

Web 2.0

Die gesellschaftlichen Aspekte des neuen Webs

Stefan Schöttelndreyer

Matr. Nr. 652351

Martin Schmidt

Matr. Nr. 664006

Zusammenfassung: Das World Wide Web ist im Wandel. Der Begriff Web 2.0 umfasst verschiedene Techniken und neue Anwendungsmöglichkeiten, die im Zusammenhang mit dieser Metamorphose stehen. Diese und die damit verbundene Bewusstseinsveränderung der Internetbenutzer sollen hier aufgezeigt werden.

1 Einführung

Die Zeit, in der das Internet vom Großteil der Bevölkerung als Medium für besonders technisch versierte Anwender belächelt wurde, ist vorbei. Die durch das Internet zur Verfügung gestellten Dienste sind aus unserem Leben kaum noch wegzudenken. E-Mails ersetzen schon seit einiger Zeit Briefe und Faxe, ja sogar das Telefon wird über das "Netz der Netze" geroutet. Durch die stetig steigende Anzahl von Webdiensten wird das Netz für immer mehr Menschen interessant. "Online" zu sein bedeutet heute nicht viel mehr als vor einigen Jahren einen Telefonanschluss zu besitzen. Das ist nur einer der Gründe, warum sich das World Wide Web im stetigen Wandel befindet. Neue Webdienste wie beispielsweise Instant Messaging, Webforen, Blogs und Wikis verändern die Wahrnehmung des Webs durch den Benutzer. Er wird aktiv in die Mitgestaltung des Webs eingebunden und kommuniziert auf neue Arten. Dieses neue Gefühl der Mitbestimmung und die ihm zugrunde liegenden Techniken bilden das Web 2.0. Diese Ausarbeitung wird diese Wahrnehmungsveränderung des Internets anhand von Beispielen aufzeigen.

2 Die Geschichte des World Wide Web

Zu Beginn des World Wide Web standen die Inhalte und nicht etwa die Darstellung der Informationen im Vordergrund. Um nun die übermittelten Daten logisch beschreiben zu können entwickelte der Brite Tim Berners-Lee 1989 am CERN in Genf ein Format zur Repräsentation dieser Daten – eine auf SGML basierende Auszeichnungssprache, die später den Namen "HyperText Markup Language" erhielt. Auch der erste Webbrowser wurde im Jahr 1990 von Berners-Lee, damals noch unter dem Namen "WorldWideWeb", vorgestellt. Ende des Jahres 1993 erschien mit dem Namen "Mosaic" der erste Browser, der außer Text auch Grafiken auf einer Seite anzeigte.

Die Fähigkeit, das Layout von Webseiten mittels HTML zu definieren wurde erst später durch Softwarehersteller geprägt, die ihre Vorherrschaft im Bereich Web-Technologien ausbauen wollten. Um bestimmte Layouteffekte zu erzielen, vermischten Webentwickler häufig Tabellen, HTML-Elemente und eingebettete Grafiken. Es entstand die so genannte "Tag Soup", eine wirre Aneinanderreihung von Tags, die von den Fehlertoleranzmechanismen der Browser meist unterschiedlich interpretiert wurden. Die Nachteile dieser Vorgehensweise liegen klar auf der Hand; die Darstellung der Webseiten ist abhängig von Browser oder Betriebssystem, benachteiligte Personengruppen, die beispielsweise auf Screen Reader angewiesen sind werden ausgegrenzt, et cetera.

Um die wachsende Menge von Seiten effizient bearbeiten und verwalten zu können, wurden Content-Management-Systeme (CMS) entwickelt, die datenbankbasiert Seiten dynamisch zur Laufzeit generieren. Die Arbeit an großen Seiten von mehreren Autoren wurde durch eine komplexe Benutzerverwaltung erleichtert und der automatisch generierte Code entspricht immer öfter gültigen Webstandards. Foren und Wikis sind exemplarische Vertreter dieser Anwendungen. Wollte man tatsächlich eine Versionierung im Web vornehmen, so müsste diese Technik mit der Versionsnummer 1.5 bedacht werden.

Seit dem Jahr 2005 kündigt sich eine Renaissance im World Wide Web an. Es ist aber keine neue Technologie, die das Web 2.0 einläutet, eher ein von der Gemeinschaft der Webanwender empfundenes Gefühl der Veränderung von Regeln und Geschäftsmodellen. In den folgenden Kapiteln werden einige dieser neuen Modelle, die den Umgang mit dem Web beeinflussen, behandelt.

3 Weblogs

Bis vor einigen Jahren benötigte man noch ein sehr spezielles Fachwissen um Webseiten im Internet zu veröffentlichen. Die zunehmende Anzahl derer, die dieses Wissen nicht aufbringen konnte veranlasste Softwarehersteller Anwendungen zu entwickeln, mit denen sich auch weniger technisch orientierte Menschen im Internet präsentieren konnten. Diese Weblogsoftware zeichnet sich durch ihre einfache Bedienbarkeit aus und ist zudem – im Gegensatz zu herkömmlichen Content Management Systemen – auf die speziellen Bedürfnisse der Blogger ausgerichtet. Das Wort Weblog oder kurz Blog setzt sich aus Web und Log zusammen. Obwohl schon Mitte der 1990er Jahre in den USA die ersten Weblogs entstanden, vergingen viele Jahre bis sich auch hierzulande das Bloggen etablierte. Ebenfalls erst sehr spät nahmen die traditionellen Medien die Macht der Weblogs wahr.

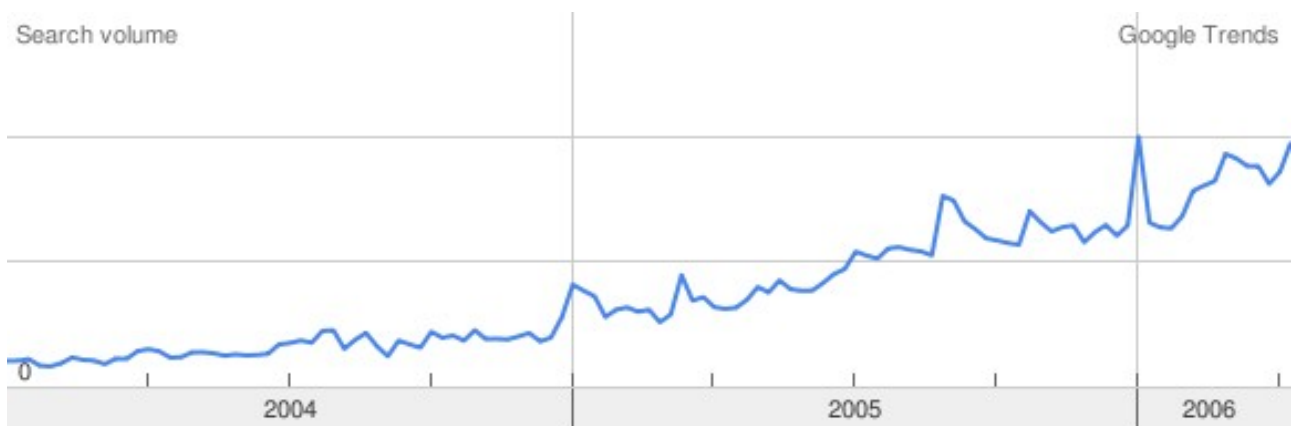


Abbildung 1: Suchaufkommen des Wortes "Blog" in Deutschland

Ein Weblog wird definiert als eine regelmäßig aktualisierte Webseite, die Informationen beinhaltet, welche in umgekehrter chronologischer Reihenfolge präsentiert werden. Blogs enthalten oft eine hohe Linkdichte und vertreten in der Regel eher persönliche Meinungen als neutrale Standpunkte. Zudem können Blogbeiträge kommentiert und mit Trackback Funktionen versehen werden.

Zuweilen wird im Zusammenhang mit Blogs auch von "Graswurzel-Journalismus" oder "partizipativem Journalismus" gesprochen. Dies kann durchaus für die Gemeinde der Blogger – wenn auch nicht für jeden Einzelnen – durchaus zutreffend sein. Die Definition für partizipativen Journalismus von Shayne Bowman und Chris Willis lautet folgendermaßen:

The act of a citizen, or group of citizens, playing an active role in the process of collecting, reporting, analyzing and disseminating news and information. The intent of this participation is to provide independent, reliable, accurate, wide-ranging and relevant information that a democracy requires. Aus: We Media - How audiences are shaping the future of news and information, Shayne Bowman, Chris Willis, S. 9, 2003

3.1 Geschichtliche Entwicklung

Das erste Weblog entsteht Anfang der 1990er Jahre in der Frühzeit des World Wide Web. Erstellt wurde es von Tim Berners-Lee, dem Erfinder des Webs. Um diese Thematik ging es auch in seinem Blog; er listete chronologisch alle neuen Webseiten auf.

