählt werden, die sich zwischen Sprachenwelten bewegen. In Tabelle 9 ist die Top-10 Liste der Bookmarks dieser Gruppe im Auswertungszeitraum dargestellt.

Mit dieser Gruppe sind deutschsprachige Nutzer ausgewählt worden. In der Top-10 finden sich vier deutschsprachige Webseiten. Auf Position fünf erscheint eine englischsprachige Webseite zum Webdesign, die auch schon in der Heise-Gruppe auf Position eins zu sehen war. Die Spiegel-URL auf Position vier zeigt eine teilweise Überschneidung

⁴⁶National Security Agency. Ein Nachrichtendienst der USA, der für Überwachung und Entschlüsselung elektronischer Kommunikation zuständig ist.

| | Σ | URL |
|----|----------|---|
| 1 | 438 | http://dict.leo.org/ |
| 2 | 83 | http://de.wikipedia.org/wiki/Hauptseite |
| 3 | 80 | http://dict.leo.org/?lang=de&lp=ende |
| 4 | 73 | http://www.spiegel.de/ |
| 5 | 51 | http://www.smashingmagazine.com/2007/01/19/53-css-techniques-you-couldnt-live-without/ |
| 6 | 48 | http://www.pandora.com/ |
| 7 | 48 | http://www.flickr.com/ |
| 8 | 39 | http://dict.leo.org/?lp=ende&lang=de&agent=firefox-de&search=%s |
| 9 | 39 | http://www.youtube.com/ |
| 10 | 36 | http://www.studivz.net/ |

Tabelle 9: Gruppe 3: **dict.leo** Domain-Kriterium Top-10 über 60 Tage ab dem 02.01.2007

der Usergruppen Spiegel und dict.leo. Auf den Positionen sieben, neun und zehn finden sich Verweise auf die Startseiten sehr bekannter Web2.0-Anwendungen. Sie legen nahe, dass es sich bei den ausgewählten Usern nicht um Experten, sondern eher um Otto-Normal-Verbraucher handelt. Wieder finden sich Verweise auf identische Webseiten auf mehreren Positionen der Top-10 Liste. Diesem Problem entsprechen die Einträge auf den Positionen eins, drei und acht. Sie verweisen auf die Webseite des Gruppen-Kriteriums, das Deutsch-Englisch Wörterbuch von dict.leo. Ein weiteres Online-Wörterbuch oder weitere Übersetzungswerkzeuge tauchen im Beispielintervall nicht auf.

Das dict.leo-Kriterium scheint sein Ziel nicht so recht zu erfüllen. Weitere Sprach- und Übersetzungswebseiten, die zu erwarten gewesen wären, tauchten im Auswertungszeitraum nicht in der Top-10 Liste auf. Das dict.leo Kriterium liefert aber gute Ergebnisse, wenn man einfache User auswählen und beobachten möchte.

Die Farecast Gruppe

In der Farecast-Gruppe sollten mobile User erfasst werden. Die Top-10 Liste dieser Gruppe im Auswertungszeitraum ist in Tabelle 10 dargestellt.

An Position eins findet sich der Verweis auf die Webseite des Domain-Kriteriums Farecast. Auf Rang vier findet sich der Priceprotectr, ein Dienst, mit dem User Produktpreise überwachen können. Price Protection ist dabei die Zusicherung eines Verkäufers, zukünftige Preisnachlässe innerhalb einer bestimmten Frist an derzeitige Käufer zu erstatten. Der Webservice ist ein Überwachungsdienst, mit dem User vom Aufwand befreit werden, selbst die Preise der von ihnen erstandenen Artikel zu prüfen. An Position sieben findet sich mit Kayak ein Dienst, der den Farecast-Service in Deutschland anbietet. Auch der Verweis von Position neun beschäftigt sich mit Methoden, Flugtickets günstig zu erwerben. Keinen offensichtlichen Bezug zum Zielthema scheinen die Verweise der Positionen zwei, drei, fünf und zehn zu haben. Auf Position acht findet sich eine Investitionsanleitung mit einem unbeabsichtigtem Bezug zum Kriterium der Gruppe. Es sollten mobile User ausgewählt werden, zusätzlich scheinen diese User aber auch Sparer zu sein. Indiz

| | Σ | URL |
|----|----------|---|
| 1 | 1162 | http://www.farecast.com/ |
| 2 | 290 | http://pipes.yahoo.com/ |
| 3 | 264 | http://www.eliteskills.com/free_education/?foo=x |
| 4 | 262 | http://www.priceprotectr.com/ |
| 5 | 240 | http://www.visual-literacy.org/periodic_table/periodic_table.html |
| 6 | 215 | http://www.smashingmagazine.com/2007/01/19/53-css-techniques-you-couldnt-live-without/ |
| 7 | 187 | http://www.kayak.com/ |
| 8 | 179 | http://www.morningstar.com/Cover/Classroom.html |
| 9 | 173 | http://blog.auiinteractive.com/how-to-get-the-cheapest-flight-every-single-time |
| 10 | 170 | http://lesliefranke.com/files/reference/csscheatsheet.html |

Tabelle 10: Gruppe 4: **Farecast** Domain-Kriterium Top-10 über 60 Tage ab dem 02.01.2007

dafür sind die Verweise auf den Priceprotectr und diese Investitionsanleitung.

Das Farecast-Kriterium liefert gute Resultate. Geht man davon aus, dass User, die Webseiten zu Flugpreisen speichern, auch häufig fliegen, dann sind mobile User ausgewählt worden. Es zeigt sich in den Ergebnissen dieser Gruppe ein weiteres Thema, zu dem diese Ergebnisse noch besser passen. Das Farecast-Kriterium wählt preisbewusste User aus die auf Schnäppchen und Angebote achten und sich auch für Investitionsstrategien interessieren.

Die Pageflakes-Gruppe

| | Σ | URL |
|----|----------|---|
| 1 | 664 | http://www.pageflakes.com/ |
| 2 | 257 | http://www.netvibes.com/ |
| 3 | 194 | http://pipes.yahoo.com/ |
| 4 | 154 | http://www.smashingmagazine.com/2007/01/19/53-css-techniques-you-couldnt-live-without/ |
| 5 | 129 | http://www.pageflakes.com/Default.aspx |
| 6 | 106 | http://www.go2web20.net/ |
| 7 | 100 | http://www.miniajax.com/ |
| 8 | 91 | http://www.bubbl.us/ |
| 9 | 87 | http://www.eliteskills.com/free_education/?foo=x |
| 10 | 83 | http://www.readwriteweb.com/archives/top_100_alternative_search_engines.php |

Tabelle 11: Gruppe 5: **Pageflakes** Domain-Kriterium Top-10 über 60 Tage ab dem 02.01.2007

Die Pageflakes-Gruppe sollte User erfassen, die Personalisierung nutzen. Die Top-10 Liste der Auswertung dieser Gruppe befindet sich in Tabelle 11.

Pageflakes selbst steht dort an Position eins und fünf. An Position zwei findet sich der Verweis auf Netvibes, Usern wird dort ähnlich wie bei Pageflakes eine personalisierte Startseite angeboten. An den Positionen vier, neun und zehn tauchen URLs auf, die schon in vorher beschriebenen Top-Listen anderer Gruppen zu sehen waren und keinen erkennbaren Bezug zum Zielthema haben. Mit den Codeschnipseln von MiniAjax an Position 7 können User eigene Webseiten personalisieren. An Position zehn findet sich der Verweis auf eine Webseite, die alternative Suchmaschinen vorstellt. Auf der Position sechs findet sich mit Go2Web20 ein sehr interessantes Web2.0-Verzeichnis, in dem man auch mittels Schlagwörtern suchen kann. Bei Bubbl an Position acht handelt es sich um ein Webanwendung, mit der Mind-Maps erstellt werden können. Mind-Maps dienen der kreativen Visualisierung und Organisation von Informationen.

Das Pageflakes-Kriterium liefert im Auswertungszeitraum nur einzelne Verweise auf Webseiten, deren Themen einen Bezug zur Personalisierung aufweisen. Ein bei den ausgewählten Usern dominanteres Thema scheinen Web2.0 Anwendungen zu sein.

Die Spotback-Gruppe

| | Σ | URL |
|----|----------|---|
| 1 | 94 | http://spotback.com/ |
| 2 | 72 | http://pipes.yahoo.com/ |
| 3 | 44 | http://www.miniajax.com/ |
| 4 | 44 | http://www.smashingmagazine.com/2007/01/19/53-css-techniques-you-couldnt-live-without/ |
| 5 | 39 | http://www.winaddons.com/top-300-freeware-software/ |
| 6 | 35 | http://www.smashingmagazine.com/2007/02/09/83-beautiful-wordpress-themes-you-probably-havent-seen/ |
| 7 | 33 | http://www.visual-literacy.org/periodic_table/periodic_table.html |
| 8 | 32 | http://www.picnik.com/ |
| 9 | 32 | http://www.wikiseek.com/ |
| 10 | 31 | http://www.bubbl.us/ |

Tabelle 12: Gruppe 6: **Spotback** Domain-Kriterium Top-10 über 60 Tage ab dem 02.01.2007

Auch das Spotback-Kriterium sollte Nutzer erfassen, die sich mit Personalisierung befassen. Die Ergebnisse sind in Tabelle 12 zu sehen.

MiniAjax an der Position drei und Bubbl an der Position zehn fand sich auch in der Top-Liste der Pageflakes-Gruppe. Auf Position zwei finden sich wieder Yahoo-Pipes, die eine graphische Umsetzung des von Unix und Linux bekannten Pipeline-Konzepts im Web-Kontext sind. Auf Position acht findet sich eine Web2.0-Anwendung, mit der User digitale Photos bearbeiten können. Wikiseek auf Position neun ist eine Suchmaschine, die in der Wikipedia sucht. Ein direkter thematischer Bezug der referenzierten Webseiten zur Personalisierung lässt sich kaum feststellen. Abgesehen von Spotback, dem Domain-Kriterium der Gruppe selbst, findet sich noch MiniAjax an Position drei. Mit gutem

Willen lässt sich der Verweis auf die Liste freier Windows-Software von Position fünf noch zur Personalisierung rechnen, da User ihre Windows-Arbeitsumgebung damit an ihre eigenen Vorstellungen anpassen können.

