

Soziale Annotation

Seminararbeit

vorgelegt am

Lehrstuhl für Praktische Informatik IV
Prof. Dr.-Ing. Wolfgang Effelsberg
Universität Mannheim

im
Juli 2006

im Rahmen des Seminars

Seminar Bridging the Semantic Gap
im SS 2006

Betreuer
Prof. Dr. Heiner Stuckenschmidt

von
Hoang Thi Thanh Thuy

Inhaltsverzeichnis:

Inhaltsverzeichnis

Abbildungsverzeichnis

1. Einführung
 - 1.a Metadaten
 - 1.b Motivation
 - 1.c Geschichte & Begriffe
2. Folksonomy
 - 2.a Del.icio.us
 - 2.b Flickr.com
 - 2.c Probleme bei Tag Wahl
 - 2.d Vor- und Nachteile von Folksonomy
3. Bemerkungen zur Folksonomy
4. Google vs. Del.icio.us
5. Weiter Forschungs- und Anwendungsgebiete
6. Fazit
7. Quellenangaben

Abbildungsverzeichnis

Abb. 1: Übersicht der Metadaten [selbst erstellt]

Abb. 2: Startseite von del.icio.us [2]

Abb. 3: Benutzerfreundliche Buttons bei del.icio.us [2]

Abb. 4: Beliebtesten Tags von del.icio.us [2]

Abb. 5: Startseite von flickr.com [3]

Abb. 6: Feature Whatsinyourbag [3]

Abb. 7: Beliebtesten Tags von flickr.com [3]

1. Einführung

Bei der sozialen Annotation geht es darum, dass der Benutzer nicht vom Vorgang der Annotation abgeschlossen ist, sondern dass er aktiv bei der Gestaltung der Metadaten mitmacht. Somit ist eine Folksonomy vom Benutzer für den Benutzer. Natürlich bringt eine soziale Annotation auch Nachteile mit sich, wie zum Beispiel dass die Daten keine feste Struktur haben. Nichts desto trotz findet die soziale Annotation großen Anklang bei den Benutzern des World Wide Web. Die meisten Erkenntnisse hier würden von [7] und [8] gewonnen.

1.a. Einführung: Metadaten

Metadaten sind Daten über Daten, diese können administrativ, strukturiert und beschreibend sein. Wir konzentrieren uns hier auf beschreibende Metadaten. Diese Metadaten können wiederum durch professionelle, automatische oder soziale Annotationen beschrieben werden.

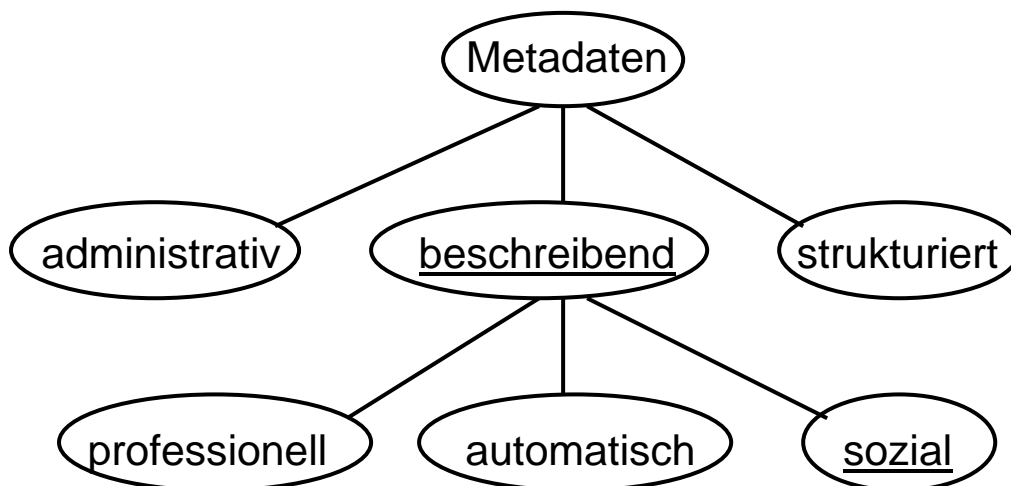


Abb. 1: Übersicht der Metadaten [selbst erstellt]

Eine professionelle Annotation ist sehr zeitaufwendig und mühselig herzustellen. Dementsprechend ist es im World Wide Web sehr ungünstig, obwohl die Qualität einer professionellen Annotation sehr gut ist, auch wenn sie für einen Laien zu komplex erscheint. Bei einer

professionellen Annotation spricht man auch von einer traditionellen Annotation.