Mitte der 1990er Jahre entstanden die ersten Online-Tagebücher, auf denen ihre Betreiber regelmäßig Einträge verfassten, jedoch gab es bis zum Jahr 1999 nur etwa 25 Webseiten, die unter die heutige Auffassung von Blogs fallen. Der amerikanische Philosoph und Programmierer Jorn Barger benutzte erstmals 1997 den Begriff "Weblog" auf seiner Seite "robot wisdom", auf der er seine Touren durch das Web "loggte". Die Kurzform "Blog" wurde kurz darauf vom Webdesigner Peter Merholz eingeführt. Die flächendeckende Ausbreitung von Blogs geschah 1999 mit der Entwicklung von frei nutzbaren

Webanwendungen, mit denen es möglich war einen Blog ohne technisches Hintergrundwissen zu pflegen.

3.2 **Trackback**

Ein wichtiger Bestandteil von Blogs ist die so genannte Trackback-Funktion. Entwickelt wurde sie für die Blog-Software "Movable Type" der Firma "Six Apart". Mittels des Trackbacking ist es möglich, eine Vernetzung von Blogs untereinander zu schaffen, die sich auf den jeweiligen Artikel beziehen. Prinzipiell ist Trackback nichts anderes als ein automatischer Benachrichtigungsdienst. Wird ein Artikel in einem Blog veröffentlicht, der sich auf einen anderen Artikel in einem anderen Blog bezieht, so können Leser des Originalartikels mittels der "Trackbacks" – wie die eingehenden Trackback-Nachrichten genannt werden – die sich auf den Originalartikel beziehenden Blogbeiträge der anderen Blogs verfolgen. Bei älteren Weblog-Anwendungen musste die Trackback-URL auf den Artikel, auf den Bezug genommen werden sollte dem Blogbeitrag manuell hinzugefügt werden. Modernere Software bietet hingegen "Trackback Auto Discovery" an, eine automatische Trackback-Erkennung.

Das Trackback ist aus technischer Sicht nicht gerade eine bahnbrechende Funktion, die dem technisch versierten Leser an eine Erweiterung des Hyperlinks erinnern lässt. Sie birgt jedoch enorme Vorteile. Auf diese Weise werden Diskussionen, anders als z.B. in Foren, dezentral geführt - jeder Blogger schreibt in seinem eigenen Blog. Dies erhöht zum einen die Ausfallsicherheit, zum anderen hat es den Vorteil, dass die Beiträge qualitativ hochwertiger sind, da sich niemand in seinem eigenen Blog durch unqualifizierte Äußerungen hervortun möchte.

3.3 **Einsatzmöglichkeiten**

Im Gegensatz zu Foren oder Mailinglisten sind die Diskussionen der "Blogosphäre", wie sich die Gemeinschaft der Blogger nennt, nicht auf bestimmte Themengebiete beschränkt. Aus diesem Grund entstehen etliche Gruppierungen, die locker untereinander vernetzt sind und deren Grenzen sich überlagern.

Die Einsatzmöglichkeiten von Blogs sind enorm. Ob als politisches Machtinstrument, authentische Kriegsberichterstattung oder Marketingfaktor. Blogs verändern unsere gesellschaftliche Umwelt schon heute mehr als wir vielleicht glauben mögen.

4 **Wikis**

Untrennbar verbunden mit dem Bewusstsein eines neuen Webs sind die Wikis, Systeme zur kollaborativen Sammlung von Informationen zu Themen jeglicher Art. Der Name „Wiki“ ist an das hawaiianische Wort „wikiwiki“ angelegt, was übersetzt „schnell“ bedeutet. Schnelle Bearbeitung ist ein wichtiges Grundprinzip der Wikis. Zentraler Verwendungszweck der Wikis ist die Sammlung, Speicherung und Darstellung von Artikeln zu beliebigen Themen. Besonderheit ist hierbei, dass die Artikel von jedermann bearbeitet werden können.

Realisiert werden Wikis mit einer Datenbank und einer HTML-Oberfläche. Die Oberfläche zeigt dem Benutzer die verfügbaren Artikel an, gibt ihm Möglichkeiten, innerhalb der Sammlung zu suchen und selbst eigene Artikel zu erstellen und bestehende Artikel zu bearbeiten

In den ersten Jahren des Webs war es nur wenigen Menschen möglich, Informationen ins Netz zu stellen. Man musste die Prinzipien des Webs kennen, sich mit der Designsprache HTML auseinandersetzen und eigenen Webpage zur Verfügung stellen, um anderen Menschen Informationen zugänglich zu machen. Die Konsequenz daraus war, dass oft nur technisch versierte Menschen in der Lage waren, eigene Webseiten anzubieten. Für Menschen, die dieses technische Wissen nicht besaßen, war es ungleich schwieriger, Informationen zu ihrem Fachgebiet im Netz zu präsentieren.

Im Jahre 1995 entwickelte Ward Cunningham, amerikanischer Software-Entwickler, ein System, das er WikiWikiWeb¹ nannte. Das Thema dieses ersten Wikis waren Softwareentwurfsmuster, an deren Entwicklung Cunningham maßgeblich beteiligt war. Um die Arbeit an diesen Entwurfsmustern zu verbessern, ersann Cunningham das Prinzip der Wikis. So war es Entwicklern aus der ganzen Welt möglich, zentral an der Entwicklung und Theorie von Entwurfsmustern zu arbeiten. Bis heute ist das WikiWikiWeb eines der bedeutendsten Wikis weltweit und wird immer noch gepflegt.

Das WikiWikiWeb von Cunningham blieb nicht das einzige seiner Art. Von der Idee begeistert, über simple

¹ Das WikiWikiWeb: <http://c2.com/cgi/wiki>

Bearbeitungsformulare auch Laien die Möglichkeit zur Informationsdarstellung zu geben, begannen auch andere, mit Cunninghams Software Wikis aufzubauen, seine Software umzuschreiben oder selbst eigene Software zu entwickeln. Die zurzeit bedeutendsten Wikis sind die Wikipedia², das MeatballWiki, Susning.Nu oder das GeneaWiki mit je mehr als 40.000 Artikeln zu vielfältigen Themenbereichen.

Wikis entwickelten sich schnell zu universellen Nachschlagewerken. Da die Wikipedia das beste Beispiel dafür ist, soll sie, auch aufgrund ihrer Größe und ihres internationalen Erfolges, hier exemplarisch betrachtet werden.

4.1 Die neuen Nachschlagewerke

Die Wikipedia ist aus dem recht erfolglosen Nupedia-Projekt hervorgegangen, das eine freie, große Enzyklopädie erstellen wollte, die mit den traditionellen Nachschlagewerken mithalten konnte. Der Grund für das Scheitern der Nupedia war die restriktive Artikelerstellung, die einen Text durch mehrere Instanzen an mehreren Fachleuten und Lektoren vorbeischickte, bevor der Text als Artikel aufgenommen werden konnte. Kurz bevor das Nupedia-Projekt eingestellt wurde, kam bei den Verantwortlichen die Idee auf, das Konzept eines Wikis zu benutzen. Mehr als Experiment gedacht, entwickelte sich die Wikipedia – wie das System jetzt hieß – immer mehr zu einem ernsthaften, dennoch freien Nachschlagewerk.

Der weltweite Erfolg und die Qualität der Wikipedia basieren auf fünf Grundsätzen, die im Artikel „Wikipedia“³ aufgestellt und im Folgenden erläutert werden:

1. Wikipedia ist eine Enzyklopädie
2. Neutraler Standpunkt
3. Verifizierbarkeit und Verzicht auf Primärrecherche
4. Urheberrecht und Freiheit der Inhalte
5. Respektvoller Umgang

Die Wikipedia betrachtet sich als Sammlung von Wissen. Das Ziel ist, möglichst viel unseres Wissens in digitale Form zu bringen, um es der Menschheit zur Verfügung zu stellen. Wichtig ist hierbei, dass die reinen, nicht zusammenhängenden Daten analysiert und interpretiert werden müssen, um daraus Wissen zu gewinnen. Demzufolge versteht sich die Wikipedia nicht als Datensammlung, sondern explizit als Sammlung von bereits aufbereitetem und analysiertem Wissen.