Mit zwei von acht für Personalisierung relevanten Links, ist das Spotback-Kriterium nicht sonderlich zielgenau. Es finden sich aber mit den Yahoo-Pipes, Picnik und Bubbl interessante Web2.0-Dienste in der Top-10 Liste der Spotback-Gruppe.

Die Personalization-Gruppe

| | Σ | URL |
|----|----------|---|
| 1 | 82 | http://www.visual-literacy.org/periodic_table/periodic_table.html |
| 2 | 78 | http://pipes.yahoo.com/ |
| 3 | 43 | http://www.youtube.com/watch?v=6gmP4nk0E0E |
| 4 | 43 | http://www.smashingmagazine.com/2007/01/19/53-css-techniques-you-couldnt-live-without/ |
| 5 | 38 | http://www.apple.com/hotnews/thoughtsonmusic/ |
| 6 | 38 | http://www.readwriteweb.com/archives/top_100_alternative_search_engines.php |
| 7 | 37 | http://www.readwriteweb.com/archives/online_video_index.php |
| 8 | 36 | http://blog.guykawasaki.com/2007/01/ten_ways_to_use.html |
| 9 | 34 | http://www.readwriteweb.com/archives/recommendation_engines.php |
| 10 | 32 | http://radar.oreilly.com/archives/2007/02/pipes_and_filte.html |

Tabelle 13: Gruppe 7: **Personalization** Tag-Kriterium Top-10 über 60 Tage ab dem 02.01.2007

Für die Personalisierungsgruppe wurde ein Tag-Kriterium zur Auswahl der User verwendet. Auch sie soll User erfassen, die sich mit Personalisierung beschäftigen. Die Ergebnisse im Auswertungszeitraum sind in Tabelle 13 zu sehen.

Wieder finden sich URLs, die schon in anderen Listen auftauchten und keinen erkennbaren direkten Bezug zum Thema haben. Dazu zählen die Verweise an Position eins, acht und neun. Die Verweise von Position zwei und zehn beschäftigen sich mit den Yahoo-Pipes, sie können von Usern genutzt werden, um sich personalisierte Internet-Newsfeeds zu erstellen. In der Top-Liste finden sich drei Artikel des Read/WriteWeb-Blogs, einem der meist gelesenen Web-Blogs weltweit,⁴⁷ in den Web-Technologien diskutiert werden. An Position acht wird die Nützlichkeit von LinkedIn, einer bekannten Social-Networking Anwendung, diskutiert.

Das Personalization-Tag-Kriterium scheint nicht besonders erfolgreich zu sein. Der große Teil der URLs im Auswertungszeitraum beschäftigt sich mit Web2.0-Themen. Der Mangel an Webseiten, die sich direkt mit Personalisierung beschäftigen, kann verschiedene Gründe haben. Personalisierung ist wahrscheinlich einfach ein schwaches Thema bei diesen Usern, so dass es vom Web2.0-Thema dominiert wird. Eine andere Erklärung

⁴⁷vgl. <http://www.technorati.com/pop/blogs/>. Am 28.04.2007 befand sich das Read/WriteWeb-Blog auf Position 36 der populärsten Blogs.

könnte ein starker Zusammenhang von Web2.0 und Personalisierung sein. Aus den gefundenen Ergebnissen lässt sich ohne weitere Analysen keine definitive Aussage treffen, warum genau die Links in Bezug auf das Zielthema wenig relevant sind.

Die Usability-Gruppe

Daten fehlerbehaftet.

| | Σ | URL |
|----|----------|---|
| 1 | 2283 | http://lesliefranke.com/files/reference/csscheatsheet.html |
| 2 | 2073 | http://www.smashingmagazine.com/2007/01/19/53-css-techniques-you-couldnt-live-without/ |
| 3 | 1791 | http://headrush.typepad.com/creating_passionate_users/2006/12/dont_make_the_d.html |
| 4 | 1738 | http://web2.wsj2.com/the_best_web_20_software_of_2006.htm |
| 5 | 1657 | http://www.visual-literacy.org/periodic_table/periodic_table.html |
| 6 | 1504 | http://www.useit.com/alertbox/9605.html |
| 7 | 1456 | http://www.humblelittlerubybook.com/book/ |
| 8 | 1428 | http://www.fleck.com/ |
| 9 | 1427 | http://www.glennferon.com/portfolio1/ |
| 10 | 1418 | http://www.hakia.com/ |

Tabelle 14: Gruppe 8: **Usability** Tag-Kriterium Top-10 über 60 Tage ab dem 02.01.2007

Die Usability-Gruppe sollte User auswählen, die sich mit der Verwendbarkeit und Nützlichkeit von Dingen beschäftigen. In Tabelle 14 sind die Ergebnisse der Auswertung im gewählten Intervall dargestellt.

Die Webseite der Position sechs hat einen direkten Bezug zur Usability von Webseiten, dort werden zehn typische Designfehler und ihre Folgen erläutert. Auch der Verweis von Position drei beschäftigt sich mit Usability und diskutiert den Zusammenhang zwischen gefühltem Projektstatus und dem Eindruck, den die Nutzeroberfläche einer Software hinterlässt. Interessant ist Position acht. Fleck gibt seinen Usern die Möglichkeit, beliebige Webseiten mit eigenen Notizen zu versehen, die sie dann auch Freunden und Bekannten zur Verfügung stellen können⁴⁸. Auch der in anderen Top-Listen schon aufgetauchte Verweis von Position fünf zu generellen graphischen Darstellungsmöglichkeiten hat einen Bezug zum Thema Usability. Position vier verweist auf ein sehr interessantes Blog Post, das sich mit einem Ranking von Web2.0-Software das Jahres 2006 beschäftigt. Auf Position zehn findet sich eine Suchmaschine, die versucht, aus den gesammelten Webseiten Bedeutungen zu extrahieren und diese für User suchbar zu machen.

In der Ergebnisliste dieses Kriteriums erscheinen im Auswertungszeitraum einige Verweise auf Webseiten, die sich offensichtlich mit Usability befassen. Es finden sich aber

⁴⁸Unter <http://extension.fleck.com/?sh=55a39e036a21588d39fe01651c73327eb819886c> findet sich ein Demonstration (geprüft 19.04.2007).

auch wieder einige Verweise zum Web2.0-Thema. Aussagen zur Güte des Kriteriums sind in dieser Gruppe schwierig, da die verwendeten Daten fehlerbehaftet waren und die in der Ergebnisliste auftauchenden URLs nicht notwendigerweise die stärksten aktuellen Themen der ausgewählten User repräsentieren.

Die Web2.0-Gruppe

Daten fehlerbehaftet.

| | Σ | URL |
|----|----------|---|
| 1 | 3950 | http://www.smashingmagazine.com/2007/01/19/53-css-techniques-you-couldnt-live-without/ |
| 2 | 3290 | http://www.miniajax.com/ |
| 3 | 3265 | http://lesliefranke.com/files/reference/csscheatsheet.html |
| 4 | 2600 | http://www.visual-literacy.org/periodic_table/periodic_table.html |
| 5 | 2540 | http://www.webdesignfromscratch.com/web-2.0-design-style-guide.cfm |
| 6 | 2079 | http://web2magazine.blogspot.com/2007/01/thanks-for-web-2.html |
| 7 | 1997 | http://www.eliteskills.com/free_education/?foo=x |
| 8 | 1685 | http://www.wisegeek.com/what-does-200-calories-look-like.htm |
| 9 | 1602 | http://bhandler.spaces.live.com/blog/cns!70F64BC910C9F7F3!1231.entry |
| 10 | 1543 | http://www.divshare.com/ |

Tabelle 15: Gruppe 9: **web2.0** Tag-Kriterium Top-10 über 60 Tage ab dem 02.01.2007

Die Web2.0-Gruppe sollte User auswählen, die sich mit Themen aus dem Web2.0-Bereich befassen. Die Top-10 Liste des Abfragezeitraums ist in Tabelle 15 dargestellt.

Ein großer Teil der gelisteten URLs befasst sich mit Themen zum Design von Webseiten. An den Positionen eins bis vier finden sich dabei Verweise auf Webseiten zu Designthemen, die auch schon in den Top-Listen andere Gruppen auftauchten. Direkt mit Design-Konzepten des Web2.0 befasst sich der Verweis an Position fünf. An Position sechs findet sich der Verweis auf ein Blog-Post mit einer nach Themen geordneten Top-100 Liste von Web2.0-Anwendungen. Eine detailliertere Diskussion der Ergebnisse dieser Gruppe scheint nicht sinnvoll, da die verwendeten Daten durch einen Fehler im Vorberechnungsschritt verunreinigt wurden.

Das Web2.0 Tag-Kriterium scheint recht erfolgreich in der Auswahl von Usern zu sein, die sich mit Web2.0-Themen befassen.

Die LWN-Gruppe

Das Ziel der LWN-Gruppe ist es, Enthusiasten und Experten aus dem Linuxumfeld zu erfassen. In Tabelle 16 sind die Ergebnisse im Auswertungszeitraum abgebildet.

In dieser Rangliste finden sich eine Reihe von Webseiten, die sich mit Linux-Themen befassen. An Position eins steht eine Webseite zu einem Buch zum Linux Kernel. Auf

| | Σ | URL |
|----|----------|---|
| 1 | 143 | http://www.kroah.com/lkn/ |
| 2 | 81 | http://lwn.net/ |
| 3 | 76 | http://www.smashingmagazine.com/2007/01/19/53-css-techniques-you-couldnt-live-without/ |
| 4 | 62 | http://lwn.net/Kernel/LDD3/ |
| 5 | 60 | http://pipes.yahoo.com/ |
| 6 | 58 | http://www.zafar.se/bkz/Articles/ClassicCompScienceTexts |
| 7 | 56 | http://www.virtualbox.org/ |
| 8 | 53 | http://www.visual-literacy.org/periodic_table/periodic_table.html |
| 9 | 51 | http://linuxondesktop.blogspot.com/2007/02/13-things-to-do-immediately-after.html |
| 10 | 49 | http://lwn.net/Articles/222773/ |

Tabelle 16: Gruppe 10: LWN Domain-Kriterium Top-10 über 60 Tage ab dem 02.01.2007

Position vier findet sich ein weiteres Buch zu Linux, dass sich mit Linux Gerätetreibern beschäftigt. Der LWN-Artikel an der Position zehn überprüft anhand des zentralen Linux Kernel Repository die allgemeine Annahme, dass Linux von Hackern in ihrer Freizeit entwickelt wird. Auch die Verweise der Positionen sieben und neun befassen sich mit Linux-Themen. Wieder tauchen schon in anderen Listen gesehene populäre Verweise auf, siehe dazu die Positionen drei, fünf und acht. Die Webseite von Position sechs legt nahe, dass die User dieser Gruppe auch ein generelles Interesse an Themen der Informatik haben. Auf der referenzierten Webseite finden sich Links zu klassischen wissenschaftlichen Artikeln der Informatik.

