

Steffen Staab

C 7 Web Science

C 7.1 Einleitung

Das World Wide Web ist ein Informationsmedium, das von seinen Benutzern gleichermaßen und quasi gleichzeitig geschaffen, konsumiert und modifiziert wird, mit der Konsequenz, dass dabei die Grenzen zwischen *Informationskonsum*, *-produktion* und *-modifikation* verwischt werden. Dies gilt ansatzweise bereits für angebotsorientierte Web-Seiten, die oft gar nicht statisch sind, sondern permanent weiterentwickelt und weitergeschrieben werden. Noch stärker gilt diese Aussage für dynamischere Inhalte wie Wiki-Seiten, Seiten in sozialen Online-Netzwerken, Webseiten mit Empfehlungsfunktionen, Like-Buttons, Kommentarfeldern, Diskussionsforen und ähnlichen Optionen. Alle diese laden den Benutzer ein, die Inhalte einerseits zu ergänzen, andererseits ihr Aussehen und ihre präsentierten Inhalte aufgrund solcher Benutzerinteraktionen anzupassen.

Darüber hinaus ist Konsum im World Wide Web auch Produktion und Modifikation, da die Auswahl des Konsumenten nicht nur lokale Änderungen an einer einzelnen Webseite zu Folge hat, sondern auch über die lokale Seite hinaus im globalen Web die Präsentation von Inhalten beeinflusst. Besonders offensichtlich ist dieser globale Zusammenhang zwischen Konsum, Produktion und Modifikation, wenn in Suchmaschinen die Ergebnisauswahl eines Nutzers die Suchergebnisse aller anderen Nutzer beeinflusst (Lit. 01). Subtiler, aber gleichzeitig viel weitreichender ist aber die Tatsache, dass die Vernetzung von Informationsdiensten, wie zum Beispiel von Suchmaschinen, sozialen Netzwerken, Werbediensten und Betriebssystemen auf mobilen Endgeräten dazu führt, dass die Aktivitäten einer Person auf einer Webseite zur Modifikation der präsentierten Inhalte und Optionen auf anderen Webseiten führen (Lit. 02).

Benutzer des Webs sind gleichzeitig Konsumenten, Produzenten und Modifikatoren von Information und beeinflussen dadurch die Informationspräsentation und die Informationssichten anderer. Daraus folgt, dass Information genauso wenig eine statische Komponente ist, wie es die Web-Plattformen und -Werkzeuge sind, die eingesetzt werden, um Information zu erzeugen und die Information in einen Kontext stellen, der ebenso auf die Interpretation des Informationskonsumenten einwirkt. Dann aber lassen sich Informationszugriff, -verwaltung, und -kontextualisierung nicht mehr statisch beschreiben, sondern es bedarf einer *systemischen Sichtweise*, die untersucht, wie das Web durch seine Benutzung mittels dieser Plattformen und Werkzeuge von seinen Benutzern und seinen Betreibern weiterentwickelt wird. Eine solche Sicht kann aber keine rein technische sein, da die Benutzer durch Wahrgenommenes in ihrem *Verhalten* beeinflusst werden, mit dem sie neue Information konsumieren, produzieren und modifizieren. Dieses Verhalten ist ebenso durch die Kognition des Benutzers geprägt wie durch seine Affekte, seine Sozialisierung sowie durch positive und negative Anreize wie etwa ökonomischen oder sozialen Gewinn. Diese Beeinflussung bleibt nicht rein zufällig, so wie in traditionellen Medien das Vorwissen und die Ansichten der Leser die Interpretation der Information bestimmen, sondern es entstehen systembedingte Strömungen und Entwicklungen, die den Interessen der Benutzer, der Betreiber oder der Gesellschaft zuwiderlaufen oder entgegenkommen können. Berners-Lee et al. (Lit. 03) beschreiben diese Zusammenhänge als „science of decentralised information structures“.

Solche systembedingten Mechanismen sind bisher wenig untersucht. Sehr selten werden diese Mechanismen quantitativ beschrieben, oft ist das auch untersuchungsbedingt nicht möglich, wenn sich Interpretationseffekte mit reichhaltigen soziologischen und psychologischen Effekten vermischen. Zunehmend werden Modelle auch auf verschiedenen Granularitätsstufen entwickelt, die systemische Zusammenhänge von Webkonsum, -produktion und -modifikation beschreiben. Da es um die Verarbeitung und Interpretation von Information geht sowie um die Beeinflussung menschlichen Verhaltens, spielen Effekte eine Rolle, die in verschiedenen wissenschaftlichen Disziplinen untersucht werden. Durch die zentrale Rolle, die diese Effekte in Bezug auf Wirkung und die Rück-

kopplung innerhalb des Systems einnehmen, ist ein *multi- und interdisziplinärer Zugang zur Beschreibung und Modellierung des Systems World Wide Web und seiner Akteure* unerlässlich.

Abbildung 1 fasst zusammen, was *Web Science* abdecken muss: Viele Akteure interagieren auf der individuellen Ebene in *Mikrointeraktionen*. Einerseits spiegeln diese Mikrointeraktionen Aspekte des Benutzerverhaltens wider, wie es durch seine kognitiven, sozialen und ökonomischen Gegebenheiten situiert ist. Andererseits reflektieren diese Mikrointeraktionen auch die Beschränkungen, die vorhandene Daten und Informationen, Anwendungen, Protokolle und Governance des Webs dem Benutzer auferlegen. Beides zusammen führt in der Masse der Benutzer zu *Makroeffekten*.

Definition:

Web Science ist die soziotechnische Wissenschaft, die untersucht, wie (i) neue oder vorhandene Technologien, (ii) Regulierungen im Web und durch Webtechnologien oder Webanwendungen und (iii) Aktivitäten der Webbenutzer das Web formen und verändern und wie umgekehrt das Web auf die Akteure zurückwirkt.

Das Ziel von *Web Science* ist es, Mikrointeraktionen und Makroeffekte zu beschreiben und zu verstehen. Idealerweise erfolgt dieses Verständnis durch quantitativ beschreibende Modelle; wo das nicht möglich ist, zumindest durch qualitative Aussagen über den Zusammenhang zwischen Mikro- und Makroebene.

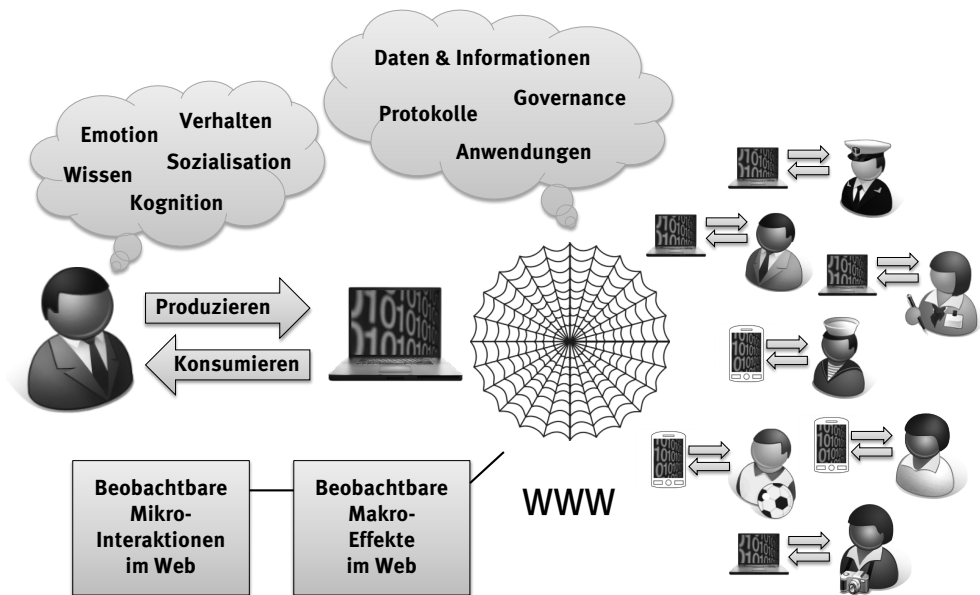


Abb. 1: Mikrointeraktionen führen zu Makroeffekten im Web. *Web Science* strebt an diese Zusammenhänge zu verstehen und zu nutzen (Quelle: URL1)