Damit dieses Wissen universell anwendbar wird, muss es in möglichst objektiver Form vorliegen, die den Sachverhalt unabhängig von ethischen, moralischen und religiösen Weltanschauungen darstellt. Implizit bedeutet das: Durch neutrale Beschreibung des Sachverhaltes wird ein Standpunkt erreicht, der jeder Ansicht ein Mindestmaß an Relevanz einräumt, ohne jedoch eine Ansicht zu bevorzugen.

Als Enzyklopädie darf die Wikipedia keine Forschung betreiben. Alle dargestellten Fakten eines Artikels müssen durch bereits durchgeführte Experimente oder durch bestehende Theorien untermauert werden. Dabei ist es wichtig, auf Primärliteratur zu verzichten und nur bereits anerkannte Sekundärquellen zu verwenden, um nur von der Gesellschaft akkreditiertes Wissen zu stellen. Das heißt auch, dass nicht bereits verifiziertes Wissen nicht eingestellt werden darf, um das Prinzip der Enzyklopädie nicht zu verletzen.

Die Inhalte der Wikipedia werden unter der GNU-Lizenz für freie Dokumentation⁴ (GFDL) veröffentlicht. Diese besagt in den Grundzügen, dass unter dieser Lizenz stehende Texte weiterverarbeitet und –verbreitet werden dürfen. Die Lizenz erlaubt ausdrücklich kommerzielle Nutzung, verpflichtet den Verwerter aber dazu, alle vom Ursprungstext abgeleiteten Dokumente mit Namen der vorherigen Autoren zu versehen und ebenfalls unter die GFDL zu stellen. Jeder Autor eines Textes bindet sich mit dem Eintrag des Artikels in die Wikipedia an diese Lizenz. Die sehr streng gehaltene GFDL ermöglicht es den Wikipedia-Inhalten, für alle frei zu bleiben und von der exklusiven Verwertung eines einzelnen Rechteinhabers ausgenommen zu

² <http://www.wikipedia.org>

³ <http://de.wikipedia.org/wiki/Wikipedia>

⁴ <http://www.gnu.org/licenses/fdl.html#TOC1>

werden.

Der fünfte und letzte Grundsatz hat weniger mit den Artikeln selbst zu tun, sondern bezieht sich mehr auf die zu jedem Artikel mögliche Diskussion. Zu jedem Artikel gibt es eine Diskussionsseite, auf der jeder Betrachter des Artikels Kommentare schreiben kann. Dies dient dazu, dass Kritik und Verbesserungsvorschläge zu einem Artikel davon separat gehalten werden, um die Übersichtlichkeit zu wahren. Der Grundsatz ruft nun zu respektvollem Diskutieren auf den Diskussionsseiten auf. Da kontroverse Diskussionen im Internet leicht zu einem „Flamewar“⁵ werden, macht dieser Grundsatz Sinn und wird auch von den Administratoren der Wikipedia durchgesetzt (z.B. durch Löschung beleidigender Beiträge und Verbannung der beleidigenden Benutzer).

Ausgehend von diesen Prinzipien ist mittlerweile eine Artikelsammlung entstanden, die es an Umfang und Qualität durchaus mit kommerziellen und traditionelleren Enzyklopädien wie der Encyclopaedia Britannica oder dem Brockhaus aufnehmen kann. Die Möglichkeit für jeden Menschen, anonym und ohne aufwendige Registrierung an den Inhalten der Wikipedia zu arbeiten, hat dazu geführt, dass die Wikipedia eine Breite an Wissensgebieten aufweist, die sonst nur mit jahrzehntelanger Recherche aufgebaut werden kann. Durch diese Attraktivität begeistert und animiert, beteiligen sich Fachleute aus Bereichen wie Literatur, Kunst, Geschichte, Mathematik und Informatik, Technik, Wirtschaft, Geographie und Sprachwissenschaften an der Pflege und Erweiterung der Artikel.

Dieses „Team“ von über 100.000 ständiger Autoren sorgt für stets aktuelle und korrekte Artikel, die sich an die Grundsätze einer Enzyklopädie halten: Ausführliches Wissen, das für alle verständlich sein soll.

4.2 Gemeinsame Beteiligung

Die gemeinsame Beteiligung ist ein wichtiger Aspekt der Wikis und der Wikipedia im speziellen. Am Anfang eines Artikels steht allerdings der Erstautor. Im Allgemeinen hat dieser Erstautor einen erstaunlichen Anspruch an sich selbst. Obwohl man schnell einen Artikel einbringen kann, beschäftigen sich die meisten Autoren vorher intensiv mit dem Thema. Oft weisen die Artikel große Ähnlichkeit zu wissenschaftlichen Ausarbeitungen auf. Die Autoren recherchieren in renommierten Quellen, Fachbüchern oder anderen Publikationen. Nach der Recherche wird der Text verfasst, oft unterteilt in Einleitung, Hauptteil und Zusammenfassung. Das Literaturverzeichnis zeigt die verwendeten Quellen. Verweise auf Internetseiten bringen die Möglichkeit, zusätzliche Informationen einzuholen.

Nachdem der Autor seinen Text in einen neuen Artikel eingetragen hat, beginnt der „*Tanz der Gehirne*“⁶: Der Artikel steht nun in der Öffentlichkeit und wird von Mitgliedern der Wikipedia-Community gelesen. Die ersten Diskussionen über den Inhalt entstehen, so führt zum Beispiel eine einseitige Betrachtungsweise zu einem Ruf nach mehr Objektivität. Sachliche Fehler werden erkannt, oft direkt gepaart mit Quellenangaben, die den Fehler aufzeigen und berichtigen können. Weitere mögliche Aspekte zu dem entsprechenden Thema werden eingebracht oder zusätzliches Bild-, Ton- oder Videomaterial bereitgestellt. Nicht selten werden diese Verbesserungen direkt vorgenommen, da jeder Benutzer die Seite editieren kann. Rechtschreibfehler oder Ausdrucksschwächen werden meist direkt von berufenen Autoren korrigiert. So steigert sich die Qualität der Artikel kontinuierlich.

Angemeldete Autoren (d.h. bei Wikipedia namentlich registriert) erlangen über die Einstellung von Texten und den Diskussionsbeitrag eine gewisse Reputation. Diese Reputation bringt den Autoren oft eine erhöhte Kompetenz ein, die sich auch auf die Entscheidungsfindung niederschlägt. Denn die Inhalte der Wikipedia werden von der Community selbst verwaltet. Die meisten Änderungen werden in einem Konsens-Verfahren vorgenommen, das oft eine kontroverse Diskussion beinhaltet. In diesen Diskussionen werden renommierte Mitglieder der Wikipedia-Gemeinde eher als entscheidende Instanz betrachtet als ein anonymer Schreiber. Verschiedene Einrichtungen geben weiterhin Anreiz, die Qualität der Artikel zu erhöhen. So kann ein Artikel mit dem Prädikat „Lesenswert“ ausgezeichnet werden. Das diese Auszeichnung vorher eine Diskussion der Gemeinde benötigt, ist im Blick auf die soziale Struktur der Wikipedia nicht verwunderlich. Sollte die Qualität des Artikels außergewöhnlichen Maßstäben genügen, kann er dann auch in die Liste der „Exzellenten Artikel“ aufgenommen werden. Diese Beachtung geht dann auch wieder auf den Autor zurück, der diesen Artikel „lesenswert“ bzw. „exzellent“ gestaltete.

⁵ „Flamewar“: Eine Diskussion, die in unsachlicher und beleidigender Weise stattfindet.

⁶ Aus: [8]

Gelöscht werden Artikel im Übrigen sehr selten, da dazu ein Administrator eingreifen muss. Falls ein Artikel als so schlecht oder unangebracht eingestuft wird, dass er nicht mehr sinnvoll überarbeitet werden kann, muss zunächst ein Löschantrag gestellt werden. Falls innerhalb einer gewissen Zeit kein Widerspruch eingeht, wird der Artikel gelöscht. In den meisten Fällen jedoch entwickelt sich um den Löschantrag eine rege und kontroverse Diskussion, die die Löschung verhindert oder zumindest verzögert.