Damit haben sechs von zehn Verweisen der Top-10 Liste des Auswertungszeitraums einen offensichtlichen Bezug zum Zielthema. Das LWN-Kriterium ist also ein gutes Kriterium, um User mit Linux-Interessen zusammenzufassen.

Zusammenfassung

Aus den dargestellten Ergebnissen lassen sich Schlüsse zu den Eigenschaften des Systems ziehen. Wie in den Top-10 Listen gezeigt wurde, lassen sich mit diesem System interessante Webseiten zu Themen von Interesse finden. Dabei hängt die Relevanz der Ergebnisse stark vom verwendeten Kriterium ab. Wenn mit dem Kriterium User ausgewählt werden, bei deren Interessen das Zielthema als dominantes und aktives Thema eine Rolle spielt, dann finden sich unter den Top-Listen im Auswertungszeitraum Verweise auf relevante Webseiten. Handelt es sich beim Zielthema aber um ein Thema, das eng mit einem dominanten Thema verbunden ist, dann folgen auch aus einem gut gewählten Kriterium kaum relevante Ergebnisse in den Top-Listen. Ein Beispiel für ein dominantes Thema ist der Web2.0 Komplex. In fast allen Ergebnislisten tauchen Links zu diesem Thema auf. Hingegen scheint das Thema der Personalisierung die Delicious-User eher weniger zu beschäftigen. Häufig befassen sich User, die sich für Personalisierung interessieren, auch mit dem Web2.0. Dieser Zusammenhang führt dazu, dass die Ergebnislisten

der Personalisierungsgruppen nur wenige für Personalisierung relevante Links enthalten.

Der Zusammenhang zwischen Kriterium und Zielthema ist nicht immer der beabsichtigte, wie z.B. an den Ergebnissen der Farecast-Gruppe deutlich wird. Es sollten besonders mobile User erfasst werden. Die Auswertung der URLs zeigte aber, dass eher Schnäppchenjäger selektiert wurden, also User die immer nach dem günstigsten Angebot suchen.

In der Betrachtung der Ergebnisse konnten drei systembedingte Probleme festgestellt werden. Ihre Art und Gründe für ihr Auftreten werden folgend kurz erklärt :

Mehrere URLs mit gleichem Ziel

In verschiedenen Gruppen tauchten in den Top-Listen verschiedene URLs auf, die auf die gleiche Webseite verweisen. Besser wäre es, diese URLs zu einer Einheit zusammenzufassen, um so eine realistische Repräsentation der Popularität der Webseite im Zeitabschnitt zu bekommen. Das hier gewählte System identifiziert Webseiten aber über URLs und sieht somit keine Möglichkeiten zur Zusammenfassung von URLs gleicher Ziele vor. Das Problem der mehrdeutigen Bezeichner für ein Ziel ist auch vom Tagging bekannt. Bei Schlagwörtern kommt es durch Pluralformen, Abkürzungen und Schreibungsunterschieden zu unterschiedlichen Zeichen, die auf die gleiche Bedeutung verweisen. Bei den URLs führen alternative Domainnamen und URL-Parameter zum gleichen Effekt.

Irrelevante populäre URLs

Die in den verschiedenen Gruppen am häufigsten gespeicherten Bookmarks überschneiden sich in einigen Teilen. Es sind also neben relevanten URLs auch populäre, aber für diese Gruppe irrelevante URLs in den Top-Listen vorhanden. Die irrelevanten Verweise erscheinen aufgrund der gewählten Vorberechnung. Da alle Lesezeichen eines Users ausgewertet werden, sobald er dem gruppenbildenden Kriterium mit einem Bookmark genügt, sind in der Ergebnismenge URLs zu mehreren Themen vorhanden. Somit werden in den Gruppen auch Links anderer Themen aggregiert und in die Rangfolge eingeordnet.

Schwache Themen

Wenn das Zielthema bei den ausgewählten Usern eine untergeordnete Rolle spielt, lassen sich mit dem beschriebenen Auswertungssystem kaum relevante URLs in den Top-Listen finden.

In diesem Abschnitt wurde die Möglichkeit zur einfachen Auswertung anhand eines Beispielintervalls und der in Abschnitt 6.2 definierten Usergruppen gezeigt. Die Ergebnislisten enthielten Top-10 Ranglisten der populärsten URLs aus den verschiedenen Gruppen über einen 60-Tage-Intervall vom 02.01.2007. Damit sind die Möglichkeiten zur Auswertung von Bookmarkingereignissen mit dem vorgestellten System noch nicht abgeschlossen. Im nächsten Abschnitt werden die Auswertungsmöglichkeiten erweitert, um Bookmarkingereignisse graphisch darstellen zu können.

6.4 Graphische Auswertung

Der letzte Abschnitt zeigte eine einfache Auswertungsmöglichkeit, die Top-Listen von Bookmarkingereignissen über frei wählbare Zeitabschnitte erstellt. Hier wird eine Erweiterung des Auswertungssystems beschrieben, die eine graphische Darstellung von Bookmarkingereignissen möglich macht. Bookmarkingereignisse werden damit in ihrer Entwicklung visualisierbar.

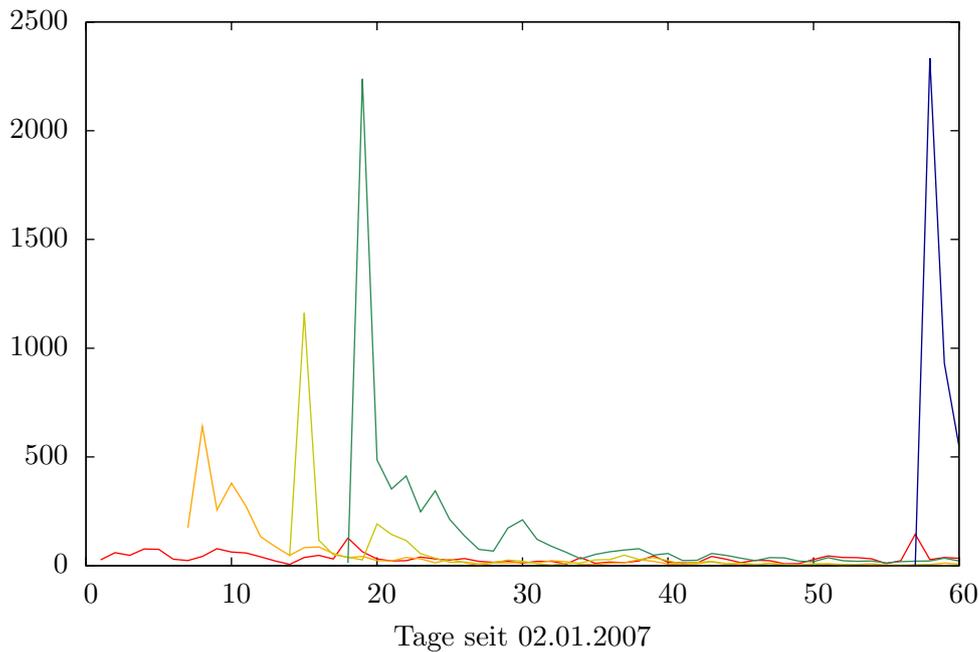
Der Vorverarbeitungsschritt des Auswertungssystem aggregiert Bookmarkingereignisse und speichert diese in der Datenbank, um ein performantes Ranking über beliebige Auswertungsintervalle, zum Abfragezeitpunkt sicher zu stellen. Die in den Analysetabellen gespeicherten Daten können auch für eine graphische Auswertung genutzt werden. Ziel dabei ist es, interessante Entwicklungen zu visualisieren.

Als Parameter der Auswertung stehen zum einen die definierten Beispielgruppen, zum anderen der frei wählbare Auswertungsintervall zur Verfügung. Zur Visualisierung von Bookmarkingereignissen über Zeiträume bieten sich Diagramme an. In ihnen lässt sich aber nur eine begrenzte Zahl von Zahlenreihen sinnvoll darstellen. Zur Darstellung muss also eine beschränkte Anzahl von URLs ausgewählt werden. Diese Auswahl soll hier automatisiert getroffen werden. Eine automatische Wahl von Kandidaten zur graphischen Auswertung kann über verschiedene Mechanismen realisiert werden. Folgend werden zwei Methoden zur Kandidatenauswahl beispielhaft vorgestellt. Die erste Variante beschreibt eine Summen-Rangfolge. Die zweite Möglichkeit zur Kandidatenauswahl benutzt eine Einzelwert-Rangfolge.

Die Summen-Rangfolge

Eine Möglichkeit zur automatisierten Kandidatenauswahl ist die Verwendung des oben beschriebenen Threshold-Algorithmus. Mit ihm lässt sich eine Top- k Liste der URLs über den Gesamtintervall berechnen. Der Threshold-Algorithmus wird hier wie oben für eine effektive Berechnung der Summe von Bookmarkingereignissen über den Gesamtintervall genutzt. Damit ist die entstehende Top- k Liste eine Summen-Rangfolge. Für alle k Elemente dieses Rankings können dann vorberechnete Datenreihen aus den Ein-Tages-Intervallen der Analysetabelle zur Visualisierung genutzt werden.