Im Folgenden wird beispielhaft dargelegt, wie verschiedene Disziplinen das System *Web* und damit Informationskonsum, -produktion und -modifikation prägen und welche Schwierigkeiten und erste Ansätze es gibt, diese zu modellieren.

C 7.2 Webregulierung (Web Governance)

Die unmittelbarste Form der Beeinflussung von Benutzerverhalten ist durch Regulierung gegeben. Im Web und im Internet wird hierbei oft von *Web* bzw. *Internet Governance* geredet: „Internet governance is the development and application by Governments, the private sector and civil society, in their respective roles, of shared principles, norms, rules, decision-making procedures, and programmes that shape the evolution and use of the Internet.“ (Lit. 04)

Im Web gilt natürlich auch die klassische Regulierung durch geltende *Gesetze*. So betreffen Paragraphen des Bürgerlichen Gesetzbuches, das Verträge behandelt, prinzipiell auch den Abschluss von Verträgen im Web, wenn die Vertragspartner deutschem Recht unterliegen. Allerdings lassen sich Gesetze im grenzüberschreitenden World Wide Web oft nur schwierig, manchmal nur teilweise und manchmal auch gar nicht durchsetzen, was vor allem beim Urheberrecht oft deutlich wird. Außerdem verbleibt natürlich die Subsumtion zu leisten, anhand derer die Anwendbarkeit einer Rechtsnorm auf eine Situation im Web beurteilt werden muss, was oft nur nach erfolgter Rechtsprechung durch ein übergeordnetes Gericht möglich ist.

Neben den klassischen nationalen Gesetze und zwischenstaatlichen Vereinbarungen treten im Web grundlegend weitere Regulierungsmechanismen:

Die Regulierung der Web-Infrastruktur

Die Regulierung der Web-Infrastruktur, vor allem des Internets, durch operative und standardisierende Institutionen wie ICANN, IETF, W3C und ähnliche: Hier werden technische Standards vereinbart, die unmittelbar die Handhabbarkeit von Webseiten und -anwendungen betreffen. Zudem verwaltet ICANN die Addressangaben im WWW, z. B. vergibt sie Top-Level-Internetadressen, wie *.name*, *.org* etc. Standardisierungen der Netzwerkprotokolle sind ebenso nötig, wie Kollaboration unter den Internet-Providern. Am Deutlichsten wird ein Fehlen der Kollaboration durch (un)absichtliche Falschkonfiguration. Zum Beispiel führten falsche Angaben der pakistanischen Telekom zur globalen Nichterreichbarkeit von YouTube. Dies wird auch als *IP-Hijacking* beschrieben (URL8).

Die spezifische Wettbewerbssituation

im Bereich der (mobilen) Internet-Provider: Indirekt werden auf diese Weise auch die Verhaltensweisen der Benutzer verändert. Zum Beispiel führen die im Vergleich zu Deutschland hohen Pauschalgebühren (*Daten-Flat-Rates*) für mobilen Datenverkehr in Brasilien im Jahr 2012 zu anderen Benutzungsmustern durch die brasilianische Bevölkerung als in Deutschland im gleichen Zeitraum. Allgemein ist die Infrastruktur in den sich entwickelnden Ländern aufgrund der geringer ausgeprägten Netzwerkeexternalitäten oft weniger gut ausgebaut und teurer als in den Industrieländern mit entsprechenden Folgen einer geringeren Verfügbarkeit des Webs für die Bevölkerung (vgl. Lit. 05).

Die Regulierung der Web-Benutzer durch Software (vgl. Lit. 06, Lit. 07)

Die Software, die verwendet wird (bisweilen als Teil einer dedizierten Hardware), reguliert auf verschiedenen Wegen den Konsum-Produktions-Zyklus. Der Regulierungsaspekt wird unten weiter ausgeführt.

C 7.2.1 Regulierung bzw. Standardisierung durch monopol- und oligopolartige Strukturen

Die Regulierung durch eine Standardisierungsorganisation wie das W3C führt nicht automatisch zum Erfolg einer Webtechnologie. Ein gutes Beispiel stellt hier die Verbreitung von Metadaten in Webseiten dar. Mit der Semantic-Web-Sprache RDF lassen sich Metadaten beschreiben. Eine sprunghaft gestiegene, starke Verbreitung solcher Metadaten in Webseiten erfolgte aber nicht bereits mit der Standardisierung von RDF durch das W3C. Erst durch die Übereinkunft von Google,

Bing (Microsoft) und Yandex, Metadaten für die Präsentation von Suchmaschinenergebnissen zu nutzen (vgl. auch URL2 und Lit. 08), wuchs die Motivation für Webseitenbetreiber, Metadaten als Teil ihrer Strategie für die Suchmaschinenoptimierung zur Verfügung zu stellen, was die entsprechend umfangreiche Verbreitung von Metadaten zur Folge hat.

C 7.2.2 Regulierung durch dedizierte Umgebungen

Auch wenn es im Web an einigen Stellen monopol- oder oligopolartige Strukturen gibt – Suchmaschinen nehmen in vielen Ländern eine solche Stelle ein – so bleibt das Web doch eine *offene Infrastruktur*, in der auch ein Unternehmen mit heute beherrschender Position im Suchmaschinenmarkt wie Google morgen potenziell neue Wettbewerber haben könnte. Zittrain (Lit. 07) beschreibt diese Offenheit – zunächst bei Rechnern im Allgemeinen, dann im Internet und Web im Speziellen – als Hauptantriebsfeder für die Innovationsfreudigkeit im Internet. Mit der Offenheit einher gehen ausgeprägte Sicherheitsprobleme, wie Phishing, Trojaner, Viren etc., die Daten ausspähen und zerstören und auf diese Weise Schaden anrichten.

Dies hat zur Folge, dass es für viele Nutzer eine starke individuelle Motivation gibt, vermeintlich sicherere Plattformen, auf denen nur gewisse Anwendungen ausgeführt werden können, zu bevorzugen: iPad und iPhone mit dem entsprechend von Apple reguliertem Anwendungsplatz sind solche Beispiele. Aber auch auf Plattformen wie sozialen Netzwerkanwendungen sind entsprechende Regulierungen von Anwendungen (Apps) denkbar. Zum Beispiel wäre es möglich, dass Facebook Spieleanwendungen der Firma Zynga bevorzugt, an der Facebook einen erheblichen Anteil besitzt.