4.3 Probleme und Missbrauch

Dies führt uns nun zu den Problemen der Wikipedia, die als Hauptkritikpunkte genannt werden. Der Weg der Entscheidungsfindung der Wikipedia wird mangels hierarchischer Instanzen oft als anarchistisch bezeichnet. Es gibt keine Eingriffe seitens der Administratoren, die gegen den Willen der Gemeinde laufen würden. Die Organisatoren – die Wikimedia Foundation⁷ – haben nur eine recht vage Roadmap festgelegt (Fünf Grundsätze, siehe oben.). Aufgrund dieser fehlenden Regulierung hat sich in der Wikipedia-Gemeinde ein soziales Umfeld entwickelt, dessen Entscheidungen für neue Mitglieder schwer nachzuvollziehen sind. Kritiker werfen den Handlungen innerhalb der Wikipedia-Gemeinde daher folgerichtig oft einen Mangel an Transparenz vor. Die teilweise vorherrschende Inkonsequenz in der Entscheidungsfindung und die mangelnden Strategien, Konflikte zu vermeiden bzw. zu beenden, sind weitere Kritikpunkte. So werden z.B. die Administratoren von allen Mitgliedern gewählt, obwohl Abstimmungen generell aber als ungeeignet angesehen werden. Hier fehlen regulierende Instanzen aus der Wikimedia Foundation.

Viele Wissenschaftler werden von der sozialen Dynamik der Wikipedia abgeschreckt. Das Nichtbeachten traditioneller Publikationsabläufe ist für sie ebenso ungewohnt wie die lasche Einhaltung bestimmter, in wissenschaftlichen Bereichen angewandeter Richtlinien wie Textstruktur und Quellenverzeichnis. Trotz verschiedener Maßnahmen, Wissenschaftlern die Wikipedia schmackhaft zu machen – z.B. offizielle Aufrufe in „Nature“ oder Konferenzen über die Wikipedia-Abläufe – kann die Wikipedia es sich nicht leisten, auf die Mitarbeit der Wissenschaft zu verzichten. Dies würde sie über kurz oder lang traditionellen Enzyklopädien – die auf wissenschaftliche Mitarbeit zurückgreifen – unterlegen machen. Man könnte in diesem Zusammenhang von einer Sammlung Laienwissens sprechen – eine gigantische Sammlung zwar, aber nichts desto trotz Laienwissen, das von wissenschaftlicher Seite nicht unbedingt akkreditiert ist. Eine umfassende Restrukturierung ist also vonnöten, unter Umständen auch gegen die Meinung der Gemeinde.

Nahe liegende Probleme wie der oft befürchtete Vandalismus kommen erstaunlicherweise seltener vor, als allgemein angenommen. Umfassende technische Maßnahmen zur Artikelsicherung und Versionshistorie geben die Möglichkeit, angerichtete Schäden schnell zu beheben. Zwar können Benutzer nur aufgrund schwerwiegender Vergehen ausgeschlossen werden, trotzdem hält sich der Missbrauch in Grenzen. Die Selbstregulierungskräfte der Wikipedia beheben Schäden schnell und oft dauerhaft.

Das Anwenden der GFDL als Lizenz für die Inhalte erschwert den Umgang mit den freien Texten. Die GFDL – ursprünglich als Lizenz für Softwaredokumentation vorgesehen – bietet der Wikipedia immer weniger Spielraum. Sie wird im Allgemeinen als zu sperrig angesehen, da sie das Weiterverbreiten der Texte in anderen freien Projekten erschwert. Auch muss sie bei jeder Verwendung eines Artikels außerhalb der Wikipedia mit angegeben werden. Die neun Buchseiten umfassende Lizenz sprengt so oft genug den Rahmen eines einzelnen Artikels. Der Text der Lizenz liegt weiterhin nur in englischer Sprache vor. Rechtlich gültige Übersetzungen gibt es bisher keine. Dies könnte zu Problemen führen, da die Wikipedia mittlerweile über 100 Sprachen unterstützt, in denen bereits Artikel verfasst wurden. Bestrebungen, die Lizenz zu reformieren, sind zwar im Gange, können jedoch nur von der Free Software Foundation durchgeführt werden.

Das letzte Problem, das ich hier anführen möchte, ist die Finanzierung der Wikipedia. Aufgrund immer höher steigender Zugriffszahlen müssen immer mehr Server angeschafft werden, um die reibungslose Benutzung der Wikipedia zu gewährleisten. Die dafür ausgegebenen Mittel steigen jedes Quartal an. Allein im ersten Quartal 2006 wurden 44 Server im Wert von ca. 138.000 \$ erworben⁸. Die Wikimedia Foundation finanziert sich als Non-Profit-Organisation weitestgehend über Spendengelder. Zurzeit reicht das noch aus, aber der Zeitpunkt wird bald kommen, da sich die Stiftung etwas Neues einfallen lassen muss. Zahlreiche, immer dringlicher wirkende Aufrufe zum Spenden machen dies deutlich. Das Ziel der Wikimedia

⁷ <http://www.wikimediafoundation.org>

⁸ http://wikimediafoundation.org/wiki/Resolution_Hardware_Purchase_Feb-06

Foundation wird auch in Zukunft sein, keinerlei Werbung zuzulassen. Um dies nachhaltig zu erreichen, muss sich die Stiftung verstärkt an Großspender halten. Aktuelle Verhandlungen mit Google gehen in diese Richtung, dürfen aber nicht die einzige Maßnahme sein.

5 Folksonomy

Der Begriff Folksonomy ist ebenfalls eine Wortkreuzung, und zwar aus den Worten "folk" und "taxonomy". Die Bedeutung kann in etwa mit "Klassifizierung durch die Leute" übersetzt werden. Im Hinblick auf das World Wide Web ist darunter eine Art gemeinschaftliche Indexierung zu verstehen, bei der bestimmte Objekte wie z.B. Blogeinträge, Fotos oder Lesezeichen mit Deskriptoren - den so genannten "Tags" - versehen werden. Bestimmte Regeln zur Indexierung bestehen meist nicht, es wird davon ausgegangen, dass sich bei einer möglichst großen Nutzergruppe sinnvolle und beschreibende Allokationen ergeben. Tags werden zudem nicht hierarchisch aufgebaut, um die Struktur so einfach wie möglich zu halten. Die Motivation hinter einer solchen gemeinschaftlichen Indexierung ist eine bessere Kategorisierung von Informationen, was im ständig wachsenden Web einen enormen Vorteil für den Anwender bedeuten kann. Dies wird unter anderem dadurch erzielt, dass diejenigen, die die Informationen kategorisieren sie auch benutzen. Außerdem sollen Zusammenhänge, die dem Einzelnen nicht auffallen sichtbar gemacht werden.

Ein großer Nachteil beim "taggen" wird der Ungenauigkeit der Schlagwörter zugeschrieben. Synonyme wie z.B. Holland und Niederlande werden nicht in Zusammenhang gebracht, es gibt Tags mit gleicher Bedeutung wie z.B. TV und Fernsehen und die Begriffe können im Singular sowie Plural vermerkt sein. Zudem kommen noch die landessprachlichen Unterschiede, um nur einige Fallstricke dieser Technik zu nennen.

Die Anwendung dieser Art von Klassifizierung ist vielseitig. Der wohl größte Social Bookmarking Dienst del.icio.us prägte wohl als einer der ersten bereits Ende des Jahres 2003 das "Tagging" in seinem Webdienst, um die Lesezeichen seiner Benutzer zu indexieren. Eine weitere sehr bekannte Webanwendung und Pionier auf dem Gebiet des kollaborativen Klassifizierens startete im Februar 2004 mit dem Namen Flickr. Hier werden Tags dazu verwendet, digitale Bilder mit Metadaten anzureichern. Es spielt dabei keine Rolle, von wem das Bild eingestellt wurde. Jeder Anwender ist berechtigt, Bilder mit Tags zu ergänzen, die in seinen Augen essentiell sind. Je mehr Menschen sich daran beteiligen, umso präziser werden die Bilder katalogisiert.

Die Bedeutsamkeit und das Potenzial der Folksonomy-Bewegung wird durch den Kauf der Webdienste del.icio.us und Flickr im Jahr 2005 durch den Suchmaschinenriesen Yahoo! hervorgehoben.