Das folgende Beispiel verdeutlicht die Effekte dieses Vorgehens. Im Beispiel werden Daten der Web2.0-Gruppe graphisch dargestellt. Um die Darstellung übersichtlich zu halten, wird $k = 5$ gesetzt. Als Auswertungsintervall wurden sechzig Tage ab dem 02.01.2007 gewählt. Die Ergebnisse dieser Abfrage sind in Abbildung 25 dargestellt. In der Legende finden sich neben der Zuordnung von Farben zu URLs auch die Ergebnisse der Summenfunktion. Vier der dargestellten URLs zeigen einen ähnlichen Verlauf der Bookmarkingereignisse. Innerhalb von ein, zwei Tagen werden sie sehr häufig gespeichert. Nach dieser Startphase fallen die Speicherhäufigkeiten fast genau so schnell wieder ab, um sich dann auf niedrigem Niveau zu stabilisieren oder ganz zu verschwinden. Die fünfte, rote Kurve zeigt diese Eigenschaften nicht. Ihre Speicherhäufigkeiten variieren nur geringfügig auf einem Niveau von ca. 40 Speicherungen pro Tag über den gesamten Intervall. Damit ist dieser Web2.0 Design Guide eine Art Dauerbrenner. Jeden



| Legende | |
|---------------------------------------|--|
| — | 2144 http://www.webdesignfromscratch.com /... ^a |
| — | 2861 http://www.visual-literacy.org /... ^b |
| — | 2497 http://www.eliteskills.com /free_education/?foo=x |
| — | 6247 http://www.smashingmagazine.com /... ^c |
| — | 3821 http://www.miniajax.com |

^a... web-2.0-design-style-guide.cfm

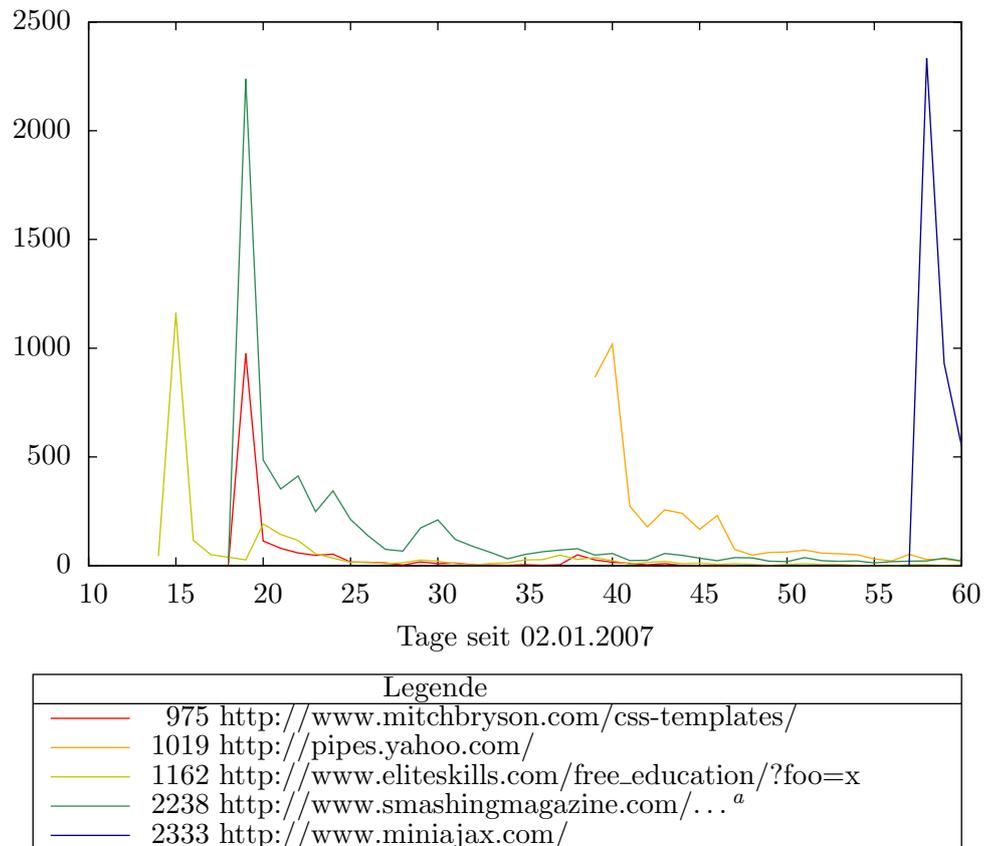
^b... periodic_table/periodic_table.html

^c... 2007/01/19/53-css-techniques-you-couldnt-live-without/

Abbildung 25: Web2.0-Gruppe, Top-5 nach Summen-Rang für 60 Tage ab 02.01.2007

Tag speichert eine ungefähr gleichbleibende Anzahl von Nutzern der Web2.0-Gruppe im Auswertungszeitraum diesen Guide. Auch diese Webseiten können für Markt-, Trend- und Zukunftsforscher interessant sein. Ziel war allerdings das Entdecken von neuen interessanten Bookmarks zum Thema.

Das Auftauchen von Dauerbrennern ist durch die Summenbildung über den Gesamtintervall bedingt. Je länger der dabei gewählte Auswertungsintervall ist, desto größere Chancen haben Dauerbrenner, in die Darstellung aufgenommen zu werden. Soll die Auswahl von Dauerbrennern zur Darstellung verhindert werden, muss ein anderer Auswahlmechanismus genutzt werden.



^a...2007/01/19/53-css-techniques-you-couldnt-live-without/

Abbildung 26: Web2.0-Gruppe, Top-5 nach Einzelwert-Rang für 60 Tage ab 02.01.2007

Die Einzelwert-Rangfolge

Das Auftreten von Dauerbrennern lässt sich durch das Verändern der Auswahlmethode vermeiden. In der zuerst vorgestellten Auswahl durch Summen-Rangfolge kamen Dauerbrenner durch die Summenfunktion, d.h. die Aggregation aller Speicherhäufigkeiten einer URL über den Auswertungsintervall, in die Top-Liste und wurden so zur Darstellung ausgewählt. Wird statt dieser Summenfunktion als Auswahlkriterium ein Rangfolge über die Ereigniszähler der Ein-Tages-Intervalle erstellt, kann die Auswahl von Dauerbrennern vermieden werden.

Diese Rangfolge wird durch Auswerten von Ein-Tages-Intervallen erstellt. Für alle im Auswertungsintervall liegenden Tage werden nacheinander Top- k Listen abgefragt. URLs werden mit ihren Ereigniszählern in einer Datenstruktur vermerkt. Existiert eine URL schon in dieser Struktur, wird ihr Zähler ausgetauscht, wenn er höher als der vorhandene Zähler ist. Nachdem alle Ein-Tages-Intervalle verarbeitet wurden, kann durch Sortieren der Datenstruktur eine Kandidatenliste ohne Dauerbrenner erstellt werden. Damit wurde

eine Einzelwert-Rangfolge über die URLs einer Gruppe im Abfragezeitraum erstellt.

Für ein Beispiel wird wieder die Web2.0-Gruppe verwendet. Der Abfrageintervall ist auf sechzig Tage ab dem 02.01.2007 eingestellt. Abbildung 26 zeigt die Ergebnisse dieser Abfrage. In der Legende finden sich neben den URLs die Werte aus der Rangberechnung, auf deren Basis die URLs zur Darstellung ausgewählt wurden. Der Vergleich mit Abbildung 25 zeigt, dass zwei andere URLs ausgewählt wurden. Der Dauerbrenner ist zwar aus der Darstellung verschwunden, aber auch das Periodensystem visueller Darstellungsmöglichkeiten wurde gegen eine andere URL ausgetauscht.

Folgend wird noch ein weiteres Beispiel der graphischen Auswertung unter Verwendung der Einzelwert-Rangfolge gegeben. Der Abfrageintervall wurde auf 15 Tage verkürzt. Als Startzeitpunkt wurde der 29.03.2007 gewählt. Die Ergebnisse dieser Abfrage sind in Abbildung 27 zu sehen. Die ausgewählten URLs zeigen das schon in Abbildung 25 vorgefundene Muster: Ein schneller Anstieg von Bookmarkingereignissen, der innerhalb von ein bis zwei Tagen wieder stark abfällt, um sich auf niederem Niveau einzupendeln.

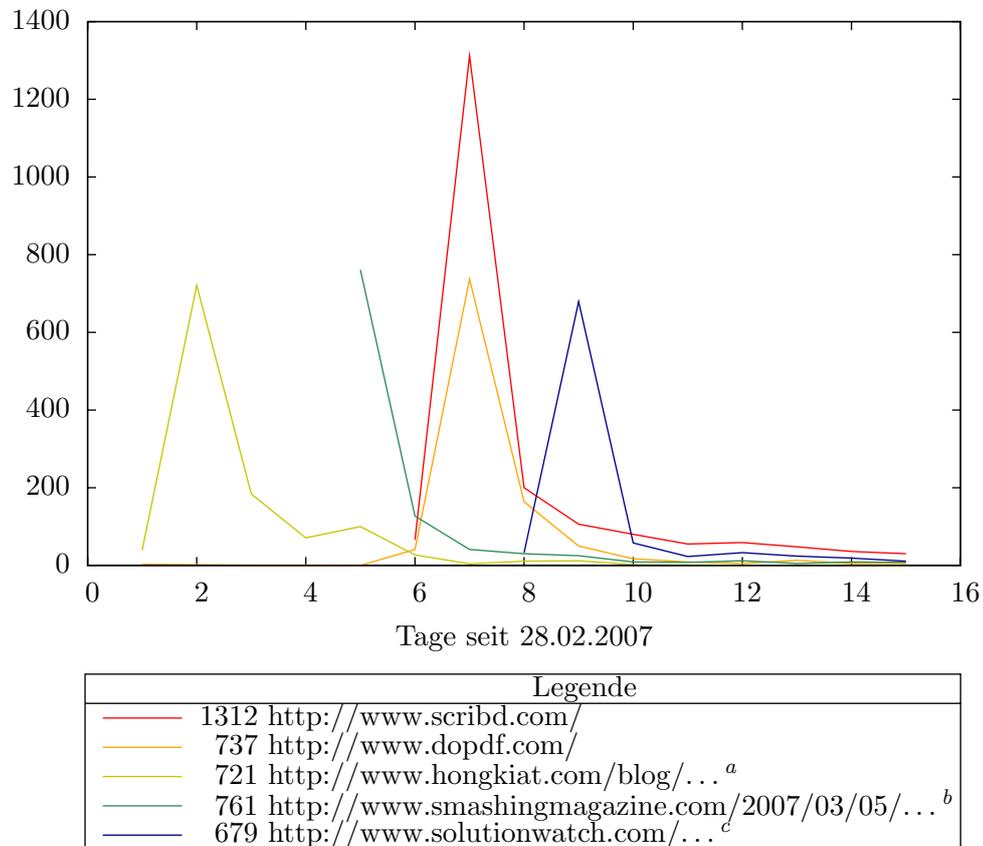
Interessant an den URLs aus der Abbildung 27 ist, dass sich zu den letzten drei in der Legende aufgezählten URLs Erstellungsdaten feststellen lassen. Bei diesen URLs handelt es sich um Verweise auf Onlineartikel, die ein Datum tragen. In der Tabelle 17 sind die Erstellungsdaten der entsprechenden Artikel aufgeführt. Ein Vergleich mit den Speicherkurven der URLs in Abbildung 27 zeigt, dass die Artikel sehr schnell gefunden und gespeichert wurden. Dabei wurde der große Teil der gespeicherten Bookmarks innerhalb von zwei Tagen nach erscheinen des Artikels angelegt. Im Fall der Web Design Übersicht des Smashing Magazins verlief dieser Prozess sogar innerhalb eines Tages, so dass in Abbildung 27 für diese URL kein Anstieg zu erkennen ist. Bedingt durch die Wahl von einem Tag als Basiszeitpunkt des Vorverarbeitungsschritt des Systems, lässt sich nur ein plötzliches Auftauchen und sofortiges Abfallen der Speicherereignisse dieser URL feststellen.