Solche *dedizierten Umgebungen* wie das iPhone oder Facebook, in denen bestimmte Anwendungen (Apps) ausgeführt werden, regulieren, indem sie beschränken, (i) wer Apps programmieren darf und (ii) mit welchen Fähigkeiten diese Apps ausgestattet werden dürfen. In einigen Fällen dient dies der Sicherheit der Benutzer, z. B. um das Ausspähen von Daten zu erschweren, in anderen Fällen scheint es aber lediglich dazu zu dienen, die marktbeherrschende Stellung der Firma, die diese Umgebung bereitstellt, zu untermauern, zu monetarisieren und den Benutzer mit seinen Daten innerhalb dieser Umgebung festzuhalten. Ein Wechsel der Umgebung hat ja für den Benutzer oft sehr hohe Migrationskosten zur Folge, zum Beispiel dadurch, dass neue Profildaten, Kontaktdaten etc. neu eingegeben werden müssen.

C 7.2.3 Regulierung in dedizierten Umgebungen

Während wir mit *Regulierung durch dedizierte Umgebungen* die Regulierung der Ausführung von Software ansprechen, die typischerweise automatisiert erkannt und gegebenenfalls unterbunden wird, reden wir von *Regulierung in dedizierten Umgebungen*, wenn gesetzesähnliche Regelungen in solchen Umgebungen die Benutzerinhalte und -aktivitäten betreffen, ohne dass das jeweilige nationale Recht dies nahelegt.

Ursache hierfür sind in der Regel *kulturelle Unterschiede*. Die Darstellung von Nacktheit und Gewalt ist in unterschiedlichen Kulturkreisen verschieden stark verpönt bzw. unterscheidet sich in ihrer Art. Im Sommer 2012 betraf dies zum Beispiel die Darstellung eines Titelbilds des Zeitmagazins auf Facebook, das einen männlichen Penis abbildete, was den Betreiber von Facebook zur Folge zum Abblocken dieses Eintrags führte. Beispiele für Zensur im Netz gibt es nach wie vor genug.

Andere Regelungen betreffen weniger die Inhalte als die Aktivitäten. „Gold farming in massively multiplayer online games“ bedeutet, dass ein Spieler sehr viel Zeit darauf verwendet, seine Spielfigur in einem Online-Spiel so aufzubauen, dass sie in der virtuellen Welt Dinge besitzt, die dort von Bedeutung sind. Dieser Spieler kann diese virtuellen Gegenstände oder Charaktere an-

schließend gegen echtes Geld verkaufen (Lit. 09). Während die Online-Spiele typischerweise normalen virtuellen Handel erlauben und fördern, ist es dem Vergnügen der meisten Spieler abträglich, wenn sich Spieler mit viel echtem Geld eine präferierte Position im Spiel erkaufen können. Deshalb verbieten die Nutzungsbedingungen dieser Spiele *Gold Farming*. Man kann hier beobachten, dass *virtuelle Gesetze* eingeführt werden, die kein Gegenstück in der realen Welt haben, deren Missachtung allerdings oft schwer zu eruieren und damit zu regulieren sind (Lit. 09).

Analoge Probleme betreffen die Betreiber von Websites und Suchmaschinen. Erstere sind daran interessiert, bei den wenigen, aber häufig benutzten Suchmaschinen auf vorderen Plätzen der Suchergebnisliste zu gelangen. Hierfür gibt es diverse Methoden der Suchmaschinenoptimierung. Einige dieser Methoden sind vom Suchmaschinenbetreiber durchaus gewünscht, da sie die Liste der Suchergebnisse verbessern, z. B. die Angaben von Metadaten (vgl. Lit. 08). Andere Methoden hingegen werden abgestraft, was bis zur Entfernung aus dem Suchindex führen kann. Eine Methode, die für unethisch gehalten wird und die in der Regel zur Klassifizierung der Zielwebseite als Web Spam führt, ist das *Link Farming*, also das Betreiben von Web Servern alleine zu dem Zweck, den Suchmaschinen eine Zielseite zu empfehlen (vgl. Lit. 10).

Obgleich das Betreiben von Link-Farmen an sich legal ist, üben die Suchmaschinenbetreiber dennoch durch die Einstufung als Spam eine regulative Kontrolle aus. Dies mag einerseits nötig sein, damit die Suchalgorithmen weiter neutral funktionieren, andererseits hat das quasi ethische und direkt ökonomische Konsequenzen, da die Kontrolle völlig im Belieben der jeweiligen Firma liegt und erhebliche Auswirkungen auf die Außenwirkung und damit auf den finanziellen Erfolg einer Organisation haben kann. Analog gilt das unter umgekehrten Vorzeichen, wenn Privatpersonen versuchen, vom Suchmaschinenindex entfernt zu werden, wenn die referenzierten Inhalte ihre *Persönlichkeitsrechte* verletzen.

In allen diesen Fällen ist die Informationsmacht der dominierenden Suchmaschinen so groß, dass auch ohne juristische Regelung bzw. ohne weitere Legitimation Suchmaschinenbetreiber erhebliche Macht über die wirtschaftliche und/oder soziale Position von Menschen und Organisationen ausüben können. Und damit ist dieser ethisch-moralisch-juristische-ökonomische Fragenkomplex auch noch nicht abschließend behandelt, neben die gesellschaftliche Diskussion tritt auch die Frage, wie die weiteren technischen Entwicklungen die Normsetzung und das Verhalten der Akteure in Zukunft beeinflussen dürfen.

C 7.3 Benutzerinteraktion und -verhalten

Wie in der Einleitung dieses Kapitels dargelegt, liegt der Schwerpunkt von *Web Science* im Verstehen des Zyklus aus Konsum und Produktion von Information im Web. Dementsprechend gehört die Planung eines klassischen organisatorischen Informationsprozesses, der Firmendaten, z. B. einen Katalog, nach außen transparent macht, nicht primär in das Gebiet *Web Science*. Jedoch sollte auch die Rückwirkung auf das Ökosystem Web mitbedacht werden, indem z. B. ein solcher Katalog der Verbreitung eines ausgewählten Vokabulars auch außerhalb der eigenen Webseite dienlich sein kann.

Wenn es um die Fragestellung der Mikrointeraktion von Benutzern im Web geht, stellen wir deshalb exemplarisch die Bereitstellung von benutzergenerierten Inhalten und die sonstige wahrnehmbare Benutzeraktivität (z. B. beim Konsumieren von Information) in den Vordergrund.

C 7.3.1 Benutzergenerierte Inhalte und Daten

Neben die klassische Bereitstellung von Inhalten über selbstverwaltete Webseiten, das sogenannte *Web 1.0*, sind seit dem Ende der 90er Jahre zunehmend Formate getreten, die es erleichtern, diese Inhalte ansprechend einzupflegen und zu präsentieren, z. B. in speziellen Content-Management-

systemen. Für spezielle Zwecke wie Blogs, Photos oder Videos gibt es *Web 2.0-Anwendungen* mit passenden Workflows, die die Verwaltung und Präsentation eigener oder fremder Inhalte vorsehen. Außer der Bereitstellung wenig strukturierter Inhalte wie Text und Foto erlauben viele dieser Anwendungen auch die Verwendung *strukturierter, oft semantisch erschlossener Daten*, auf denen zusätzliche Apps aufsetzen können und die dann auch unter den Begriff *Web 3.0* gefasst werden. So lesen Apps bei Facebook strukturierte Daten aus, z. B. Geburtsdaten von Freunden, und diese Daten können gesammelt und in einer Sicht, z. B. einem Geburtstagskalender, integriert werden. Andere Apps integrieren solche Daten aus verschiedenen Webanwendungen. So berechnet Klout eine Kennzahl, die über verschiedene Social-Media-Anwendungen hinweg Daten ausliest und daraus eine Kennzahl berechnet, wie groß der eigene Einfluss auf andere Benutzer dieser Netzwerke war. In einigen Fällen werden solche strukturierte Daten darüber hinaus offen in Semantic-Web-Formaten zur Verfügung gestellt, zum Beispiel um Webseiten in Suchmaschinenrankings besser zu positionieren (Lit. 08).