Eine automatische Annotation ist günstiger wie eine professionelle Annotation, aber bei der automatischen Annotation muss das System erst einmal trainiert werden, trotz gutem Training ist das Ergebnis nicht 100% fehlerfrei.

Beide Annotationen haben einen gravierenden Nachteil, der Benutzer ist vom Vorgang der Annotation getrennt und kann nicht darauf Einfluss nehmen.

Dies ist bei einer sozialen Annotation nicht der Fall, denn hier macht der einzelne Benutzer oder eine Community die Annotation für den Benutzer.

1.b. Einführung: Motivation

Der Benutzer muss erst einmal motiviert werden selber zu annotieren. Ein Grund hierfür ist, dass er sein eigenes Wissen und die Erkenntnisse, die er gewonnen hat mit anderen teilen möchte. Ein weiterer Motivationsgrund ist, dass der Benutzer leichter Dokumente im Web wieder finden kann und damit auch für andere das Finden der Dokumente leichter macht.

Des Weiteren ist es so wie Shelly Powers gesagt hat „This is the user’s web now, which means it’s my web and I can make the rules.“
Somit sollte sich jeder einzelne Benutzer dazu aufgefordert fühlen das Web aktiv mit zu gestalten und selber zu annotieren.

1.c. Einführung: Geschichte & Begriffe

Mosaic ist der Großvater von allen modernen Web-Browsern und hat auch die Hotlist erfunden. Die Hotlist ist aber nur eine Auflistung von Links. Genauso wie die Favoriten beim Internet Explorer. Danach kamen die Suchmaschinen wie Yahoo! oder Google, welches ein dynamisches Bookmarking Service bietet bzw. Bookmarks nach Nachfrage. Was jetzt kommt sind so genannte Sozial Link Manager, die vom Benutzer für den Benutzer gestaltet werden. Den Sozial Link Managern nennt man auch Folksonomy. Der Begriff Folksonomy

wird auf Thomas Vander Wal zurückgeführt [5], Folksonomy besteht aus dem Wort Folk und Taxonomie, welches schlicht bedeutet dass es vom Benutzer selber klassifiziert ist. Merkmale der Folksonomy sind das sie keine Hierarchie wie bei einer Ontologie haben und das sie eine freie Struktur haben, die nicht festgelegt worden ist. Ein weiterer Begriff der Sozialen Annotation ist der Begriff Tag, welches eine 1-Wortbeschreibung bzw. Schlüsselwort ist. Ein Merkmal von Tags ist das sie nicht kontrollierbar sind, da jeder einzelne Benutzer jeden Tag auswählen kann den er möchte.

2. Folksonomy

Das Wort Folksonomy hat sich in den meisten Dokumenten durchgesetzt. Begriffe wie „folk classification“, „ethnoclassification“, „distributed classification“ oder „social classification“ werden auch im Zusammenhang mit Merkmalen der sozialen Annotation, wie der freien Struktur genannt. Die Begriffe meinen aber alle das gleiche. Beispiele für eine Folksonomy sind del.icio.us, flickr.com, wikipedia.org, 43Things.com, ...