Zusammenfassung

Während die Summen-Rangfolge URLs in absteigender Rangfolge der Fläche unter ihrer Ereigniskurve auswählt, werden bei der Einzelwert-Rangfolge URLs in absteigender Rangfolge ihrer Kurvenspitzen gewählt. Beide Auswahlmethoden haben Vor- und Nachteile. Während der Summen-Rang auch URLs mit niedriger, konstanter Bookmarkingfrequenz zur graphischen Darstellung auswählt, werden beim Einzelwert-Rang auch URLs ausgewählt, die nur sehr kurzfristig relevant sind. Die Anwendung dieser Auswahlmechanismen zeigte im Beispiel einen typischen Verlauf von Bookmarkingereignissen. Dabei steigt die Zahl der Speicherungen innerhalb eines kurzen Zeitraums von ein bis zwei Tagen auf ein Maximum, fällt dann im Zeitraum von ein bis drei Tagen sehr stark ab und stabilisiert sich auf niederem Niveau oder verschwindet ganz.

6.5 Bewertung der Ergebnisse

In den beiden vorhergehenden Abschnitten 6.3 und 6.4 wurden Auswertungsmöglichkeiten von Bookmarkingereignissen gezeigt. Dieser Abschnitt versucht die gefundenen



^a...revealing-methods-of-drawing-web-20-logos/

^b...45-fresh-clean-and-impressive-designs/

^c578/a-roundup-for-developers-developers-developers/

Abbildung 27: Web2.0-Gruppe, Top-5 nach Einzelwert-Rang für 15 Tage ab 28.02.2007

| Datum | URL |
|------------|---|
| 01.03.2007 | http://www.hongkiat.com/blog/revealing-methods-of-drawing-web-20-logos/ |
| 05.03.2007 | http://www.smashingmagazine.com/2007/03/05/45-fresh-clean-and-impressive-designs/ |
| 08.03.2007 | http://www.solutionwatch.com/578/a-roundup-for-developers-developers-developers/ |

Tabelle 17: Erstellungsdaten der Online-Artikel aus Abbildung 27

Ergebnisse und das System zu bewerten. Im ersten Unterabschnitt werden die Fähigkeiten und Grenzen des Systems kritisch betrachtet. Der zweite Unterabschnitt befasst sich mit der Relevanz des Systems für die Markt-, Trend- und Zukunftsforschung.

6.5.1 Kritische Systembetrachtung

In diesem Abschnitt werden die Fähigkeiten und Grenzen des beschriebenen Auswertungssystems einer kritischen Betrachtung unterzogen. Die folgenden Betrachtungen gliedern sich dabei in Überlegungen zur Relevanz der URLs zu den Zielthemen und Betrachtungen zu Performanceeigenschaften des Systems.

Relevanz

Über Tag- und Domain-Kriterien werden im System Gruppen gebildet. Dabei ist eine geschickte Wahl des Kriteriums notwendig, um die User eines Themas von Interesse zu selektieren. Alle User die, das gewählte Kriterium mit mindestens einem Bookmark erfüllen, werden in die Gruppe aufgenommen. Der Vorverarbeitungsschritt aggregiert dann alle Bookmarks von jedem dieser User, um die Auswertung zum Abfragezeitpunkt zu beschleunigen. Damit lassen sich wie in Abschnitt 6.3 gezeigt, relevante Ergebnislisten generieren. Trotzdem tauchen in diesen Listen auch irrelevante URLs auf. Irrelevante URLs werden aus zwei Gründen gelistet :

Schlechtes Kriterium

Ein Kriterium ist dann ein schlechtes Kriterium, wenn es seine Zielvorgaben in der Anwendung nicht erreicht. Kriterien werden verwendet, um User mit bestimmten Interessen auszuwählen. Die Wahl eines Tag- oder Domain-Kriteriums ist auch die Interpretation eines semantischen Zeichens. Wenn der Auswählende⁴⁹ in seiner Interpretation des Kriteriums als semantisches Zeichen nicht mit einem großen Teil der User übereinstimmt, wird das Kriterium sein Ziel nicht erfüllen können. Dann wird die Auswertung der Bookmarkingereignisse dieser Gruppe viele irrelevante Resultate liefern.

Methodische Gründe

Selbst mit sehr guten Kriterien werden irrelevante URLs in den Auswertungslisten auftauchen. Durch die Aggregation aller Bookmarks eines Users im Vorverarbeitungsschritt fließen auch Bookmarks aus anderen Interessensgebieten in die Auswertung ein. Irrelevante URLs tauchen in den Top-Listen der Auswertung auf, wenn sich ein großer Teil der User einer Gruppe in Interessen überschneidet, die, bezogen auf das Zielthema des Kriteriums, irrelevant sind.

Doubletten

Bedingt durch verschiedene URLs mit dem gleichen Ziel wird sowohl die Relevanz der Top-Listen als auch die Einordnung einer Webseite in die Rangfolge durch Du-

⁴⁹Jemand der Kriterien auswählt und Delicious-User zu Gruppen zusammenzufassen.

bletten verschlechtert. Basis des verwendeten Systems sind zwar die URLs, von Interesse sind aber die referenzierten Webseiten. Mit diesem System kann nur eine Annäherung an die ideale Rangfolge der Popularität von Webseiten erreicht werden.

Spam

Eine weitere Gefahr für die Relevanz der URLs in den Auswertungslisten geht von Social Spam aus. Social Bookmarking ist schon zum Ziel von Spammern geworden. Umso mehr User Social Bookmarking Systeme verwenden, desto interessanter werden diese Plattformen für Manipulateure. Sie versuchen die Aufmerksamkeit der User auf in der Regel irrelevante Inhalte zu lenken (vgl. MisterWong 2007). Sind diese Manipulationen erfolgreich, können zusätzliche irrelevante Links in den Auswertungslisten auftauchen. Bisher sind Spammer in Social Bookmarking Plattformen wenig erfolgreich. Einer der Gründe dafür ist sicher auch die Barriere eines User Accounts, der zum Speichern von Bookmarks in der Regel benötigt wird.

Performance

Die Verarbeitung und Analyse der Bookmarking Daten fordert teilweise beachtlichen Aufwand an Rechenzeit. Das beschriebene System wurde zwar auf Verarbeitungsgeschwindigkeit hin optimiert, seiner Skalierbarkeit sind aber Grenzen gesetzt. An dieser Stelle sollen keine theoretischen Betrachtungen des Rechenaufwands angestellt werden. Folgend werden aber ein paar Zusammenhänge zum Laufzeitverhalten und möglichen Grenzen des Systems dargestellt.

Eine interessante Maßzahl ist die Menge neuer Bookmarks pro Zeiteinheit, die das System aufnehmen kann. Würden z.B. alle zwei Millionen Delicious-User aktiv werden und jeden Tag ein oder zwei Bookmarks speichern, dann könnte das Einfügen von neuen Bookmarks ins System zum Engpass werden⁵⁰.

Problematischer als die Anzahl neuer Bookmarks kann die Anzahl der in die Analysetabellen einzutragenden Zähler werden. Durch die Definition einer neuen Gruppe erhöht sich der Aufwand des Vorverarbeitungsschritts. Werden viele Gruppen mit vielen Usern erzeugt, steigen der notwendige Aufwand im Vorverarbeitungsschritt und auch der Platzbedarf der Datenbank sehr schnell an.

6.5.2 Markt-, Trend- und Zukunftsforschung

Das beschriebene System zur automatisierten Auswertung von Social Bookmarks ist ein Prototyp und dient der Demonstration der Verfahrensweise. Ein in der Praxis einsetzbares System könnte sich der gleichen Prinzipien bedienen, um die Methoden der Markt-, Trend- und Zukunftsforschung zu unterstützen.

In Abschnitt 2.5 auf Seite 16 sind die Anforderungen an ein System zur computergestützten Auswertung von User Generated Content mit dem Ziel der Unterstützung von

⁵⁰Zur Zeit werden unter 150.000 Bookmarks täglich auf Delicious gespeichert (vgl. Keller 2007).

Methoden der Markt-, Trend- und Zukunftsforschung formuliert worden. Diese Überlegungen wurden im Abschnitt 5.1 ab Seite 57 weiter ausgearbeitet, um auf ihrer Basis das beschriebene System zu entwickeln.

Diese Anforderungen gliedern die folgende Betrachtung der Ergebnisse und den Versuch einer Einschätzung der Nützlichkeit der mit dem System möglichen Auswertungsmethoden von User Generated Content am Beispiel der Delicious Bookmarks.