Zentral bei dem Einstellen solcher Inhalte und Daten ist dabei die Auskunft über deren Herkunft (engl. *Provenance*). *Provenance* (vgl. URL3) erlaubt es, das Verhalten anderer nachzuvollziehen und diese Information über die Herkunft von Information bei der Interpretation zu berücksichtigen. Dabei kann Vertrauen eine Rolle spielen, oft allerdings sind Fakten, zum Beispiel die Angabe einer Arbeitslosenquote, nicht kontextfrei wahr, sondern sind in Abhängigkeit von Bezugsrahmen (Jahr, Personengruppe, Lokation) und Erhebungsmethode (z. B. die der OECD vs. die des Arbeitsamtes) zu verstehen. Aktuelle Anwendungen wie das Wikidata-Projekt bei Wikipedia (URL4) stellen deswegen eine unmittelbare Verknüpfung zwischen Fakten und ihrer Herkunft her und nutzen diese Verknüpfung zur Auswahl und Präsentation.

C 7.3.2 Modellierung von Interaktions- und Verhaltensmustern

Die Modellierung von Interaktions- und Verhaltensmuster ist ein etabliertes Thema in der Webforschung. Modellierung wurde zunächst vor allem eingeführt, damit Webanwendungen den Benutzern Vorschläge machen können, welches Informationsobjekt für sie am interessantesten ist (Lit. 11), wie sie am besten navigieren können (Lit. 12) oder welches Produkt für sie am geeignetsten sein könnte (Lit. 43). Diese frühesten Modelle des Benutzerverhaltens, die in *Empfehlungssystemen* realisiert wurden und werden (vgl. B 14 Empfehlungssysteme), waren aber restringiert auf punktuelle Entscheidungen des Benutzers, z. B. ein Produkt zu kaufen oder nicht. Durch die Modellierung solcher punktuellen Entscheidungen erzielt man kein Verständnis über die Art und Weise, wie Benutzer im Web ihre Zeit einteilen und wie sie verschiedene Webanwendungen nutzen.

Neue Ansätze gehen darüber hinaus, indem sie ein tiefgehendes Verständnis der psychologischen und sozialen Verhaltensmuster anvisieren. Zum Beispiel haben Lehmann et al. (Lit. 13) festgestellt, dass selbst einfache Charakterisierungen des Benutzerverhaltens wie Verweildauer und Häufigkeit der Nutzung einer Webanwendung geeignet sind, verschiedene Arten von Benutzerverhalten und Anwendungen zu kategorisieren. Zum Beispiel werden *Suchmaschinen* häufig benutzt, aber eher mit kürzerer Verweildauer, hingegen *soziale Netzwerke* oft und mit langer Verweildauer und Reiseportale eher selten – nämlich genau zur Erledigung der Aufgabe *Reisebuchung*. Ob Benutzer also mit einem Angebot zufrieden sind und eine *Benutzerbindung* an das Angebot eingetreten ist, lässt sich nicht alleine aufgrund dieser Muster bestimmen, sondern nur im Vergleich zur Nutzung ähnlich gearteter Angebote.

Über Verweildauer und Häufigkeit der Nutzung können aber auch diffizilere Unterschiede in den Verhaltensweisen der Benutzer festgestellt bzw. verwertet werden. Die Sozialwissenschaften haben für kollaborative Systeme verschiedene *Verhaltenstypen* beschrieben, z. B. Lurkers, Trolls oder Elitists. Neuere Untersuchungen befassen sich damit, die Klassifikation in diese Nutzertypen automatisiert aus dem beobachtbaren Verhalten im Web abzuleiten (vgl. Lit. 14, Lit. 15). Zum Beispiel wird hierfür betrachtet, wer Diskussionen anstößt und wer wie oft und in welchen sozialen

Zusammenhängen antwortet. Aus der Betrachtung dieses individuellen Verhaltens und der Zusammensetzung einer *Online Community* mit variierenden Anteilen verschiedener Benutzertypen lässt sich dann voraussagen, ob z. B. eine *Online Community* verstärkt von der Abwanderung ihrer Benutzer betroffen sein könnte (Lit. 16).

Einen andern Weg gehen Golbeck et al. (Lit. 17), die direkt aus veröffentlichten Profilbeschreibungen in sozialen Netzwerken und der Häufigkeit bestimmter Aktionen (z. B. der Betätigung eines Like-Buttons) Korrelationen zu *Persönlichkeitsmerkmalen* (z. B. Intro- vs. Extrovertiertheit) ableiten. Während dieser Ansatz zunächst analytisch ist, lassen sich Webanwendungen denken, die die Bereitstellung von Informationen von den Persönlichkeitsmerkmalen des jeweiligen Benutzers abhängig machen (Lit. 17). Zum Beispiel könnten extrovertierte Benutzer mehr Information von unbekanntenen Personen angeboten bekommen als introvertierte Benutzer.

Darüber hinaus beeinflusst das Verhalten eines Benutzers auf vielerlei Art und Weise, was er von anderen Benutzern angeboten bekommt. Offensichtlich ist dies bei Interaktionen zwischen verschiedenen Benutzern, die *Vertrauen* erfordern. Vertrauen in eine andere Person, etwas (nicht) zu tun (Lit. 18), impliziert ein Risiko. Dieses Risiko beinhaltet einen Verlust (oder auch einen Gewinn) und eine Wahrscheinlichkeit, dass dieser Verlust auftritt. In der persönlichen Interaktion in der *realen* Welt erfolgt eine Einschätzung der Vertrauenswürdigkeit aufgrund des aktuellen und des vergangenen Verhaltens des Gegenübers. Im Web lässt sich aufgrund der Vielzahl der Benutzer und der Kürze der Interaktionen diese Einschätzung oft nicht sinnvoll treffen. E-Commerce-Plattformen behelfen sich an dieser Stelle oft mit summarischen Einschätzungen, zum Beispiel einer Kennziffer für *Reputation* (vgl. Lit. 03), die aus der Bewertung früherer Benutzerinteraktionen (z. B. erfolgreiche oder fehlgeschlagene Käufe) abgeleitet wird.

Aktuell zielen die Entwickler von Webanwendungen allerdings auch auf präzisere Voraussagen für das wahrscheinliche Verhalten von Benutzern und passen dementsprechend ihre Produktangebote an: Unter *Preisdiskriminierung* versteht man die Bereitstellung von Kaufofferten, die davon abhängen, was der Benutzer zuvor getan hat, zum Beispiel ob er direkt zu einer Webseite navigiert ist oder ob er die Webseite von einem Vergleichsportal aus angesteuert hat. In letzterem Fall gibt es eine Tendenz, ihm ein günstigeres Kaufangebot zu machen (Lit. 19).