2. a. Folksonomy: del.icio.us

The screenshot shows the del.icio.us homepage. At the top left is the del.icio.us logo. On the right, there are links for 'popular | recent' and 'login | register |'. Below the logo, there are three main sections: 'keep' (your favorite websites, music, books, and more in a place where you can always find them), 'share' (your favorites with family, friends, and colleagues), and 'discover' (new and interesting things by browsing popular & related items). In the center, there is a search bar with the text 'discover favorites:' and a 'search' button. To the right of the search bar is a 'sign up now' section with fields for 'username', 'password', 'password again', and 'email', and a 'register' button. Below the search bar, there is a section titled 'What's a tag?' with the text 'A tag is just a word that describes an item saved on del.icio.us. Learn more >'. At the bottom, there is a section titled 'the del.icio.us hotlist' with a sub-header 'also see: popular | recent | tags'. This section contains a list of items, each with a title, a 'save this' link, a list of tags, and a 'people' count. The items are grouped by time: 'hot right now', '1 hour ago', '2 hours ago', and '3 hours ago'. The first item in the 'hot right now' group is '20 Amazing Facts About Voting in the USA' with 50 people, tags 'politics voting bush usa history', and 'first posted by Zuljin ... on june 27'. The first item in the '1 hour ago' group is 'Welcome to ConnectBeam' with 48 people, tags 'social tagging bookmarks web2.0 collaboration', and 'first posted by traui ... on june 26'. The first item in the '2 hours ago' group is 'Favicon.ico Maker' with 157 people, tags 'favicon tools web design icons', and 'first posted by pwever ... on sept 30, 2004'. The first item in the '3 hours ago' group is 'Google"Pro' with 49 people, tags 'google business interview news search', and 'first posted by niyagawa ... on june 30'.

Abb. 2: Startseite von del.icio.us [2]

del.icio.us ist ein sozialer Bookmarkmanager. Es ist im Dezember 2003 erstmals online gegangen und ist das verbreitetste Folksonomysystem. Hier gibt man einen Suchbegriff bzw. Tag ein und bekommt eine Liste von Links, die Links sind aber versehen mit Tags und einem persönlichen Kommentar, welche in zeitlichen Einheiten unterteilt sind. Man hat hier aber nicht nur die Möglichkeit

nach Tags zuzusuchen, sondern man kann hier auch nach einzelnen Benutzern suchen und kann sehen, was dieser als seine favorisierten Links gekennzeichnet hat. Vom Benutzer erkennt man auch, was für Tags er gewählt hat und somit seine Favoriten auf das eigene Interessensgebiet etwas verkleinern, um dann seinen Pfad zu folgen.

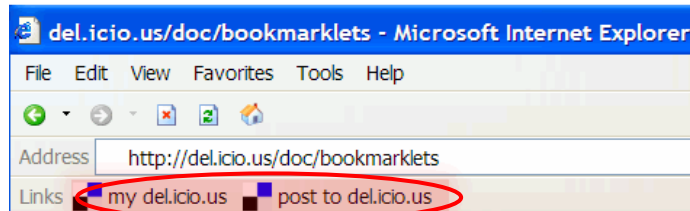


Abb. 3: Benutzerfreundliche Buttons bei del.icio.us [2]

Bei del.icio.us kann man dann 2 Buttons installieren (siehe Abb. 3). Somit kann man beim Durchstöbern des Webs die Seiten, die man interessant findet, einfach zu seinen Favoriten hinzufügen, indem man auf den Button „my del.icio.us“ klickt. Dann kann man zu seinem ausgewählten Link einen Tag und wenn man will, noch einen persönlichen Kommentar hinzufügen. Damit ist der Link für alle Benutzer zugänglich, auch wenn diese den Link nicht getagt haben. Bei der Tag-Wahl kann man sich, falls der Link schon existiert, bei den Common Tags orientieren. Common Tags sind die Tags, die von den meisten Benutzern für diesen Link gewählt worden sind. Eine weitere Möglichkeit, sich bei der Tag-Wahl zu orientieren, sind die beliebtesten Tags, popular tags (siehe Abb. 4).