Themenvielfalt, Trendführende Daten und Aktualität

Die Themenvielfalt ist eine notwendige Bedingung, um mit dem System eine Vielzahl von Fragen und Interessen befriedigen zu können. Sie wurde durch die Auswahl von Social Bookmarking als auszuwertender User Generated Content sicher gestellt. Besonders interessant sind Delicious Bookmarks auch, weil in Delicious Early Adopters aktiv sind (vgl. Abschnitt 2.2.2 auf Seite 6 und Abbildung 5 auf Seite 8). Die Ergebnisse einer computergestützten Auswertung sind um so wertvoller, desto aktueller die verwendeten Daten sind. Diese Aktualität kann nur durch eine kontinuierliche Datensammlung sicher gestellt werden. Dabei ist mir die Brisanz eines solchen Sammelns von Daten, dem massiven Auslesen von Informationen eines Web2.0-Dienstes, deutlich bewusst. Die hier verwendete Methode kann nur im Kontext akademischer Forschung gerechtfertigt werden. Wollte man z.B. Delicious Bookmarks in vergleichbarem Umfang und kommerziellem Kontext bearbeiten, müsste man mit Delicious, bzw. Yahoo in Verhandlung treten.

Exploration mit thematischem Fokus

Markt-, Trend- und Zukunftsforscher arbeiten themenorientiert. Somit muss ein System zur computergestützten Auswertung und Exploration von User Generated Content auch diesem Themenfokus gerecht werden. Die Verwendung von Kriterien zum Bilden von Usergruppen mit den sich anschließenden Auswertungsmöglichkeiten ist der Versuch diesen, Anforderungen gerecht zu werden. Die exemplarische Auswertung in Abschnitt 6.3 zeigt, dass dieser Ansatz in Bezug auf Zielthemen relevante Resultate bringt und damit diese Bedingung erfüllen kann.

Vorteile gegenüber manueller Auswertung

Um von praktischem Wert für Markt-, Trend- und Zukunftsforscher zu sein, muss ein System computergestützter Auswertung von User Generated Content Vorteile im Vergleich mit Methoden manueller Auswertung bringen. Unter manuellen Auswertungen werden hier Methoden verstanden, bei denen der Markt-, Trend- oder Zukunftsforscher selbst Daten sammelt und analysiert. Im Delicious-System werden den Usern verschiedene Explorationswege angeboten (vgl. Abschnitt 4.3, S.47) bei denen wertvolle Informationen gewonnen werden können. Um mit diesen Methoden zu aussagekräftigen Ergebnissen mit Bezug zu Themen von Interesse zu gelangen, muss ein beachtlicher Aufwand betrieben werden. Das hier vorgestellte Auswertungssystem kann einen großen Teil dieses manuellen Aufwands durch automatisierte Methoden und themenbezogene Aggregation von Bookmarkingergebnissen reduzieren. Relevante Resultate werden mit dem beschriebenen System

jedoch nur für gute Kriterien und dominante Themen erreicht. Verbesserungen des Systems zur Erhöhung der Ausbeute relevanter URLs, insbesondere für schwache Themen sind wünschenswert. Trotzdem hat das hier vorgestellte Auswertungssystem auch in seiner aktuellen Imperfektion einen Anwendungswert für Markt-, Trend- und Zukunftsforscher. Sie können durch dieses System in ihrer Arbeit mit trendführenden und den Zielthemen relevanten, aktuellen Daten unterstützt werden. Sinn des Systems ist dabei die Aggregation von aktuellen Daten und deren Aufbereitung; deren Interpretation und Bewertung bleibt Aufgabe der Markt-, Trend und Zukunftsforscher.

Damit wurde gezeigt, dass das hier beschriebene System Markt-, Trend- und Zukunftsforscher in ihrer Arbeit unterstützen kann. Sie können in dieser Art von User Generated Content, in Daten aus Social Bookmarking Systemen, mit den beschriebenen Auswertungsmethoden relevante Informationen entdecken. Selbst wenn dieses System die Arbeit von Markt-, Trend- und Zukunftsforschern vereinfacht, bleiben genügend Probleme bestehen. Selbst mit einem identifizierten Trend lassen sich häufig nur schlechte Prognosen seiner weiteren Entwicklung abgeben. Der nächste Abschnitt beschäftigt sich mit diesem Problem und versucht kurz Gründe zu benennen, warum Trendforschung fehlschlagen kann.

6.5.3 Die Kluft

Das frühzeitige Erkennen interessanter neuer Produkte und Ideen ist schwierig. Das beschriebene System kann Markt-, Trend- und Zukunftsforscher bei dieser Aufgabe unterstützen. Selbst wenn Trends, Ideen oder vielversprechende Produkte identifiziert werden können, lassen sich aufgrund ihrer momentanen Popularität allein kaum Aussagen zum weiteren Verlauf ihrer Verbreitung in der Gesellschaft machen.

Neue Produkte oder Ideen breiten sich nach Rogers (1995) in Form einer Glockenkurve in einer Population aus (vgl. Abbildung 3 auf Seite 7). Bei Betrachtung dieser Kurve ist aber zu bedenken, dass sie den Fall einer erfolgreichen Annahme und Diffusion darstellt. Viele Ideen und Produkte erreichen aber nie den Massenmarkt. Moore (2006) zeigt Risse in der Diffusionskurve auf. Jeder dieser Risse ist eine Gefahr für einen weiteren positiven Verlauf der Diffusion des Produkts oder der Idee. Einer dieser Risse ist dabei besonders groß, Moore nennt ihn „Die Kluft“ (vgl. Abbildung 28). Sie tut sich zwischen dem Markt der Early Adopters und der Frühen Mehrheit auf. Den Grund für diese Kluft sieht Moore in den sich stark unterscheidenden Einstellungen von Early Adopters und Früher Mehrheit zu nicht-kontinuierlichen Innovationen (engl. 'disruptive innovations'). Early Adopters können durch ihre Begeisterungsfähigkeit für Neues über Hemmnisse und Schwierigkeiten der Innovation hinwegsehen und scheuen keinen Aufwand, um in den Genuss ihrer Vorteile zu kommen. Die Frühe Mehrheit dagegen ist durch praktische Überlegungen der Machbarkeit und des Aufwands motiviert. Nicht-kontinuierliche Innovationen haben es sehr schwer sich in ihrem Markt durchzusetzen.

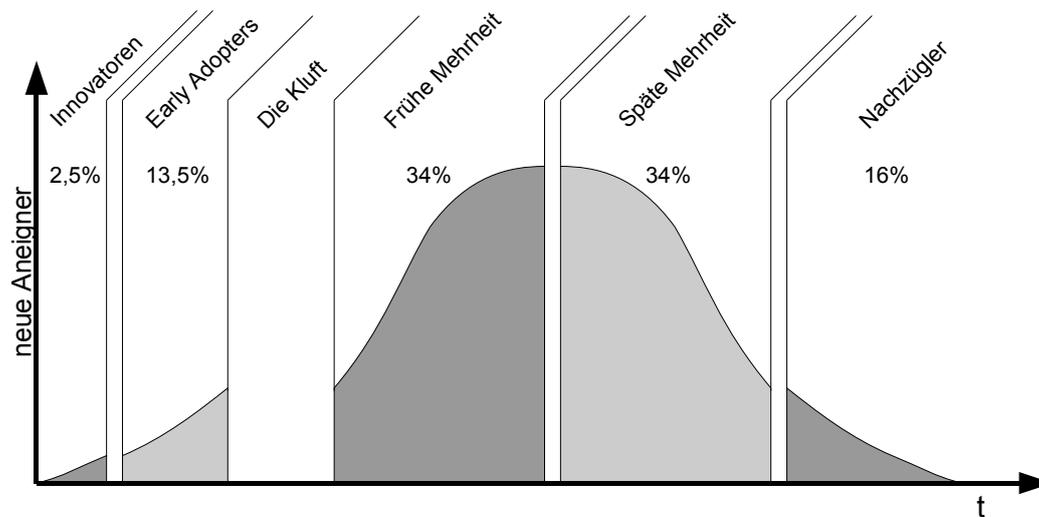


Abbildung 28: Die Kluft nach Moore (2006)

6.6 Zulässigkeit der Datenverwertung

Der hier verwendete Ansatz, das Sammeln und Auswerten von User Generated Content, wirft Fragen nach der Zulässigkeit des Vorgehens auf. Häufig befinden sich diese Daten in frei oder geschützt zugänglichen Anwendungen auf kommerziell betriebenen Plattformen. Diese Datenbanken auszulesen und auszuwerten, könnte gegen das Recht auf Datenbanken verstossen.

Auf der anderen Seite könnten sich die User einer solchen User Generated Content Anwendung durch diese Art der Informationssammlung und Auswertung in ihrer Privatsphäre gestört fühlen. Mit dem Wissen um eine solche Auswertung konfrontiert, könnte sich bei ihnen das Gefühl einstellen, überwacht zu werden.

Der Ruf nach Privatsphäre und die Empörung über deren Verletzungen sind häufig laut und groß. Verwunderlich ist aber die Eigenart betroffener Nutzer, schon für kleinste Anreize in ihrer Empörung zu verstummen und persönliche Informationen preiszugeben.

Eine mögliche Argumentation zur Rechtfertigung von Abgreifen und Auswerten von User Generated Content basiert auf der Unterscheidung von privaten und öffentlichen Räumen oder Sphären. Man kann ungeschützt zugängliche User Generated Content Plattformen als öffentlichen Raum verstehen. Damit handeln deren User bewusst öffentlich und machen sich damit beobachtbar. Nach dieser Argumentation wäre der gewählte Ansatz keine Verletzung der Privatsphäre von Usern. Die hier beschriebene Auswertungsmethode, die Bookmarkingereignisse aggregiert und somit anonyme Resultate liefert, wirft keine Probleme im Bereich verletzter Privatsphäre auf. Schwieriger in diesem Kontext zu rechtfertigen wäre das automatische Aufspüren von interessanten Usern. Aber solange Delicious-User Pseudonyme verwenden und diese nur schlecht zu realen Personen in Beziehung gesetzt werden können, bleibt das Problem verletzter Privatsphäre

marginal.