C 7.4 Makrostrukturen in sozialen Inhalten und Informationsstrukturen

Mikrointeraktionen individueller, isoliert voneinander agierender Benutzer lassen sich auf Meso- und Makroebene in Form von *Verteilungen* über Aktivitäten und daraus resultierenden Strukturen zusammenfassen. Im Web finden sich Makrobeobachtungen zum Beispiel zu:

- Häufigkeitsverteilungen von Wort-Verwendungen, z. B. in Dokumenten oder in benutzererzeugten Inhalten oder in Suchmaschinenanfragen
- Häufigkeitsverteilungen von Beurteilungen, z. B. Notenvergabe für Produkte
- Verteilungen von Linkstrukturen, z. B. Hypertextlinks oder Freundschaftsbeziehungen
- Größenverteilungen von Clustern und Cliques oder Beinahe-Cliques, z. B. Größe von *Threads* in Online-Diskussionen.

Leitfragen zu diesen Verteilungen sind:

- Gibt es Modelle für Mikrointeraktionen, die diese Verteilungen mittels weniger Annahmen erklären?
- Lassen sich Aussagen zu Erwartungswerten (1. Moment der Verteilung), Medianen, Varianz (2. Moment), oder der Schiefe dieser Verteilungen (3. Moment) natürlicherweise aus Mikromodellannahmen ableiten?
- Welche Empfehlungen für Einzelpersonen, für Gruppen oder für gesamtgesellschaftliche Entwicklungen lassen sich aus diesen Makrobeobachtungen gewinnen?

C 7.4.1 Makrobeobachtungen sozialer Benutzerinteraktionen im Web

Das wohl auffälligste Merkmal von Makrobeobachtungen sozialer Benutzerinteraktionen im Web ist, dass selten Gleich-, Normal- oder Exponentialverteilungen auftreten, wie man es erwarten würde, wenn die Modelle für die Mikrointeraktionen der einzelnen Benutzer unabhängig voneinander wären. In der Tat ist es aber so, dass an vielen Stellen Variationen von und Kombinationen von Potenzgesetzen beobachtet werden.

Die einfachste Art dieser Beobachtung erfolgt bereits bei der Frage, welche Worte wie oft in einer Dokumentensammlung auftreten, und bei der Zählung, in wie vielen Dokumenten einer Sammlung ein bestimmtes Wort auftaucht (Lit. 20). Hierdurch erhalten verschiedene Worte eine entsprechende Spezifität. Das Abweichen von solchen Spezifitäten deutet auf eine Besonderheit hin. Zum Beispiel konnten Ginsberg et al. (Lit. 21) zeigen, dass vermehrte Anfragen in Websuchmaschinen nach *Grippe* als Vorboten einer entsprechenden Epidemie gedeutet werden können, da sie mit später berichteten Krankendaten korrelierten. Ebenso lassen sich Microblogs wie Twitter mit Geolokation daraufhin untersuchen, ob Dengue vermehrt erwähnt wird, was auf ein Aufflackern einer entsprechenden Ausbreitung des Dengue-Virus hindeutet (Lit. 22).

Über die Auswertung der *Erwartungswerte* hinaus lassen sich Häufigkeitsverteilungen von Worten nutzen, um Anomalien aufzudecken. Dellschaft & Staab (Lit. 23) berichten, dass Spammer andere Mikrointeraktionen bei der Auswahl ihrer Worte durchführen als *normale* Benutzer. Daher kann man alleine aufgrund der andersgearteten Verteilung und ohne jegliche Analyse der Inhalte Spammer und Nicht-Spammer unterscheiden, sofern nur genügend viele Daten vorliegen.

Seit Ende der 90er Jahre gab es eine umfangreiche Analyse von Netzwerkstrukturen des World Wide Webs. Wiederholt wurde festgestellt, dass Linkstrukturen im Web *skalenfrei* sind, die Anzahl der Links pro Knoten einem *Potenzgesetz* folgen und sich ein Gesamtnetzwerk mit im Mittel kleinem mittlerem Abstand zwischen zwei beliebigen Knoten ergibt. Das Netzwerk bildet eine *Small World* (Lit. 27). Ähnliche Aussagen lassen sich auch für andere Strukturen im Web finden, z. B. für Freundschaftsbeziehungen in sozialen Netzwerken (Lit. 25). Die Entwicklung von fundamentalen Modellen für die Mikrointeraktion, die diese beobachtbaren Makroeffekte erklären, ist noch bei weitem nicht abgeschlossen (vgl. Lit. 26).

C 7.4.2 Systemische Effekte bei Informationskonsum und -produktion

Rückkopplungen im World Wide Web erfolgen oft beiläufig dadurch, dass Benutzer Wertungen abgeben, zum Beispiel dadurch, dass Produkte wie Bücher oder Hotels auf einer Skala bewertet und gegebenenfalls Kommentare verfasst sowie dadurch, dass eigene oder fremde Informationsinhalte annotiert werden. Ein gut untersuchtes Beispiel sind kollaborative Lesezeichensysteme und Referenzlisten wie Delicious (URL5) oder Bibsonomy (URL6), in denen jeder einzelne seine eigenen Daten organisieren kann und beiläufig von der Arbeit anderer profitiert.

Zum Beispiel kann ein Benutzer mit Namen *Staab* seine Webseite (URL7) mit den Tags *Web* und *Science* so annotieren, dass auch ein anderer Benutzer auf die Suchanfrage *Web Science* hin diese Webseite als Teil des Suchergebnisses angezeigt bekommt. Aktivitäten wie diese, aber auch das Klicken auf ausgewählte Suchergebnisse, führen zunächst zu *Systemeffekten erster Ordnung*: Die Daten der Webanwendung ändern sich, und aufgrund dieses geänderten oder erweiterten Informationsbestandes ändern sich Menge, Reihenfolge und Art der Informationspräsentation.

Darüber hinaus ergeben sich aber auch *Systemeffekte zweiter Ordnung*: Aufgrund der geänderten Ansicht erhält der Informationskonsument ein anderes Bild, andere Eingaben von Information, Bewertung oder Annotierungen werden nahegelegt, und damit ändern sich Informationsmodifikation und -produktion. Mithin ändern die produzierten Informationsinhalte des einen Benutzers das, was der nächste Benutzer erzeugt. Solche Effekte können die Qualität des Informationsgehaltes insgesamt potenziell verbessern – aber auch verschlechtern.

C 7.4.3 Informationskaskaden

Eine mögliche Folge von Systemeffekten zweiter Ordnung sind *Informationskaskaden* in Tagging-Systemen (Lit. 27). Solche Kaskaden können dadurch entstehen, dass ein Produzent von Tags beobachten kann, wie andere Produzenten vorher die gleiche (oder auch eine ähnliche) Webseite annotiert haben. Damit wächst die Wahrscheinlichkeit, dass der Informationsproduzent, der ja zunächst Informationskonsument ist, ein bereits benutztes Tag wiederverwendet. Da der nächste Produzent auch wieder Beobachter ist, ergibt sich die Tendenz, dass bereits verwendete Tags auch in Zukunft häufiger verwendet werden.