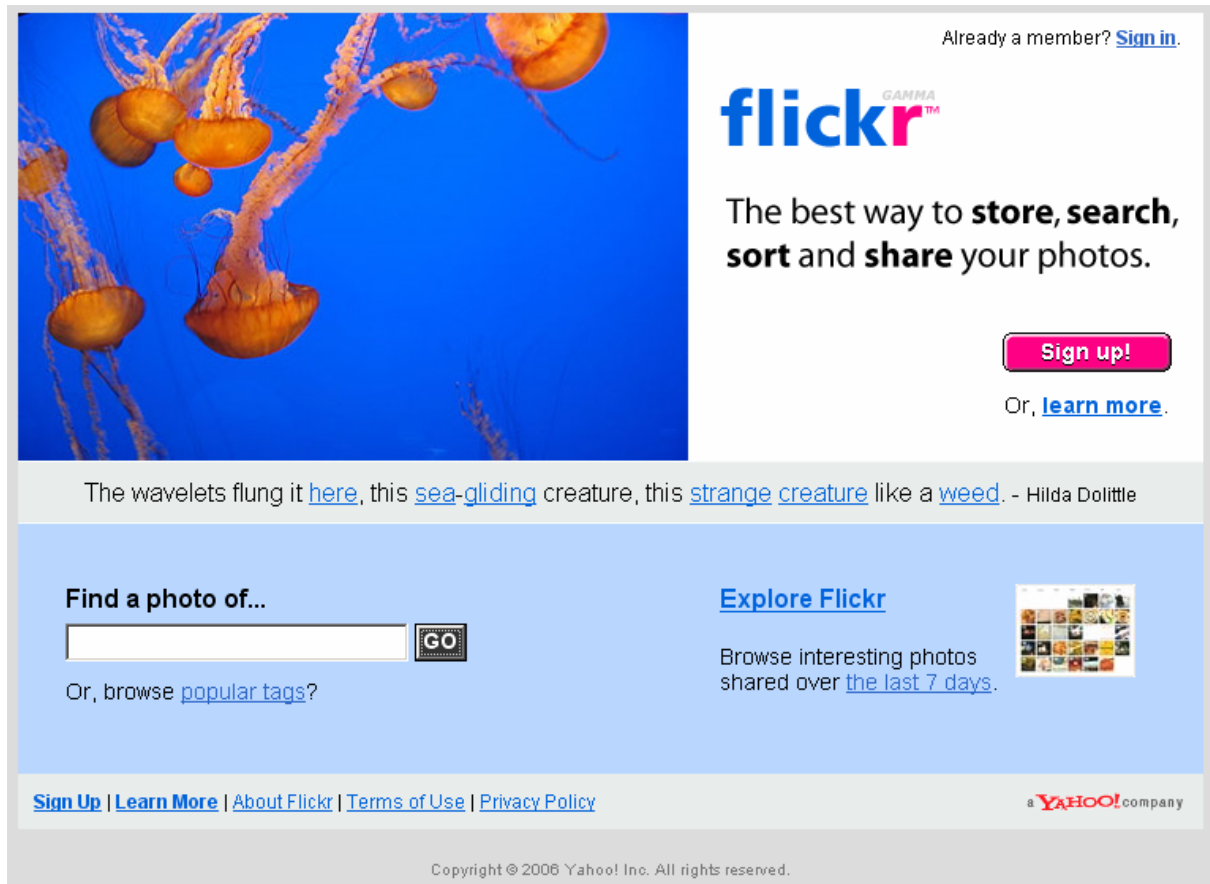


Abb. 4: Beliebtesten Tags von del.icio.us [2]

An den beliebtesten Tags sehen wir schon, wie an den Wörtern „java“, „javascript“, „linux“ und „mac“ (rot eingekreist), dass del.icio.us

meist von technisch interessierten Menschen benutzt wird. Was auch interessant ist, ist das Wort „toread“ (grün eingekreist), denn dadurch wird klar dass die Links vom Benutzer getagt worden sind. Wenn sie eine Seite lesen müssen geben sie ihm einfach den Tag „toread“.

2.b. Folksonomy: flickr.com



Already a member? [Sign in.](#)

flickr GAMMA™

The best way to **store, search, sort** and **share** your photos.

[Sign up!](#)

Or, [learn more.](#)

The wavelets flung it [here](#), this [sea-gliding](#) creature, this [strange creature](#) like a [weed](#). - Hilda Dolittle

Find a photo of... [GO](#)

Or, browse [popular tags?](#)

[Explore Flickr](#)

Browse interesting photos shared over [the last 7 days.](#)

[Sign Up](#) | [Learn More](#) | [About Flickr](#) | [Terms of Use](#) | [Privacy Policy](#)

a **YAHOO!** company

Copyright © 2006 Yahoo! Inc. All rights reserved.