6 RSS – Really Simple Syndication

RSS ist eine Technik zum Abonnieren von Informationen, die auf Webseiten veröffentlicht werden. Zu diesem Zweck bietet der Anbieter – meist eine Nachrichtenseite oder ein Weblog – einen so genannten Webfeed (auch RSS-Feed oder RSS-Channel genannt) in Form einer XML-Datei an. Der Abonnent kann sich diese Datei herunterladen oder die Informationen direkt in eigene Webseiten integrieren. Um die Informationen unterschiedlicher RSS-Feeds zu sammeln und strukturiert darzustellen wurden so genannte Aggregatorprogramme oder Feed-Reader entwickelt. Die Funktionen dieser Aggregatorprogramme werden mittlerweile auch in E-Mail Programme oder Webbrowser integriert. Der Anwender hat gegenüber der herkömmlichen Informationsgewinnung den Vorteil, dass die Informationen nicht durch zeitaufwendiges Besuchen unterschiedlichster Webseiten gesammelt werden müssen, sondern durch ein zentrales Programm abgerufen und auch archiviert werden können.

Der Name RSS ist ein Akronym, er ist jedoch in den verschiedenen technischen Spezifikationen unterschiedlich definiert. Dies lässt sich vor allem darauf zurückführen, dass verschiedene Entwicklergruppen teilweise unabhängig voneinander unterschiedliche Versionen von RSS herausgaben. Den Anfang machte Netscape im Jahr 1999 mit der Version 0.90, basierend auf dem "Resource Description Framework" (siehe Kapitel 5.2). RSS stand bei dieser Version für "RDF Site Summary". Nur vier Monate später veröffentlichte Dan Libby, der Entwickler der ersten Version die Version 0.91, welche nicht mehr auf RDF sondern auf einer Dokumenttypdefinition für XML basierte und das Format vereinfachte. Nun wurde im Zusammenhang mit RSS von "Rich Site Summary" gesprochen. Trotz der aufkommenden Verbreitung von RSS verlor Netscape das Interesse an dem Format und es bildeten sich zwei unterschiedliche Entwicklungszweige. Der amerikanische Softwarehersteller UserLand entwickelte die Version 0.9x weiter und veröffentlichte im Jahr 2000 seinerseits eine Version 0.91, die nicht mit der vorangegangenen

Spezifikation übereinstimmte. Es folgte die Version 0.92 und Entwürfe für die Versionen 0.93 und 0.94. Ebenfalls im Jahr 2000 veröffentlichte eine unabhängige Entwicklergruppe RSS 1.0. Diese Version basiert erneut auf RDF. Im Jahr 2002 veröffentlichte UserLand ihre Weiterentwicklung der 0.9x-Reihe mit der Version 2.0. Diese Version wird als "Really Simple Syndication" bezeichnet.

Aus dem Bedürfnis heraus, die Vorteile aus den verschiedenen RSS-Versionen mit zusätzlichen Weiterentwicklungen zu einem neuen Standard zu vereinen entstand 2003 ein neues Format welches nach mehreren Namenswechselln "Atom" genannt wurde. Schon in der Version 0.3 trieb Google mit seinen Produkten die Verbreitung des Standards voran, die Version 1.0 aus 2004 wurde im Dezember 2005 zu einem offiziellen Internetstandard.

7 Desktop Online

Unter dem Begriff "Desktop Online" soll nun beschrieben werden, wie der Desktop eines einzelnen Rechners mit dem Internet zusammenwächst. Seit einiger Zeit gibt es Bestrebungen, wieder zurück zur Server – Terminal-Architektur zu kommen. Server im Internet stellen Dienste – so genannte Webservices – zur Verfügung, die mit dem Browser in Anspruch genommen werden können. Die Benutzer speichern Daten nicht mehr lokal, der Speicherplatz wird ins Internet ausgelagert: Es gibt spezielle Fotodienste wie Flickr, die privaten Benutzern Online-Fotogalerien zur Verfügung stellen. Die Benutzer können ihre Bilder in Flickr einstellen und dort anderen Benutzern verfügbar machen.

Der eMaildienst GMX bietet angemeldeten Kunden ein "Media Center", in dem die Kunden Bilder, Videos und Musikdateien speichern und verwalten können. Über so genanntes Streaming können die Inhalte in die Media Center anderer kopiert werden – intern wird nur ein Link gesetzt, um das Datenaufkommen zu beschränken. Andere Dienste wie Streamload haben sich darauf spezialisiert, "Onlinefestplatten" bereit zu stellen, teilweise mit Speicherplatz im Gigabytebereich, teilweise ohne Mengenbegrenzung. Wie viele andere Web 2.0 Dienste sind auch diese Dienste zunächst kostenlos. Nur wer mehr will als einen Standardatz an Optionen, muss dafür bezahlen. So werden die Benutzer motiviert, die Angebote auszutesten. Viele Menschen machen mittlerweile von diesen Angeboten Gebrauch. Für viele bedeutet diese "Onlinefestplatte" die Möglichkeit, ein Backup sensibler Daten zu machen, um sich zum Beispiel bei einem Hausbrand abzusichern. Andere Anwender lagern weniger eigene Daten im Internet aus: Die Warenszene profitiert stark von diesen Angeboten, die auch eMaildienste mit gigantischen Postfächern wie Gmail (2 Gigabyte) oder GMX (1 Gigabyte) einschließen.

In die andere Richtung gehen Programme, die von außerhalb auf den lokalen Rechner zugreifen. Googles Desktop Suche ist ein gutes Beispiel dafür. Sie ermöglicht es dem Benutzer, über ein kleines, zu installierendes Programm, Googles ausgereifte Suchalgorithmen zu verwenden, um den Datenbestand auf dem heimischen Rechner zu durchsuchen.

7.1 Ajax

Viele dieser Webanwendungen benutzen eine Technologie, die Ajax genannt wird: Asynchronous JavaScript And XML. Diese Technologie eröffnet die Möglichkeit, vom bisherigen Blättern durch Internetseiten wegzukommen. Durch sukzessives, vom Anwender interaktiv beeinflusstes Nachladen von Informationen über speziell formatierte XML-Dateien lösen sich Webanwendungen immer mehr vom zustandsbezogenen Datenaustausch hin zum zustandslosen, von Desktopprogrammen gewöhnten Verhalten. Der User surft nicht mehr über verschiedene Seiten, sondern bleibt auf einer Seite, die durch ständiges Aktualisieren im Hintergrund ihr Aussehen an die Handlungen des Anwenders anpasst.

Einer der Vorreiter dieser Technologie ist Gmail⁹, der schon erwähnte E-maildienst von Google. Gmail orientiert sich an traditionellen Desktop-eMailclients wie MS Outlook oder Mozilla Thunderbird. Der Benutzer kann komfortabel auf seine eMails zugreifen, hat ein Adressbuch und eine Rechtschreibprüfung. Eine Besonderheit ist, dass auf viele dieser Funktionen auch per Hotkey zugegriffen werden kann. Durch hauptsächlich clientseitige Ausführung bleibt die Geschwindigkeit konstant vergleichbar mit der von normalen Desktopanwendungen. Weitere Optionen sind frei konfigurierbare eMailfilter und verschiedene Ansichten für erhaltene eMails durch mehrfache Kategorisierung. So soll die Übersicht gewahrt bleiben.

⁹ Gmail: <http://www.gmail.com>

Ein anderes Programm, das Ajax verwendet, ist AjaxWrite¹⁰, eine Online-Textverarbeitung. AjaxWrite orientiert sich stark an bekannten Textverarbeitungen wie MS Word, OpenOffice.Org oder StarOffice. Es unterstützt dabei wichtige Optionen wie Textformatierung (Schriftart, Schriftgröße, Farbe, Stil) und Ausrichtung. Listen anzulegen ist möglich wie in den vergleichbaren Desktopanwendungen, eine Undo- und Redo-Funktion unterstützt komfortables Arbeiten.

AjaxWrite kann verbreitete Textformate wie MS Word doc, OpenOffice.org's odt oder RichTextFormat speichern und öffnen. Die Kompatibilität zu den Desktopprogrammen ist recht gut, die meisten einfachen Funktionen und Textauszeichnungen werden unterstützt. Des Weiteren ist der Export in Adobes PDF-Format möglich.

Beim Öffnen von Dokumenten werden diese auf den Server geladen, der den Browser mit einem Document Object Model versorgt, das die Datei repräsentiert. Änderungen werden im Hintergrund ständig zwischen Server und Client synchronisiert. Das Speichern der Datei auf der lokalen Platte verhält sich wie ein Download, d.h. der Server bereitet eine Datei vor, die den aktuellen Bearbeitungsstand darstellt und der Browser holt diese Datei per `http_get` ab.

Weitere Dienste, die Ajax verwenden, sind Google Suggest, ein Feature in der Google-Suche, das dem Anwender mögliche Suchbegriffe anhand des bereits eingegebenen Begriffes vorschlägt, oder Google Maps, mit dem man innerhalb eines interaktiven Atlas auch Satellitenbilder ansehen kann. Wie man an den Beispielen erkennen kann, ist vor allem der Informationsriese Google an der Verwendung und Weiterentwicklung von Ajax interessiert.