Entstehende Verteilungen von Taghäufigkeiten lassen sich durch ein Modell, das *Imitation* bei der Produktion von Tags beinhaltet, besonders gut erklären (Lit. 28). Entsprechend der traditionellen Sicht auf solche Zyklen von Informationskonsum und -produktion wird durch die Empfehlung von bereits verwendeten Tags lediglich die Effizienz der Informationsproduktion verbessert, so dass der Benutzer die gleiche Annotierung nur mit geringerem Zeitaufwand betreiben würde. Neuere Resultate haben aber ergeben, dass die Struktur des Informationsraums dadurch beeinflusst wird, ja dass die erzielbare Qualität beim Information Retrieval dadurch vermutlich abnimmt (Lit. 23). Auf analoge Weise ergeben sich die Häufigkeiten bei der Verlinkung von Webseiten. Durch das Suchen nach Inhalten unter Berücksichtigung von *Reputation* wie *Pagerank* (Lit. 29) oder *Authority*, mit anschließender Kreierung von Links profitieren bereits hochgewichtete Seiten überproportional (*rich get richer*; vgl. Lit. 30, Lit. 31, Lit. 32). Auch hier ist der Effekt der unabhängigen Bewertung von Reputation vermischt mit dem Effekt der *Imitation*, was die Information über die Reputation selbst verwässert.

Insbesondere initiale Fehlinformationen können in so einer Informationskaskade überproportionale Aufmerksamkeit erlangen. Deutlich wird dies am Fall von Bewertungsportalen, in denen Kunden ihrer (Un)Zufriedenheit mit einem Produkt (z. B. Urlaubsreise, Smartphone etc.) Ausdruck verleihen können. Untersuchungen haben ergeben (Lit. 23, Lit. 28, Lit. 33), dass initiale Benutzerkommentare häufig überproportional lang und überproportional positiv sind. Dies lässt vermuten, dass die Hersteller oder Verkäufer des Produktes selbst diese „Benutzerkommentare“ erstellt haben. Obwohl nachfolgende Benutzer Metakommentare angeben können, d. h. auf die Frage „Wie nützlich fanden Sie diese Meinung?“ mit einer Note antworten können, führen häufige gute Bewertungen dazu, dass ein initialer Benutzerkommentar häufiger bei einer Suchanfrage in führender Position gelistet wird und eine Informationskaskade mit positiver Wirkung für den Verkäufer und mit negativer Wirkung für den Informationskonsumenten eingesetzt hat.

Diese Informationskaskaden können für alle Benutzer einer Anwendung oder des gesamten Webs gleichermaßen erfolgen. Aufgrund von Empfehlungssystemen wie oben diskutiert, kann es aber auch passieren, dass sich Gruppen herauskristallisieren mit ähnlichen Präferenzen, deren Informationszugang eingeschränkt wird auf ihre – vielleicht nur initial – bevorzugten Inhalte. Wähler von linken Parteien sehen dann im Extremfall nur noch Informationen von den von ihnen präferierten politisch links-orientierten Autoren und das Analoge gilt für die Wähler von rechten Parteien. Ein globaler demokratischer Dialog wird aufgrund solcher Systemeffekte eher eingeschränkt. Solche Makroeffekte wurden bei der Analyse von Querverweisen zwischen Blogs mit politischen Inhalten beobachtet (Lit. 34): Als Systemeffekt gibt sich eine Tendenz zu *Meinungsmonokulturen*. Obwohl das Medium *Web* selbst im Prinzip bestens dafür geeignet ist, den Gedankenaustausch zu fördern, kommt in einem naiv organisierten Informationsraum diese Fähigkeit nicht zum Tragen – vielmehr tritt das Gegenteil ein (Lit. 35).

C 7.5 Soziale Maschinen

Die bisherige Beschreibung von Benutzerinteraktionen und daraus resultierenden Makrostrukturen ist durch *individuelle Ziele*, zum Beispiel Informationsbedürfnisse, und daraus abgeleitete Ak-

tivitäten und Webangebote geprägt. Bereits vor der Entwicklung des Webs gab es allerdings Newsgruppen und Foren im Internet (z. B. im Usenet), die zwar diesen individualistischen Bedürfnissen Rechnung trugen, die aber Aspekte sozialer Reglementierung beinhalteten, um ein *übergeordnetes Gruppenziel* zu erreichen, zum Beispiel einen Mindestgrad an Informationsqualität. Implizit wurde dies teilweise durch *soziale Kontrolle* erzielt, zum Beispiel durch Anfeindungen von Personen, die die *Netiquette* verletzten (Lit. 36), explizit durch *Moderation* von Newsgruppen, die die Beiträge auf ihre jeweilige Eignung filterte, bevor die Beiträge publiziert wurden.

Mit den Programmiermöglichkeiten von Webanwendungen haben sich solcherlei *Möglichkeiten, im Internet in großen Gruppen zu kollaborieren*, weiterentwickelt und ausdifferenziert:

Gruppen können sich ad hoc bilden und *intrinsische Ziele* gemeinsam verfolgen. Ein prominentes Beispiel hierfür ist die Plattform *GuttenPlag*, die gegründet wurde, um kollaborativ einen Plagiatsfall zu untersuchen.

Akteure können versuchen, die *intrinsische Motivation* von Webanwendern für ihre eigenen Ziele zu nutzen. Zum Beispiel beschreiben Ahn et al. (Lit. 37), wie die kognitiven Fähigkeiten von Personen, die sich zwecks Registrierung für einen Webdienst als menschliche Akteure von Computerprogrammen unterscheiden müssen, für die Konvertierung von gedrucktem Text in eine digitale Repräsentation genutzt werden können.

Akteure können versuchen, die *soziale Motivation* von Webanwendern für ihre eigenen Ziele zu nutzen. Zum Beispiel gibt es in *Online Communities* wie im SAP Community Network (SCN) ein Punktesystem, nach dem sich Experten von neuen Benutzern abheben können – ohne dass dies unmittelbar mit einer sonstigen Belohnung verbunden ist (Lit. 38).

Akteure können Webnutzer *extrinsisch motivieren*, z. B. durch Bezahlung, um Aufgaben kollaborativ zu lösen. Ipeirotis (Lit. 39) beschreibt zum Beispiel, für welche Art von kollaborativen Aufgaben der populäre Marktplatz *Amazon Mechanical Turk* zahlende Auftraggeber und bezahlte Webnutzer zusammenbringt.

Die genannten Ziele und Motivationslagen sind hierbei nicht disjunkt. Vielmehr können sie miteinander überlappen und sich verstärken. Zum Beispiel im Fall von Wikipedia kann man argumentieren, dass sowohl das intrinsische Ziel eine Rolle spielt, das eigene Wissensgebiet in der Enzyklopädie qualitativ hochwertig repräsentiert zu sehen, als auch dass erfolgreiche und häufige Beiträge zu erhöhter sozialer Reputation führen können. Letzteres kann sozial motiviert sein, also zu einer Aufwertung als Editor führen mit speziellen Befugnissen, Konflikte auf Wikipedia zu entscheiden. Was aber auch passieren kann, ist, dass durch ungeschickt gesetzte Anreize die Qualität des Ergebnisses nicht steigt, sondern sinkt. Dies wurde zum Beispiel bei Aufgaben auf der Amazon-Mechanical-Turk-Plattform beobachtet, wo eine verbesserte Bezahlung die Qualität der Ergebnisse oft nicht verbessert, sondern verschlechtert (Lit. 40).