Abb. 5: Startseite von flickr.com [3]

flickr.com ist ein Fotomanager, man kann hier private Fotos, Fotos von Webseiten oder auch Softwareanwendungen hochladen. Meist wird flickr.com aber benutzt um eigene Bilder zu repräsentieren. flickr.com sollte ursprünglich ein Online-Spiel werden. Als man im Februar 2004 es erstmal als Vorabversion online stellte, hat man festgestellt, dass das Feature mit den Bilderhochladen, mehr Anklang fand als das Spiel an sich. Somit hat man entschlossen das Spiel sein zulassen und sich mehr auf den Fotomanager zukonzentrieren. [6] Nebenbei bemerkt flickr.com und del.icio.us gehören zu Yahoo!. [4]

Wie in del.icio.us kann man hier nach einem bestimmten Tag oder aber auch nach einem Benutzer suchen. Jeder Benutzer hat die Möglichkeit einen Kommentar zu jedem Bild zumachen, aber nur Freunde und Familienmitglieder haben die Möglichkeit noch weitere Tags zu einem Bild hinzuzufügen. Bei flickr.com haben wir noch ein besonderes Feature, welches „Whats in your bag“ heißt. Hierbei hat der Benutzer die Möglichkeit eine bestimmte Stelle im Bild zumarkieren und mit einem Kommentar zu versehen.



Abb. 6: Feature Whatsinyourbag [3]

Auch hier haben wir bei der Tag Wahl mit den beliebtesten Tags eine Orientierungsmöglichkeit. (siehe Abb. 7)

africa amsterdam animal animals april architecture art australia baby barcelona
 beach berlin birthday black blackandwhite blue boston building bw california
 cameraphone camping canada canon car cat cats chicago china
 christmas church city clouds color concert day dc dog dogs england europe
 family festival film florida flower flowers food france friends fun
 garden geotagged germany girl graduation graffiti green halloween hawaii hiking
 holiday home honeymoon hongkong house india ireland island italy japan june kids
 lake landscape light london losangeles macro march may me mexico moblog
 mountain mountains museum music nature new newyork newyorkcity newzealand
 night nikon nyc ocean paris park party people portrait red river roadtrip
 rock rome san sanfrancisco school scotland sea seattle show sky snow spain
 spring street summer sun sunset sydney taiwan texas thailand tokyo toronto
 travel tree trees trip uk urban usa vacation vancouver washington water
 wedding white winter yellow york zoo

Abb. 7: Beliebtesten Tags von flickr.com [3]

2.c.Folksonomy: Probleme bei Tag Wahl

Wie wir oben schon sehen gibt es gewisse Probleme bei der Tag Wahl, wie zum Beispiel „flower“ und „flowers“ (rot eingekreist). Zwei Wörter meinen das gleiche, aber werden als verschiedene Tags gesehen, bzw. flickr erkennt nicht ob der Tag Mehrzahl oder Einzahl ist. Es gibt aber auch die Problematik, wenn ein Tag verschiedene Sachen definieren soll, bzw. man kann den Tag nicht eindeutig zuordnen. Ein Beispiel hierfür wäre „apple“, „apple“ kann der Macintosh sein oder ein Apfel, man könnte aber auch den Big Apple damit verbinden. Synonyme werden nicht erkannt.

Obwohl flickr.com auch in Japan und Frankreich ist, beherrscht die englische Sprache die Tags. Problematisch könnte es dann mit der chinesischen oder russischen Sprache geben, da sie ja andere Zeichen haben. Aber auch mit der deutschen Sprache könnte es Probleme geben, wegen dem ä, ü und ö, diese werden ja dann als ae, ue und oe geschrieben.

Auch ein Problem ist das es Benutzer gibt die einfach mehrer Wörter aneinander reihen, wie zum Beispiel „Womanwithaflower“. Man

könnte aber auch „Women.with.a.flower“ oder „Women-with-a-flower“ schreiben, hierbei taucht das Problem mit den Satzzeichen auf.

Es können aber auch Rechtschreibfehler oder Spaßbezeichnungen auftauchen, diese sind aber kein großes Problem, da eine Folksonomy ein selbst regulierendes gemeinschaftliches Netzwerk ist. Das heißt wenn ein Einzelner einen Rechtschreibfehler oder eine Spaßbezeichnung macht fällt das nicht ins Gewicht, da die meisten Benutzer so etwas nicht machen. Es würde nur ins Gewicht fallen, wenn mehr als die Hälfte der Benutzer den gleichen Fehler oder die gleiche Spaßbezeichnungen machen würde und die Wahrscheinlichkeit das so etwas eintritt ist sehr gering.