7.2 Privates Grid-Computing

Ein anderer Aspekt des Themas ist das private Grid-Computing. Millionen Rechner stehen täglich lange Zeit still bzw. schöpfen ihre Leistung nicht komplett aus. Bürorechner, die während der Mittagspause nicht benutzt werden und im normalen Gebrauch nur ein einfaches Officeprogramm laufen lassen, sind gemessen an ihrer Laufzeit nie wirklich ausgelastet. Einige weltweite Projekte machen sich diese ungenutzte Rechnerzeit zu Eigen, um damit weit reichende Berechnungen durchzuführen, die sonst teure Supercomputerzeit beanspruchen würden.

Dazu können sich Benutzer Programme herunterladen, die dann als Daemon oder Bildschirmschoner angemeldet werden. Diese Programme holen sich über das Internet Datenpakete, die sie dann in freier CPU-Zeit auswerten und nach Auswertung wieder an den Server zurücksenden, um sich das nächste Paket zu holen.

Dass dieses „private“ Grid-Computing Sinn machen kann, zeigen Projekte wie Seti@Home¹¹ oder Folding@Home¹². Seti@Home wird von der University of California, Berkley entwickelt und betrieben. Es dient der Suche nach extraterrestischem Leben. Dazu werden Emissionen, die ein Radioteleskop empfangen hat, in kleine Datenpakete unterteilt und den privaten Rechnern zur Auswertung überlassen. Dabei werden Gaußsche Anstiege und Gefälle, die auf Radioquellen hinweisen, und verschiedene Pulsfolgen (Triplet) gesucht, die Hinweise auf extraterrestisches Leben sein könnten. Die durchschnittliche Rechenleistung der verteilten privaten Computer beläuft sich auf ca. 200 Teraflops (vgl. BlueGene/L mit 280 Teraflops, schnellster Supercomputer der Welt). Bisher konnte eine Rechenzeit von knapp 2,3 Millionen Jahren erzielt werden.

Folding@Home ist ein anderes Projekt, das von der Stanford University betreut wird. Nach gleichem Muster wird hier die Faltung von Proteinen berechnet. Faltung von Proteinen meint hier den räumlichen Aufbau von Aminosäuren, die sich auf verschiedene Weisen zu einem Protein zusammensetzen können. Der fehlerhafte Aufbau eines Proteins soll Ursache für Krankheiten wie Krebs, Alzheimer und BSE sein. Die Ergebnisse der Forschung werden über die Stanford University der Öffentlichkeit frei zugänglich gemacht.

7.3 Privatsphäre & Sicherheit

Das wohl größte Problem der Verschmelzung von Desktop und Internet ist die Privatsphäre. Wie kann der

¹⁰ AjaxWrite: <http://www.ajaxwrite.com>

¹¹ <http://setiathome.ssl.berkeley.edu/>

¹² <http://folding.stanford.edu/>

Benutzer sicher sein, dass niemand anderes an seine Daten kommt? Die Vertraulichkeit der Daten muss in jedem Fall gewährleistet werden, völlig unabhängig von ihrem Aufbewahrungsort. Daten, die lokal gespeichert werden, sind relativ einfach zu schützen. Daten, die hingegen aufwendig zwischen Server und Client transportiert werden, lassen sich oft einfach abfangen. Hier ist es nötig, die Kommunikation sicher zu gestalten, beispielsweise über verschlüsselte Kanäle (SSL, TLS, SSH). Weiterhin wird es nötig sein, die Authentifizierung des Benutzers sicher zu handhaben. Private Daten gehen nur den Autor bzw. Besitzer etwas an, andere Benutzer des Systems dürfen darauf keinen Zugriff erlangen, indem sie beispielsweise Bugs in der Software ausnutzen, sei es bewusst oder unbewusst.

Auch die geographische Lage der Server darf nicht außer Acht gelassen werden. In den USA gelten seit jeher und besonders seit den letzten Jahren weitaus schwächere Datenschutzgesetze als in der EU. Wie wird damit umgegangen, vor allem vor dem Hintergrund ausländischer Daten (Privatdaten beispielsweise eines Europäers, die in den USA gespeichert werden)? Bevor hier nicht wirkliche Aufklärungsarbeit bzw. Garantie der Privatsphäre steht, wird sich die Mehrzahl der Anwender diesen Systemen nicht anvertrauen.

8 Semantic Web

Die Inhalte des World Wide Web wurden über Jahrzehnte optimiert, um von Menschen meist visuell erfasst zu werden. Die Auswertung und Interpretation dieser Inhalte durch Computerprogramme gestaltet sich hingegen nach wie vor als schwierig. Zwar können Maschinen Überschriften, Paragraphen oder Links erkennen, jedoch bleibt ihnen die Bedeutung meist verschlossen. An dieser Stelle setzt die Idee des Semantic Web an. Es ist nicht etwa ein neues, sondern eine Erweiterung des bestehenden Netzes. Die Informationen – in diesem Fall Webseiten bzw. Teile davon – werden mit Metadaten angereichert, die die Bedeutung des Inhaltes für Programme begreiflich machen sollen. Auf diese Weise werden Informationen gleichgestellt - unabhängig davon, ob sie für Mensch oder Maschine bestimmt sind.

Die erweiterbare Markup-Sprache XML und das Ressourcenbeschreibungssystem RDF sind zwei wichtige Werkzeuge für die Entwicklung des Semantischen Netzes, auf die im Folgenden näher eingegangen wird.

8.1 XML

Ausgezeichnet werden die Daten in XML wie auch in anderen Auszeichnungssprachen mit Tags. Tags beinhalten Informationen über die von ihnen eingeschlossenen Daten, sind also so genannte Metadaten. Im Gegensatz zu der im World Wide Web stark verbreiteten Auszeichnungssprache HTML gibt es in XML keinen endlichen, bestimmten Satz von Tags für die Strukturierung eines Dokuments. Es können also Tags selbst definiert werden, um die Struktur oder Semantik von Elementen eines Dokuments zu bestimmen. XML kann also als ein Satz von Regeln verstanden werden, mit denen es möglich ist eigene Auszeichnungssprachen zu kreieren, die in selbst definierten Tags angeordnet sind und aus selbst definierten Strukturen bestehen.

XML-Dokumente sind ähnlich wie HTML- oder XHTML-Dokumente hierarchisch aufgebaut und weisen eine Baumstruktur auf. Als Knoten des Baumes können auftreten:

- Elemente – in Form von Start- und Endtag (`<tag-name>`,`</tag-name>`), bzw. Empty-Element-Tag (`<tag-name />`)
- Attribute – Zusatzinformation vorkommend als Wertepaar im Start- oder Empty-Element-Tag (`<tag-name attribut-name="attribut-wert" />`)
- Steueranweisungen – `<?name daten?>`
- Kommentare – `<!-- Kommentar -->`
- Text – repräsentiert als normaler Text oder als CDATA-Abschnitt (`<![CDATA[text]]>`)

CDATA ist ein Akronym für "Character Data". Man verwendet CDATA-Abschnitte um gewisse Textbereiche vom Auswerten eines Parsers zu schützen. Dieser Fall tritt ein, wenn ein Textknoten Markups enthält, die nicht als solche ausgewertet werden sollen.