Trotz dieser Rechtfertigung sind durch digitale Technologien Überwachungs- und Suchmöglichkeiten von allgemeinen und auch von persönlichen Daten stark gestiegen. Die Diskussion der Folgen, der Bedeutung dieser Möglichkeiten, ihrer Risiken und wie wir als Gesellschaft auf sie reagieren sollten, ist noch lange nicht abgeschlossen. Lawrence Lessig schätzt vor, dass wir überlieferte Werte, die durch ein verändertes technologisches Umfeld gefährdet werden, überprüfen und mittels Übersetzung geltender rechtlicher Regelungen in den neuen Kontext die Wahrung dieser Werte sicherstellen (vgl. Lessig 2006, S. 157f.).

6.7 Zusammenfassung der Ergebnisse

In diesem Abschnitt wurden die Auswertungsmethoden an Beispielen vorgestellt. Zuerst wurde eine einfache Auswertungsmethode beschrieben, die über frei wählbare Auswertungsintervalle Top- k Listen von URLs erzeugt. Als eine Art erweiterte Auswertung wurde die Entwicklung von Bookmarkingereignissen in Diagrammen visualisiert. Die an Beispielen gezeigten Methoden dienen dazu, die Leistungsfähigkeit des Systems zu erfassen und Probleme aufzuzeigen.

Folgend werden, die an den Beispielen entdeckten Ergebnisse in kurzen Stichpunkten zusammengefasst :

- Relevante Resultate werden für gute Kriterien und dominante Zielthemen erzeugt.
- Für schwache Zielthemen erzeugt das System kaum relevante Resultate.
- Der Zusammenhang zwischen Kriterium und Zielthema ist, bis er getestet wird, eine Vermutung.
- Die URLs, der mit dem Kriterium ausgewählten User, können in einen unerwarteten, plausiblen Bezug zum Kriterium stehen.
- Das beschriebene System erzeugt irrelevante Resultate aufgrund von:
 - URLs zu populären Themen
 - getrennter Zählung verschiedener URLs mit gleichem Ziel
 - Schwachen Zielthemen
- Es wurde ein typischer Verlauf in der Entwicklung von Bookmarkingereignissen festgestellt, bei dem eine URL innerhalb von ca. drei Tagen durch ihr Maximum schreitet, um danach nur von wenigen Usern neu gespeichert zu werden.

7 Ausblick

Das beschriebene System zur automatisierten Unterstützung von Methoden der Markt-, Trend- und Zukunftsforschung kann Erweiterungen und Verbesserungen erfahren.

Ein Ziel solcher Erweiterungen kann es sein, die Relevanz der Ergebnislisten zu verbessern. Recht einfach scheint eine Erweiterung um zusammengesetzte Kriterien zur Gruppenbildung zu sein. Statt nur ein Kriterium zu erlauben, könnten Kombinationen von Kriterien verwendet werden, um mit Schnittmengen oder Vereinigungen von Usergruppen Zielthemen besser zu spezifizieren und mehr oder auch weniger User zu Auswertungsgruppen zusammenzufassen.

Ein anderer Ansatz zur Verbesserung der Relevanz von Ergebnislisten besteht in der möglichen Einführung eines nachgelagerten Filters für die Elemente der Ergebnislisten. Mit der Unterstützung von URL-Tag-Clouds ließen sich Ergebnisse themenbasiert filtern. Dabei könnten z.B. Ergebnislisten von mit Tag-Kriterien gebildeten Gruppen vor der Ausgabe mit dem gleichen Tag gefiltert werden, so dass die Ergebnisliste nur noch URLs enthält, in deren Tag-Cloud auch das Kriteriums-Tag vorkommt. Um die Ergebnisliste noch spezifischer zu filtern, könnte die Bedingung gestellt werden, das Gruppen-Tag unter den ersten Top- k Tags jeder URL-Tag-Cloud zu finden. Diese Vorgehensweise ließe sich auch auf Domain-Kriterien übertragen. Damit sollten sich auch für schwache Zielthemen relevante Ergebnisse erzeugen lassen.

Auf Basis der über 46 Millionen gesammelten Delicious Bookmarks können weitere Untersuchungen durchgeführt werden. Einige der möglichen Fragestellungen folgen hier stichpunktartig :

- Eigenschaften und Früherkennung von Dauerbrennern⁵¹
- Tagstrukturen aufdecken
- User- und Tagclustering
- Aufarbeitung für und Umsetzung von allgemeinen semantischen Such- und Rankingmethoden
- Erkennen von Lead Usern (von Hippel 1986) und Tipping Points (Gladwell 2002) in der Entwicklung von Bookmarkingereignissen.
- Entdecken von Userähnlichkeiten für Empfehlungssysteme⁵²

⁵¹Dauerbrenner sind kontinuierlich von neuen Usern gespeicherte URLs. Sie wurden unter der Überschrift: *Der Summen-Rang* in 6.4 erwähnt.

⁵²Damit könnten den Usern von Social Bookmarking-Systemen andere User ähnlicher Interessen als Informationsquelle vorgeschlagen werden. Dieses Vorschlagsystem könnte den Aufwand zum Entdecken interessanter User deutlich reduzieren.

Literatur

- Amitay, E., D. Carmel, M. Herscovici, R. Lempel, und A. Soffer (2004, Dezember). Trend detection through temporal link analysis. *J. Am. Soc. Inf. Sci. Technol.* 55(14), 1270–1281. Online: <http://portal.acm.org/citation.cfm?id=1063169> [08.11.2006]. DOI: 10.1002/asi.20082. 57
- Ansoff, I. H. (1975). Managing Strategic surprise by response to weak signals. *California Management Review* 18(2), 21–33. 9, 10, 12
- Ansoff, I. H. (1980). Strategic Issue Management. *Strategic Management Journal* 1(2), 131–148. 12
- Bakos, Y. und E. Brynjolfsson (2000). Aggregation and Disaggregation of Information Goods. In: *Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property*. 23, 24
- Begelman, G., P. Keller, und F. Smadja (2006). Automated Tag Clustering: Improving search and exploration in the tag space. In: *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*. Online: <http://www.rawsugar.com/www2006/20.pdf> [07.03.2007]. 56
- Benkler, Y. (2002, Dezember). Coarse's Penguin, or, Linux and The Nature of the Firm. *The Yale Law Journal* 112, 369–446. 18, 31
- Benkler, Y. (2006, Mai). *The Wealth of Networks : How Social Production Transforms Markets and Freedom*. Yale University Press. Online: <http://www.amazon.co.uk/exec/obidos/ASIN/0300110561/> [07.03.2007]. 32, 33
- Benkler, Y. und H. Nissenbaum (2006, Dezember). Commons-based Peer Production and Virtue. *The Journal of Political Philosophy* 14, 394–419. 32
- Bleicken, J. (1995). *Die athenische Demokratie* (4. Aufl.). UTB für Wissenschaft. 27
- Brynjolfsson, E. und X. M. Zhang (2007). Innovation Incentives for Information Goods. *Innovation Policy and the Economy* 7, 99–123. Online: <http://seeit.mit.edu/Publications/brynjolfsson-zhang6-14-06.pdf> [03.03.2007]. 24
- Cattuto, C., V. Loreto, und L. Pietronero (2006, Mai). Collaborative Tagging and Semiotic Dynamics. Online: <http://arxiv.org/abs/cs/0605015> [27.10.2006]. 56
- Clarke, M. und S. M. N. Islam (2004). *Economic Growth and Social Welfare* (1. Aufl.). Elsevier. 19, 20
- Coarse, R. H. (1937, November). The Nature of the Firm. *Economica, New Series* 4, 386–405. 26
- Cunningham, W. (2005, November). Correspondence on the Etymology of Wiki. Online: <http://c2.com/doc/etymology.html> [01.05.2007]. 35

- Delicious (2007, 12.März). Popular tags on del.icio.us. Online: <http://del.icio.us/tag> [12.03.2007]. 47
- Dubinko, M., R. Kumar, J. Magnani, J. Novak, P. Raghavan, und A. Tomkins (2006). Visualizing tags over time. In: *WWW '06: Proceedings of the 15th international conference on World Wide Web*, New York, NY, USA, S. 193–202. ACM Press. Online: <http://portal.acm.org/citation.cfm?id=1135777.1135810> [27.10.2006]. DOI: 10.1145/1135777.1135810. 56, 57, 58, 64, 65, 66
- Encyclopædia Britannica, Inc. (2003). *The New Encyclopaedia Britannica* (15. Aufl.), Volume 1. 14
- Fagin, R., A. Lotem, und M. Naor (2003, Juni). Optimal aggregation algorithms for middleware. (4), 614–656. 66
- Gladwell, M. (2002). *Tipping Point - Wie kleine Dinge Großes bewirken können* (2. Aufl.). Goldman. 94
- Glance, N. S., M. Hurst, und T. Tomokiyo (2004). BlogPulse: Automated Trend Discovery for Weblogs. Online: <http://www.blogpulse.com/papers/www2004glance.pdf> [12.04.2007]. 56
- Golder, S. A. und B. A. Huberman (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science* (2), 198–208. 44, 46
- Green, H. (2005, 9. Dezember). Yahoo snaps up del.icio.us. *Business Week Online*. Online: http://www.businessweek.com/the_thread/blogspotting/archives/2005/12/yahoo_snaps_up.html [04.04.2007]. 48
- Hammond, T., T. Hanny, B. Lund, und J. Scott (2005). Social Bookmarking Tools(I). *D-Lib Magazine* (4). Online: <http://www.dlib.org/dlib/april05/hammond/04hammond.html> [01.04.2007]. 45
- Horx, M. und P. Wippermann (1996). *Was ist Trendforschung*. Econ Verlag. 10, 58
- Hotho, A., R. Jäschke, C. Schmitz, und G. Stumme (2006a, Oktober). Emergent Semantics in BibSonomy. In: C. Hochberger und R. Liskowsky (Hrsg.), *Informatik 2006 - Informatik für Menschen. Band 2*, Volume P-94 of *Lecture Notes in Informatics*, Bonn. Online: <http://www.kde.cs.uni-kassel.de/stumme/papers/2006/hotho2006emergent.pdf> [01.04.2006]. Proc. Workshop on Applications of Semantic Technologies, Informatik 2006. 57
- Hotho, A., R. Jäschke, C. Schmitz, und G. Stumme (2006b, Juni). Information Retrieval in Folksonomies: Search and Ranking. In: *Proceedings of the 3rd European Semantic Web Conference*, LNCS, Budva, Montenegro, S. 411–426. Springer. Online: http://www.kde.cs.uni-kassel.de/hotho/pub/2006/seach2006hotho_eswc.pdf [12.04.2007]. 57