Gemeinsam ist diesen kollaborativen Aufgabenlösungen, dass sie einerseits die Arbeitskraft von Menschen nutzen, um *kognitiv* schwierige oder *kreative* Teilaufgaben (z. B. Schreiben oder Verstehen eines Textes, Erkennen von gedruckten Worten, Übersetzung von Texten) zu lösen und andererseits Rechner und Software einbeziehen, um Ergebnisse von Teilaufgaben zu verwalten und Routineaufgaben zu übernehmen. Die Kollaboration zwischen den Benutzern verlangt dabei eine explizite oder implizite Reglementierung, die auf verschiedene Arten geleistet werden kann:

Redundanz der Lösung von Teilaufgaben erlaubt es, die Ergebnisse miteinander zu vergleichen und nur solche Ergebnisse zu akzeptieren, die von verschiedenen Leuten unabhängig voneinander erzielt wurden (Lit. 37).

Hierarchisierung von Befugnissen erlaubt es, Streitfälle an eine übergeordnete Instanz zu eskalieren (z. B. in Wikipedia; Lit. 41, aber auch in anderen Umgebungen, Lit. 44).

Stark intrinsisch motivierte Ziele wie zum Beispiel Katastrophenhilfe unterliegt einer stärkeren sozialen Kontrolle (Lit. 42).

Solche Kombinationen aus (i) verwaltender Software, (ii) sozialer und administrativer Strategien für die Kollaboration und (iii) Einbeziehung menschlicher Problemlösungsfähigkeiten wird auch unter den Begriff *soziale Maschine* subsumiert: „Real life is and must be full of all kinds of so-

cial constraint – the very processes from which society arises. Computers can help if we use them to create abstract *social machines* on the Web: processes in which the people do the creative work and the machine does the administration...” (Lit. 24)

Der Entwurf und Betrieb solcher sozialen Maschinen führt zu völlig neuen technisch-organisatorischen Möglichkeiten, aber auch zu Problemsituationen. Einerseits können einige vormals unlösbare Aufgaben erfolgreich bearbeitet werden, wie zum Beispiel der Fall der massenhaften Digitalisierung von gedrucktem, schwer automatisiert lesbaren Text zeigt (Lit. 37). Andererseits gibt es Fälle von sozialen Maschinen wie die der chinesischen *Human Flesh Search Engine*, die in ethisch-moralisch-juristisch fragwürdigen Situationen angewandt werden. Positive Fälle umfassen einerseits die Aufdeckung von Verbrechen wie Korruption oder Tierquälerei. Gerade in Fällen von Tierquälerei führte die Anwendung der *Human Flesh Search Engine* aber nicht nur zu einer polizeilichen Verfolgung und juristischen Ahndung, sondern auch zu einer nicht akzeptablen Selbstjustiz durch einzelne Nutzer (Lit. 32).

Gerade im Bereich der sozialen Maschinen benötigen wir also einerseits ein besseres Verständnis von Mikrointeraktionen und Makroeffekten, aber auch ein Verständnis davon, wie ethisch-moralisch bedenkliches Verhalten verhindert und gesellschaftlich positive Effekte hervorgerufen werden können.

C 7.6 Konklusion

Web Science ist auf der Suche nach expliziten und impliziten Regulierungs-, Kollaborations- und Koordinationsmethoden, die das Web auch in Zukunft zu einer produktiven und gesellschaftlich wünschenswerten Plattform für das Zusammenwirken vieler, wenn nicht aller Individuen machen. Die Forschungsarbeit der *Web Science* benötigt hierfür einen Mix an Methodeninventar: Normative Wissenschaft untersucht Normen im Web auf ihre Auswirkungen hin, deskriptive Wissenschaft klassifiziert und strukturiert die beobachteten Web-Phänomene aus der Vergangenheit und prädiktive Methoden extrapolieren diese Beobachtungen in die Zukunft. Das Wissen dieser Zusammenhänge lässt uns soziale Maschinen gestalten.

In dieser Übersicht wurden eine Reihe von Phänomenen im Web und Methoden analytisch umrissen, die Gegenstand weiterer *Web-Science*-Forschung sein werden. Dabei ist das Web allerdings so vielfältig, dass die Fragen, wie es das Verhalten der Benutzer beeinflusst und wie daraus neue Information und neues Informationsverständnis geriert wird, damit nur angerissen sind. Insbesondere wurde hier die historische Entwicklung vernachlässigt, in der sich Muster und Modelle finden lassen, wie aus der Beteiligung vieler Anwender auf der Mikroebene (z. B. viele Benutzer stellen Webseiten zur Verfügung) Probleme auf der Makroebene resultieren, die aber auch oft zu neuen Einsichten und Lösungen führen (z. B. *Pagerank*). Diese Entwicklungen zu beschreiben, zu modellieren und auf neue Aufgaben anzuwenden ist Aufgabe der neuen Disziplin *Web Science*.

Literatur

- 01 Eugene Agichtein; Eric Brill; Susan Dumais: Improving web search ranking by incorporating user behavior information. SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval, 19-26, New York, NY, USA, 2006. ACM Press
- 02 Nathan Newmann: Euro Privacy Agencies Slam Google's Privacy Violations, Call for Empowering User Control of Data. Huffington Post, October 18, 2012 (http://www.huffingtonpost.com/nathan-newman/euro-privacy-agencies-google_b_1974715.html)
- 03 Tim Berners-Lee; Wendy Hall; James A. Hendler; Kieron O'Hara; Nigel Shadbolt; Daniel J. Weitzner: A Framework for Web Science. Foundations and Trends in Web Science. now publishers, September 2006
- 04 Akash Kapur: Internet Governance – A Primer. Wikibooks (http://en.wikibooks.org/wiki/Internet_