Es gibt einige Probleme der Tag Wahl und da diese ja nicht kontrollierbar und voraussehbar ist, im Gegensatz zu einem formalen Klassifizierungssystem.

Ein möglicher Ansatz, um mehr Struktur in die Tags zu bringen sind die beliebtesten und die Common Tags, woran man sich orientieren kann, welchen Tag man wählt. Dies ist aber keine Lösung, da man sich nur orientieren kann und somit bleibt die Tag Wahl immer noch unkontrollierbar. Ein weiterer Ansatz ist, wie in Wikipedia das man Kategorien und Unterkategorien hat. Der Ansatz ist aber leider nicht überall durchführbar. Aber zurzeit gibt es keine Lösung für Unkontrollierbarkeit der Tags.

2.d. Folksonomy: Vor- und Nachteile

Der große Nachteil bei einer Folksonomy ist, dass die Tag Wahl nicht immer ideal ist und dass man keine Kontrolle bzw. keine Struktur bei den Tags hat. Was auch ein Nachteil ist, dass man das Surferverhalten jedes einzelnen Benutzers verfolgen kann, da man nach Benutzern suchen kann. Somit kann man seine Interessen herausfinden und ihn dann gezielt mit Werbung bombardieren.

Aber die Suche nach dem Benutzer hat auch seine Vorteile, wie bei del.icio.us, kann man seinem Pfad folgt und somit neue/interessante Themen und Dokumente finden kann. Natürlich hat die Folksonomy

noch weitere Vorteile, wie dass es vom Benutzer für den Benutzer ist und somit auch wenig Fachbegriffe darin enthalten sind. Ein weiterer Vorteil ist, dass man keine Kompromisse eingehen muss im Gegensatz zur einer professionellen Annotation. Dementsprechend ist eine soziale Annotation günstiger als eine professionelle.

Der Grund warum Folksonomy mehr an Beliebtheit gewinnt ist, dass alles nicht nur auf eine einzelne Person oder Gruppe basiert, sondern auf eine Community und die Bereitschaft sein Wissen mit anderen zuteilen, dementsprechend ist die Barriere mitzumachen gering. Ein weiterer Grund ist seine Einfachheit, man braucht hier kein komplexes Klassifizierungssystem um an ein gutes Ergebnis zu kommen. Laut Stewart Butterfield, einer der Erfinder von flickr.com bekommt man bei einer Folksonomy zu 90% ein korrektes Ergebnis, ist aber 10x einfacher wie ein formales Klassifizierungssystem. Viele Fachleute finden Butterfields Behauptung übertrieben, aber geben zu dass bei einer Folksonomy nicht triviale und wichtige Metadaten erkannt werden, finden aber einen Vergleich zwischen einer professionellen und sozialen Annotation irrelevant.

3. Bemerkungen zur Folksonomy

Es sind noch 5 Bemerkungen zur Folksonomy zu machen.

1. Die Sprache zwischen einer professionellen und sozialen Annotation kann unterschiedlich sein. Da der einfache Benutzer eine andere Sicht hat wie einer der professionell annotiert, können unter Umständen auch unterschiedliche Begriffe auftauchen. Vielleicht sogar Begriffe, die ein Laie nicht versteht und wo man eine Übersetzung machen muss.
2. Die Sprache vom Benutzer ändert sich, ein Beispiel hierfür wäre das Wort „googeln“, welches im Jahr 2004 in den Duden aufgenommen wurde[1]. Vor 7 Jahren war dieser Begriff noch unbekannt. Hier kann man auch von einem Vorteil der Folksonomy sprechen, da die Tags sich mit der Sprache vom Benutzer mit verändert.
3. Eine Folksonomy ist ein selbst regulierendes gemeinschaftliches Netzwerk und je mehr Benutzer an der sozialen Annotation teilnehmen, umso besser wird sie auch. Ein schönes Beispiel hierfür ist wikipedia.org.
4. Man sollte einen unterschied ziehen zwischen suchen und stöbern, denn wenn man etwas gezielt sucht geht man anders vor als wie man einfach nur im Netz rumstöbern möchte. Man kann bei einer Folksonomy interessante Themen finden, welche man beim gezielten Suchen nicht finden würde, bzw. nicht darauf kommen würde danach zusuchen. Wogegen bei einer gezielten suche man so schnell wie möglich an sein Ziel kommen möchte.
5. Handelt von der Beeinflussung der Tag Wahl. Man kann sich bei der Tag Wahl mehr oder weniger von den beliebtesten oder Common Tags beeinflussen lassen. Je nach dem wie gut oder schlecht der Tag ist, kann die Beeinflussung gut oder schlecht sein. Ideal wäre es natürlich, wenn jeder genannte Tag in den Common Tags auftauchen würde und jeder Benutzer sich dann entscheiden kann, ob der Tag gut ist oder nicht. Aber dadurch