- Hotho, A., R. Jäschke, C. Schmitz, und G. Stumme (2006c, Dezember). Trend Detection in Folksonomies. In: Y. S. Avrithis, Y. Kompatsiaris, S. Staab, und N. E. O'Connor (Hrsg.), *Proc. First International Conference on Semantics And Digital Media Technology (SAMT)*, Volume 4306 of *LNCS*, Heidelberg, S. 56–70. Springer. Online: <http://www.kde.cs.uni-kassel.de/stumme/papers/2006/hotho2006trend.pdf> [12.04.2007]. DOI: 10.1007/11930334_5. 57
- Howe, J. (2006, 17. Dezember). Your Web, Your Way. *Time Magazine*. Online: <http://www.time.com/time/magazine/article/0,9171,1570815-1,00.html> [14.02.2007]. 1
- Hueck, T. (2004, Dezember). Trendforschung. Landkarten für die Zukunft. *Die Bank*, 70–73. 13, 14
- Internetworldstats (2007). Internet Usage Statistics - The Big Picture 11. Januar 2007. Online: <http://www.internetworldstats.com/stats.htm> [15.02.2007]. 2
- Jäschke, R., A. Hotho, C. Schmitz, und G. Stumme (2006, June). Wege zur Entdeckung von Communities in Folksonomies. In: S. Braß und A. Hinneburg (Hrsg.), *Proc. 18. Workshop Grundlagen von Datenbanken*, Halle-Wittenberg, S. 80–84. Martin-Luther-Universität. Online: http://www.kde.cs.uni-kassel.de/jaeschke/pub/jaeschke2006wege_gvd.pdf [12.04.2007]. 57
- Katz, I. (2006, November). Delicio.us, Joshua Schachter. Online: <http://www.guardian.co.uk/weekend/story/0,,1939056,00.html> [01.04.2007]. 48
- Keller, P. (2007, 23. April). delicio.us stats. Online: <http://deli.ckoma.net/stats> [23.04.2007]. 89
- Klein, G. (2003). *The Power of Intuition - How to use your gut feelings to make better decisions at work*. Currency, Doubleday. 10, 11, 13
- Kristóf, T. (2006, Juni). Is it possible to make scientific forecasts in social sciences? *Futures* 38(5), 561–574. Online: <http://www.sciencedirect.com/science/article/B6V65-4HRDY8H-1/2/863ac7bc19ed074cacb7f11b9370e310> [14.02.2007]. DOI: 10.1016/j.futures.2005.09.004. 11
- Lessig, L. (2002). *The Future of Ideas*. Vintage Books. 29
- Lessig, L. (2004). *Free Culture*. The Penguin Press. Online: <http://www.free-culture.cc/freeculture.pdf> [21.03.2007]. 23
- Lessig, L. (2006). *Code version 2.0*. Basic Books. 18, 33, 93
- Macgregor, G. und E. McCulloch (2006). Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. *Library Review* 55(5). Online: <http://eprints.rclis.org/archive/00005703/> [08.11.2006]. 56

- Malone, T. W. (2004). *The Future of Work*. Harvard Business School Press. 26, 28, 29
- Mankiw, N. G. und M. P. Taylor (2006). *Principles of Economics*. Thomson Learning. 20, 21
- Marlow, C., M. Naaman, D. Boyd, und M. Davis (2006, Mai). Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. In: *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*. Online: <http://www.rawsugar.com/www2006/29.pdf> [08.11.2006]. 44, 46
- McDaniel, C. und R. Gates (2006). *Marketing Research Essentials* (5. Aufl.). John Wiley & Sons. 3, 4, 8
- MisterWong (2007, 14.März). Mister Wong Anti-Spam Regeln. Online: <http://www.mister-wong.de/blog/mister-wong-anti-spam-regeln> [24.04.2007]. 89
- Monitor110 Inc. (2007). Online: <http://www.monitor110.com/index.php> [01.03.2007]. 15, 16
- Moore, G. A. (2006). *Crossing the Chasm* (First Collins Business Essentials Edition 2006 Aufl.). 91, 92
- Nielsen//Netratings (2006a, September). The Hottest Online Brands in 2006. Online: http://www.nielsen-netratings.com/pr/pr_060914_unitedkingdom.pdf [12.02.2006]. 2, 3
- Nielsen//Netratings (2006b, August). User-Generated Content Drives Half of U.S. Top 10 Fastest Growing Web Brands, According to Nielsen//Netratings. Online: http://www.netratings.com/pr/PR_060810.PDF [12.02.2006]. 3
- O'Reilly, T. (2005, 30.September). What Is Web 2.0?. Online: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> [24.04.2007]. 2
- Pfadenhauer, M. (2004, Mai). Wie forschen Trendforscher? Zur Wissensproduktion in einer umstrittenen Branche [41 Absätze]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-line Journal]* 5(2). Online: <http://www.qualitative-research.net/fqs-texte/2-04/2-04pfadenhauer-d.htm> [28.02.2007]. 9
- Rogers, E. M. (1995). *Diffusion of Innovations* (4. Aufl.). The Free Press. 6, 7, 91
- Saul, P. (2006, Februar). Seeing the Future in Weak Signals. *Journal of Futures Studies* 10. Online: <http://www2.tku.edu.tw/~tddx/jfs/10.3.93.pdf> [14.02.2007]. 10, 12, 13
- Schachter, J. (2005, 13.Mai). more hardware. Online: http://blog.del.icio.us/blog/2005/05/more_hardware.html [29.04.2007]. 60

- Schachter, J. (2007a, 21.März). rss has flavor. Online:http://blog.del.icio.us/blog/2007/03/rss_has_a_flavo.html [04.04.2007]. 51
- Schachter, J. (2007b, 29.März). that was fast. Online:http://blog.del.icio.us/blog/2007/03/that_was_fast.html [04.04.2007]. 47
- Scharffe, F. und Y. Ding (2006). Three Levels of Knowledge Structuration for the Web. Online:<http://sw.deri.org/~francois/pub/threeLevels.pdf> [12.02.2007]. 56
- Schmitz, P. (2006, Mai). Inducing Ontology from Flickr Tags. In: *Collaborative Web Tagging Workshop at WWW 2006, Edinburgh, Scotland*. Online:<http://www.rawsugar.com/www2006/22.pdf> [08.11.2006]. 56
- Scott, D. M. (2004, Juni). Information Arbitrage. *EContentmag.com*. Online:<http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=6609> [01.03.2007]. 16
- Shadbolt, N., T. Berners-Lee, und W. Hall (2006, Mai). The Semantic Web Revisited. *IEEE Intelligent Systems* 21, 96–101. Online:<http://www.geospatialsemanticweb.com/wp-content/uploads/2006/07/01637364.pdf> [10.04.2007]. DOI: 10.1109/MIS.2006.62. 45
- Shen, K. und L. Wu (2005, September). Folksonomy as a Complex Network. Online:<http://arxiv.org/abs/cs.IR/0509072> [31.10.2006]. 56
- Sietmann, R. (2006). Über die Ketten der Wissensgesellschaft. *CT 2006*, 190. Online:<http://www.heise.de/ct/06/12/190/> [05.04.2007]. 31
- SIGMA, G. (2007). Online:http://www.sigma-online.com/de/SIGMA_Milieus/SIGMA_Milieus_in_Germany/ [01.03.2007]. 5
- Sinus-Sociovision (2007). Online:<http://www.sinus-sociovision.de> [20.02.2007]. 5, 6
- Software Garden, Inc. (2004, 6.Juli). What is RSS? - A basic tutorial introduction to RSS feeds and aggregators for non-technical people. Online:<http://rss.softwaregarden.com/aboutrss.html> [05.04.2007]. 51
- Sunstein, C. R. (2006, August). *Infotopia: How Many Minds Produce Knowledge*. Oxford University Press, USA. 34
- United Internet Media AG (2007). Online:<http://www.webmilieu.web.de/zielgruppen> [01.03.2007]. 5
- upian Hot Links (2007). joshua quitting to work on del.icio.us full time. Online:<http://dev.upian.com/hotlinks/tag/joshuaschachter> [11.04.2007]. 48
- Vander Wal, T. (2004, August). You Down with Folksonomy?. Online:<http://www.vanderwal.net/random/entrysel.php?blog=1529> [12.04.2007]. 57

-
- von Hippel, E. (1986, Juli). Lead Users: A Source of Novel Product Concepts. *Management Science* 32, 791–805. 94
- von Krogh, G. und E. von Hippel (2006). The Promise of Research on Open Source Software. *Management Science* 52(7), 975+. Online: <http://erf.sbb.spk-berlin.de/han/5632/mansci.highwire.org/cgi/reprint/52/7/975> [07.03.2007]. 31
- Voss, J. (2006, April). Collaborative thesaurus tagging the Wikipedia way. Online: <http://arxiv.org/abs/cs.IR/0604036> [04.04.2007]. 46
- Washington, H. A. (Hrsg.) (1854). *The Writings of Thomas Jefferson*. Taylor & Maury. 22
- Wikipedia (2007, 21:19 1.April). Wikipedia Categorization. Online: <http://en.wikipedia.org/wiki/Wikipedia:Categorization/> [04.04.2007]. 46
- Wood, C. A. und R. J. Kauffman (2006, November). Revolutionary Research Strategie for E-Business: A Philosophy of Science View in the Age of the Internet. Online: http://www.misrc.umn.edu/workingpapers/fullPapers/2006/0614_113006.pdf [07.03.2007]. 56
- Yates, J. und R. I. Benjamin (1991). The Past and Present as a Window on the Future. In: M. S. S. Morton (Hrsg.), *The Corporation of the 1990s*, S. 61–92. Oxford University Press. 28