- Governance; Version vom 28. November 2011, abgerufen am 8. Oktober 2012)
- 05 Hania Farhan; Debra D'Agostino; Henry Worthington: Web Index 2012. World Wide Web Foundation (<http://thewebindex.org/2012/10/2012-Web-Index-Key-Findings.pdf>)
 - 06 Lawrence Lessig: Code: And Other Laws of Cyberspace, Version 2.0. Basic Books 2006
 - 07 Jonathan Zittrain: The Future of the Internet – And How to Stop It. 2008
 - 08 Jason Ronallo: HTML5 Microdata and Schema.org. *code4lib*, issue 16, 2012-02-03, ISSN 1940-5758 (<http://journal.code4lib.org/articles/6400>)
 - 09 Brian Keegan; Muhammad Aurangzeb Ahmad; Dmitri Williams; Jaedeep Srivastava; Noshir Contractor: Sic transit gloria mundi virtuali? Promise and peril at the intersection of computational social science and online clandestine organizations. ACM Web Science 2011: Conference Proceedings, June 14-17, 2012. Koblenz, Germany
 - 10 Nikita Spirin; Jiawei Han: Survey on web spam detection: principles and algorithms. SIGKDD Exploration Newsletter. 13, 2 (May 2012), 50-64 (<http://doi.acm.org/10.1145/2207243.2207252>)
 - 11 Paul Resnick; Neophytos Iacovou; Mitesh Suchak; Peter Bergstrom; John Riedl: GroupLens: an open architecture for collaborative filtering of netnews. Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW '94). ACM, New York, NY, USA, 175-186
 - 12 Robin D. Burke; Kristian J. Hammond; Benjamin C. Young: The FindMe Approach to Assisted Browsing. IEEE Expert 12 (4), 32-40 (1997)
 - 13 Janette Lehmann; Mounia Lalmas; Elad Yom-Tov; Georges Dupret: Models of user engagement. User Modeling, Adaptation, And Personalization. LNCS 7379, Springer 2012, 164-175
 - 14 Sofia Angeletou; Matthew Rowe; Harith Alani: Modeling and analysis of user behavior in online communities. The Semantic Web – ISWC 2011. LNCS 7031. Springer 2011, 35-50
 - 15 Matthew Rowe; Miriam Fernandez; Sofia Angeletou; Harith Alani: Community analysis through semantic rules and role composition derivation. Journal of Web Semantics, 2012
 - 16 Marcel Karnstedt; Matthew Rowe; Jeffrey Chan; Harith Alani; Conor Hayes: The effect of user features on churn in social networks. Third ACM/ICA Web Science Conference 2011, Koblenz, Germany, June 14-17, 2011
 - 17 Jennifer Golbeck; Cristina Robles; Karen Turner: Predicting personality with social media. CHI Extended Abstracts 2011, 253-262, ACM 2011
 - 18 Piotr Cofta: The Trustworthy and Trusted Web. Foundations and Trends in Web Science 2 (4), 243-381. now publishers, 2011
 - 19 Jakub Mikians; Laszlo Gyarmati; Vijay Erramilli; Nikolaos Laoutaris: Detecting price and search discrimination on the internet. Hotnets '12 – Eleventh ACM Workshop on Hot Topics in Networks, Oct 29-30, 2012, Seattle, WA, USA, ACM 2012
 - 20 Karen Spärck Jones: A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation, Volume 28 Number 1, 1972, 11-21
 - 21 J. Ginsberg; M. H. Mohebbi; R. S. Patel; L. Brammer; M. S. Smolinski; L. Brilliant: Detecting influenza epidemics using search engine query data. Nature 457, 1012-1014, 2009
 - 22 Janaína Gomide; Adriano Veloso; Wagner Meira Jr.; Virgílio Almeida; Fabrício Benevenuto; Fernanda Ferraz; Mauro Teixeira: Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. Third ACM/ICA Web Science Conference 2011, Koblenz, Germany, June 14-17, 2011
 - 23 Klaas Dellschaft; Steffen Staab: On differences in the tagging behavior of spammers and regular users. Proc. of WebSci-2010, Raleigh, US, April 2010
 - 24 Tim Berners-Lee: Weaving the Web, Harper-Collins, 1999
 - 25 Lars Backstrom; Paolo Boldi; Marco Rosa; Johan Ugander; Sebastiano Vigna: Four Degrees of Separation. Fourth ACM/ICA Web Science Conference 2012, Evanston, USA, June, 2012
 - 26 Hans Akkermans: Web dynamics as a random walk: How and why power laws occur. ACM Web Science 2012: Conference Proceedings, June 22-24, 2012. Evanston, IL, 1-10
 - 27 David Easley; Jon Kleinberg: Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Cambridge University Press, 2010
 - 28 Klaas Dellschaft; Steffen Staab: An epistemic dynamic model for tagging systems. Proc. of the 19th ACM Conference on Hypertext and Hypermedia, HT '2008, Pittsburgh, PA, USA, June 19-21, 2008. ACM 2008,

- 71-80
- 29 Sergey Brin; Lawrence Page: The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems, Proceedings of the Seventh International World Wide Web Conference, Volume 30, Issues 1-7, April 1998*, 107-117
 - 30 A.-L. Barabási; R. Albert: Emergence of Scaling in Random Networks. *Science* 286 (1999) 509-512
 - 31 D. J. De Solla Price: A General Theory of Bibliometric and Other Cumulative Advantage Processes. *J. Amer. Soc. Inform. Sci.* 27 (1976) 292-306
 - 32 Fei-Yue Wang; Daniel Zeng; James A. Hendler; Qingpeng Zhang; Zhuo Feng; Yanqing Gao; Hui Wang; Guanpi Lai: A Study of the Human Flesh Search Engine: Crowd-Powered Expansion of Online Knowledge. *Computer*, 45-53, August, 2010, IEEE Press
 - 33 Klaas Dellschaft; Steffen Staab: Measuring the influence of tag recommenders on the indexing quality in tagging systems. *Proc. of the 23rd ACM Conference on Hypertext and Social Media, HT '12, Milwaukee, WI, USA, June 25-28, 2012*. ACM 2012, 73-82
 - 34 Lada A. Adamic; Natalie Glance: The political blogosphere and the 2004 U.S. election: divided they blog. *Proceedings of the 3rd international workshop on Link discovery (LinkKDD '05)*. ACM, New York, NY, USA, 2005, 36-43
 - 35 Eli Pariser: *The Filter Bubble: What the Internet Is Hiding from You*, Penguin Press, New York, May 2011
 - 36 Christine B. Smith; Margaret L. McLaughlin; Kerry K. Osborne: Conduct Control on Usenet. *Journal of Computer-Mediated Communication* 2 (4), Wiley 1997
 - 37 Luis von Ahn; Benjamin Maurer; Colin McMillen; David Abraham; Manuel Blum: recaptcha: Human-based character recognition via web security measures. *Science* 321.5895 (2008), 1465-1468
 - 38 Georgia Demetriou; Peter Kawalek: Benefit-driven participation in open organizational social media platforms: the case of the SAP Community Network. *Issues in Information Systems* 10 (1), 601-611, 2010
 - 39 Panagiotis G. Ipeirotis: Analyzing the amazon mechanical turk marketplace. *XRDS: Crossroads, The ACM Magazine for Students*. 17.2 (2010), 16-21
 - 40 Winter Mason; Duncan J. Watts: Financial incentives and the performance of crowds. *Proceedings of the ACM SIGKDD workshop on human computation*. ACM, 2009
 - 41 Brian Butler; Elisabeth Joyce; Jacqueline Pike: Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 1101-1110, 2008 (<http://doi.acm.org/10.1145/1357054.1357227>)
 - 42 Ory Okolloh: Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. *Participatory Learning and Action* 59.1 (2009), 65-70
 - 43 Greg Linden; Brent Smith; Jeremy York; Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing* 7 (1), 76-80, 2003
 - 44 Felix Schwagereit; Ansgar Scherp; Steffen Staab: Survey on Governance of User-generated Content in Web Communities. *Third ACM/ICA Web Science Conference 2011, Koblenz, Germany, June 14-17, 2011*

URL-Verzeichnis

- URL1 http://commons.wikimedia.org/wiki/File:Human_Behavior_Co-constituting_the_Web.png
 URL2 <http://schema.org>
 URL3 http://www.w3.org/2011/prov/wiki/Main_Page
 URL4 <http://meta.wikimedia.org/wiki/Wikidata>
 URL5 <http://delicious.com/>
 URL6 <http://www.bibsonomy.org/>
 URL7 <http://west.uni-koblenz.de/>
 URL8 http://en.wikipedia.org/wiki/IP_hijacking

(abgerufen am 8. Oktober 2012)