Es wird im Allgemeinen von der Wohlgeformtheit eines XML-Dokuments gesprochen, wenn das Dokument die Auszeichnungsregeln des W3C-Standards befolgt. Dazu gehören unter anderem:

- Eine gültige Deklaration von folgender, exemplarischer Form:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
```

Das Attribut "version" muss als Wert die Nummer des verwendeten Auszeichnungsstandards beinhalten, als Wert der Variablen "encoding" wird der verwendete Zeichensatz angegeben und die Variable "standalone" gibt an, ob eine Dokumenttypdefinition vorliegt oder nicht.

- Elementnamen (Tagnamen) müssen mit einem Buchstaben, Doppelpunkt oder Unterstrich, dürfen aber nicht mit den Teilstrings "xml" oder "XML" beginnen, im weiteren Verlauf des Namens dürfen auch Zahlen, Bindestrich und Punkt verwendet werden.
- Attributnamen unterliegen der gleichen Konvention wie Elementnamen, Attributwerte müssen von Anführungszeichen umschlossen sein.
- Eine korrekte Hierarchie. Es muss genau ein Wurzelement im Hierarchiebaum des XML-Dokuments vorhanden sein. Desweiteren dürfen sich Elemente unterschiedlichen Typs nicht überlappen, Verschachtelungen hingegen sind legitim.

Für die vollständige Beschreibung des XML-Standards sei an dieser Stelle auf die XML-Spezifikation¹³ des W3C verwiesen.

Da die Tags in XML nicht wie beispielsweise in HTML meist eine bestimmte Visualisierung ihres Inhalts zur Folge haben, können Webseiten so mittels XML mit Metadaten angereichert werden, ohne das Erscheinungsbild einer Webseite zu verändern.

Programme, die mit Metadaten angereicherte Webseiten verarbeiten sind in der Lage durchaus anspruchsvolle Aufgaben zu lösen. Wird durch Tags z.B. die Postleitzahl innerhalb einer Adresse ausgezeichnet, so ist es einer Maschine – verknüpft mit den nötigen Datenbanken – möglich, die Entfernung zu dieser Adresse zu berechnen. Allerdings muss das Programm hierzu die Bedeutung der Tags kennen. Dies wird durch RDF ausgedrückt.

8.2 RDF

Das "Resource Description Framework" wurde vom World Wide Web Consortium entwickelt, um Eigenschaften von Ressourcen im World Wide Web so zu beschreiben, dass sie durch Maschinen verarbeitet werden können.

Der typische RDF-Ausdruck besteht aus einem RDF-Triple (s. Abb. 2). Die Drei Elemente sind Subjekt, Prädikat und Objekt – vergleichbar mit einem einfachen Satz. Die Elemente des Tripels stellen ein Subjekt (zum Beispiel ein Dokument oder eine Person) in einer bestimmten Relation ("wurde verfasst von", "ist Autor von", "ist verheiratet mit") zu einem Objekt (z.B. einer Webseite oder einer anderen Person).



Abbildung 2: Graph eines RDF Tripels

Subjekt, Prädikat und Objekt werden meist als "Uniform Resource Identifier" (kurz URI) angegeben. Zudem können Objekte auch als leere Ressource oder Literal angegeben werden. Ein Literal ist die einfachste Form, da nur ein Wert zugeordnet wird. Abbildung 2 enthält ein solches Literal – "Max Mustermann". Ressourcen werden in der grafischen Darstellung als Ellipsen dargestellt, Relationen als benannte Kanten und Literale als Rechtecke. Weil ein Programm die für den Menschen gleichwertigen Zusammenhänge wie beispielsweise "Mitglied im W3-Consortium ist Tim Berners-Lee" und "Tim Berners-Lee ist Mitglied im W3-Consortium" nicht ohne weiteres als solche erkennen kann, muss die Relation im Graph zusätzlich eine gerichtete Kante sein. Um den in der menschlichen Sprache häufig vorkommenden Bedeutungsverschiebungen vorzubeugen werden verschiedene URIs für die verschiedenen Bedeutungen eines Wortes verwendet. Auf diese Weise lassen sich Verwechslungen z.B. zwischen einer Parkbank und einem Geldinstitut vermeiden.

¹³ XML-Spezifikation: <http://www.w3.org/TR/2004/REC-xml-20040204/>

Das RDF-Modell kann auf verschiedene Arten dargestellt werden. Am weitesten verbreitet ist die Repräsentation in XML, der Urvater des World Wide Web – Tim Berners-Lee – entwickelte jedoch eine äquivalente aber leichter verständliche Form der Darstellung genannt "Notation 3". Im Folgenden werden beide Varianten anhand eines kleinen Beispiels vorgestellt:

```
<http://foo.bar/doc.odt> <http://purl.org/dc/elements/1.1/creator> <mailto:st@foo.bar>
```

Listing 1: Beispiel von RDF in N3

Zum reinen Verständnis ist die Repräsentation der RDF-Tripel in N3 sicherlich besser geeignet, da Subjekt, Prädikat und Objekt einfach durch Leerzeichen getrennt hintereinander geschrieben werden. Die verbreitetste Repräsentation ist jedoch die in XML-Syntax – nicht zuletzt deshalb, weil sie sich leicht in bestehende XML oder (X)HTML-Dokumente einbinden lässt und Dokumente so sehr einfach um Metadaten erweitert werden können.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://foo.bar/doc.odt">
    <dc:creator>mailto:st@foo.bar</dc:creator>
  </rdf:Description>
</rdf:RDF>
```

Listing 2: Beispiel von RDF in XML-Syntax

Der Beispielcode beginnt mit einer "Processing Instruction", die das Dokument als XML-Dokument kennzeichnet. In den folgenden Zeilen werden externe XML-Namensräume (xmlns = XML Name Space) an selbst gewählte Bezeichner (rdf, dc) geknüpft. Anschließend folgt das RDF-Tripel. An dieser Stelle sei darauf hingewiesen, dass das W3C abweichend von dieser, der so genannten "Serialisation Syntax", parallel die so genannte "Abbreviated Syntax" vorschlägt, bei der alle Metadaten als Attribute innerhalb der Tags angegeben werden um Problemen bei der Einbettung in HTML-Dokumente vorzubeugen.

In den vorhergehenden Beispielen wird ein Subjekt mit der URL (einer Spezialform der URI) "http://foo.bar/doc.odt" in Relation (http://purl.org/dc/elements/1.1/creator) mit einem Objekt mit der URI "mailto:st@foo.bar" gesetzt. Die bei diesen Beispielen benutzte Relation "creator" ist Teil eines Metadaten-Schemas namens "Dublin Core" der "Dublin Core Metadata Initiative" (im Weiteren DCMI). Die DCMI besteht seit 1994 und hat es sich zur Aufgabe gemacht, Webressourcen durch Beschreibungen zu kategorisieren. Das Ergebnis der im März 1995 in Dublin/Ohio tagenden Konferenz zu diesem Thema wurde dem Tagungsort entsprechend "Dublin Core metadata" genannt. Das Prädikat "creator" sollte nach der DCMI eine natürliche Person sein, die verantwortlich für die aktuelle Fassung des Dokuments ist. Der Verfasser oder Herausgeber des Dokuments steht in der Relation "publisher" und Mitwirkende in der Relation "contributor", um nur einige Beispiele zu nennen.

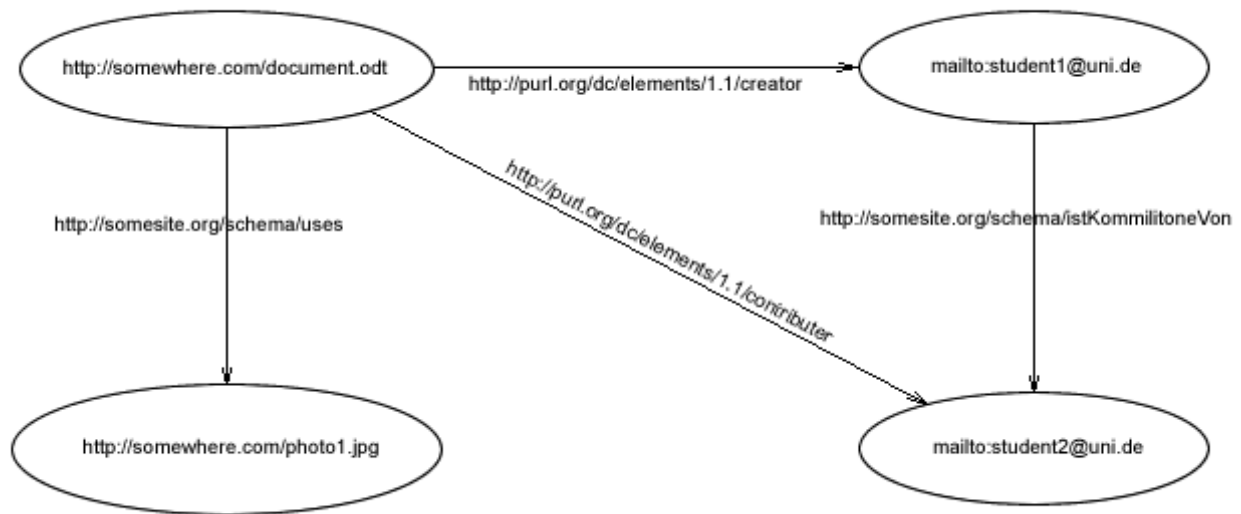


Abbildung 3: Graph eines RDF-Beispiels

Abbildung 3 soll einen kleinen Einblick verschaffen, auf welche Weise Ressourcen miteinander verknüpft werden können. Obwohl in der Realwelt häufig eine viel höhere Vernetzung zwischen Ressourcen herrscht lässt sich erahnen, welches Potenzial in einer Vernetzung von Metadaten steckt. Folgend der zum Graphen gehörende RDF-Code:

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:bsp="http://somesite.org/schema/">
  <rdf:Description rdf:about="http://somewhere.com/document.odt">
    <dc:creator>mailto:student1@uni.de</dc:creator>
    <dc:contributer>mailto:student2@uni.de</dc:contributer>
    <bsp:uses>http://somewhere.com/photol.jpg</bsp:uses>
  </rdf:Description>
  <rdf:Description rdf:about="mailto:student1@uni.de">
    <bsp:istKommilitoneVon>mailto:student2@uni.de</bsp:istKommilitoneVon>
  </rdf:Description>
</rdf:RDF>
```

Listing 3: RDF Beispiel [6]

Mittels XML und RDF lassen sich also sehr gut Zusammenhänge zwischen und zusätzliche Informationen über Ressourcen beschreiben. Dies ist nicht nur wichtig, um das weltweite Datennetz aus Sicht der normalen Nutzer zu verbessern, sondern um es auch bisher stark benachteiligten Benutzergruppen wie z.B. Blinden leichter zugänglich und benutzbar zu machen.

9 Kollektive Intelligenz

Den beschriebenen Wandel vom statischen zum dynamischen Internet kann man als eine Art Paradigmen-Wechsel von kollektivem Wissen zur kollektiven Intelligenz auffassen. Kollektives Wissen beschreibt die Vergangenheit: Das Internet bestand aus statisch definierten Webseiten, auf denen Kundige ihr Wissen darstellten. Die einzelnen Seiten waren relativ von einander abgegrenzt, bildeten also Wissensseinheiten, die keinerlei (oder vielmehr kaum) Interaktion aufwiesen. Die Definition von Wissen in diesem Zusammenhang meint faktisch und sachlich wahre Informationen über reale Zusammenhänge, die in sich stimmig sind. Dieses Wissen bedeutet zunächst, dass die Informationen bewiesen oder beweisbar, aber untereinander nicht zwangsläufig verknüpft und von statischer Natur sind.

Mittlerweile jedoch darf man im Zusammenhang mit dem Internet nicht mehr von (kollektivem) Wissen sprechen, sondern muss die Gesamtheit der Informationen als (kollektive) Intelligenz bezeichnen (kollektiv deshalb, weil das Internet als solches eine Vernetzung von Individualitäten darstellt). Intelligenz meint hier, dass die vorhandenen Information oder das vorhandene Wissen nicht mehr nur statisch vorhanden sind, sondern sich – eigenständig oder von außerhalb – verknüpfen lassen, um somit neue Informationen und somit neues Wissen aus altbekanntem zu dedizieren. Der neue, dynamische Charakter des Internets – Wikis, Weblogs (mit Trackbacking) – schafft diese Verknüpfungen. Die Fähigkeit des Internets – welches hier nicht nur die rein technische Seite, sondern auch die soziale Seite einschließt – diese Verknüpfungen aufzubauen, zu erweitern, zu verwerfen und zu aktualisieren kann man durchaus als Intelligenz bezeichnen, da genau dies die Definition von Intelligenz ist: „*Intelligenz (lat.: intelligentia „Einsicht, Erkenntnisvermögen“, intellegere „verstehen“) bezeichnet im weitesten Sinne die Fähigkeit zum Erkennen von Zusammenhängen und zum Finden von optimalen Problemlösungen.*“¹⁴

10 Zusammenfassung

Seit Entstehung der ersten Webseiten hat sich vieles im World Wide Web verändert. Zum ersten mal technisch revolutioniert wurde das Netz der Netze vor einigen Jahren durch datenbankbasierte Content Management Systeme, doch damals wagte niemand von einer neuen Version des Webs zu sprechen. Erst durch neue Modelle, veränderte Regeln und konsequente Anwendung von Techniken, die nachhaltig zur Verbesserung des Webs beitragen, wird bei den Nutzern ein Gefühl von Veränderung ausgelöst. Das Schlagwort „Web 2.0“ ist nun kein von Soziologen und Informatikern ausgelöster Hype mehr, es bezeichnet eine Entwicklung, die das Netz tatsächlich durchlaufen hat und noch durchläuft. Das Web 2.0 ist in der Gesellschaft angekommen. Inwiefern es die Gesellschaft verändern wird, bleibt abzuwarten. Sicher ist nur, dass es kein neuer, statischer Zustand ist.

11 Literatur

- [1] Berners-Lee, Tim; Hendler, James; Lassila, Ora: The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American, <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>, 2001
- [2] Bosak, Jon; Bray, Tim: Mehr Tempo auf der Datenautobahn. Spektrum der Wissenschaft 7/1999, 1999, S. 74
- [3] Bowman, Shayne; Willis, Chris: We Media – How audiences are shaping the future of news and information. 2003, S. 9
- [4] Bray, Tim: What is RDF. <http://www.xml.com/pub/a/2001/01/24/rdf.html>, 2001
- [5] Carvin, Andy; Berners-Lee, Tim: Weaving a Semantic Web. <http://www.digitaldivide.net/articles/view.php?ArticleID=20>, 2005
- [6] Champin, Pierre-Antoine: RDF Tutorial. <http://www710.univ-lyon1.fr/~champin/rdf-tutorial/rdf-tutorial.html>, 2001
- [7] Kleinz, Thorsten: Fünf Herausforderungen für die Wikipedia. Telepolis, <http://www.heise.de/tp/r4/artikel/21/21787/1.html>, 2006
- [8] Möller, Eric: Das Wiki-Prinzip: Tanz der Gehirne Teil 1. Telepolis, <http://www.heise.de/tp/r4/artikel/14/14736/1.html>, 2003
- [9] Möller, Erik: Die heimliche Medienrevolution - Wie Weblogs, Wikis und freie Software die Welt verändern. Hannover, 2005
- [10] Pilgrim, Marc: What is RSS. <http://www.xml.com/pub/a/2002/12/18/dive-into-xml.html>, 2002
- [11] Seeboerger-Weichselbaum, Michael: JavaScript – Das bhv Taschenbuch, 2000
- [12] Sixtus, Mario: Das Web sind Wir. TECHNOLOGY REVIEW 7/2005; 2005
- [13] Sixtus, Mario: Flickr, Fotos, Folksonomy. http://www.sixtus.net/article/608_0_2_0_C; 2005
- [14] Sixtus, Mario: Massenmedium - Blogosphäre: Kommunikationsgeflecht und Marketingfaktor. c't 19/05, 2005
- [15] W3C: Extensible Markup Language (XML) 1.0 (Third Edition). <http://www.w3.org/TR/REC-xml>,

¹⁴ Aus: [19]

2004

- [16] Wikipedia: Dublin Core. http://de.wikipedia.org/wiki/Dublin_Core, Stand 11.06.2006
- [17] Wikipedia: Folding@Home. http://de.wikipedia.org/wiki/Folding_at_home, Stand 19.06.2006
- [18] Wikipedia: Gemeinschaftliches Indexieren.
http://de.wikipedia.org/wiki/Gemeinschaftliches_Indexieren, Stand 19.05.2006
- [19] Wikipedia: Intelligenz. <http://de.wikipedia.org/wiki/Intelligenz>, Stand 19.06.2006
- [20] Wikipedia: Resource Description Framework.
http://de.wikipedia.org/wiki/Resource_Description_Framework, Stand 11.06.2006
- [21] Wikipedia: RSS. <http://de.wikipedia.org/wiki/RSS>, Stand 12.05.2006
- [22] Wikipedia: RSS (file format). http://en.wikipedia.org/wiki/RSS_%28file_format%29, Stand 12.05.2006
- [23] Wikipedia: Seti@Home. http://de.wikipedia.org/wiki/Seti_at_home, Stand 19.06.2006
- [24] Wikipedia: Trackback. <http://de.wikipedia.org/wiki/Trackback>, Stand 10.06.2006
- [25] Wikipedia: World Wide Web. http://de.wikipedia.org/wiki/World_Wide_Web, Stand 15.06.2006
- [26] Wikipedia: Wikipedia. <http://de.wikipedia.org/wiki/Wikipedia> , Stand 16.06.2006
- [27] Wrede, Oliver: Weblogs: DJing mit News.
http://weblogs.design.fh-aachen.de/owrede/publikationen/weblogs_debug, 2001