würden die Common Tags höchstwahrscheinlich unübersichtlich werden und aus allen Nähten platzen.

4. Google vs. Del.icio.us

Google war erstmals im September 1998 online. Der Grund warum sich Google von anderen Suchmaschinen hervorgehoben hat, ist die schnelle Lieferung der Ergebnisse. Dementsprechend ist das direkte Suchen bei Google auch besser als wie bei del.icio.us. Ein weiterer Grund warum das direkte Suchen bei Google besser ist, ist das Google global sucht, wogegen bei del.icio.us nur die lokale Umgebung absucht wird. Da nur ein Link als Ergebnis geliefert werden kann, wenn ein Benutzer es markiert bzw. einen Tag dazu genannt hat.

Aber wenn man einfach nur im Web etwas herum stöbern möchte ist del.icio.us besser, da man den Pfad von einem/mehreren Benutzer folgen kann und somit interessante Themen und Dokumente findet.

Der Algorithmus bei Google heißt PageRang und ist patentiert. Was einzig und allein bei Google zählt ist die Linkpopularität, der Inhalt der Links ist vollkommen egal. Dies sehen wir wenn wir bei Google „miserable failure“ eingeben und als erstes die Biographie von George W. Bush erhalten. Bei del.icio.us würde das nicht auftreten, da hier nur der Inhalt zählt.

Bei del.icio.us muss man sich nicht vorher erst einlesen wie bei Google, sondern kann schon bei den gewählten Tags und Kommentaren von anderen Benutzern erkennen, ob es sich zu lesen lohnt bzw. abwägen ob man es lesen möchte.

Aber da del.icio.us noch nicht so lange existiert und noch zu wenige Benutzer aktiv die Seite mit gestalten bzw. die meisten Benutzer technisch interessiert sind und dementsprechend auch die Links sind, wird del.icio.us noch eine Weile brauchen, um so bekannt zu werden wie Google es jetzt ist. [1]

5. Weiter Forschungs- und Anwendungsgebiete

Eine Verbesserungsmöglichkeit der Quantität der Ergebnisse ist, wenn das Ergebnis benutzerspezifischer wäre. Ein Benutzer könnte im Vorfeld seine Interessensgebiete bestimmen und würde dann nur Ergebnisse geliefert bekommen, welche seine Interessen wieder spiegelt. Hier muss man natürlich die Möglichkeit haben jederzeit die Gebiete zu erweitern oder zu verkleinern.

Auch eine qualitative Verbesserungsmöglichkeit der Ergebnisse ist denkbar, wenn man einem Benutzer, der oft zu Links den passenden Tag nennt, den auch vielen Benutzern übernehmen, mehr gewichtet als einen Benutzer der immer Tags von anderen übernimmt. Dadurch würden gute Tags nicht verschwinden und das Ergebnis wäre mit (noch) besseren Tags versehen.

Es wird aber noch weitere Untersuchungen gemacht wie beim FolkRang-Algorithmus [10] oder bei der Community, inwieweit sie mit einander verstrickt ist[9].

Soziale Annotation ist überall anwendbar, wo schon ein vorhandener Klassifikationsentwurf oder ein Intranet existiert, Beispiele hierfür wäre in der Universität oder in der Melodieerkennung denkbar.

6. Fazit

Man kann sagen, dass eine Folksonomy zugleich eine gute und schlechte Verwaltung der Information ist, da sie nicht kontrollierbar und ungenau ist. Aber da der Benutzer aktiv die Daten mitgestaltet, mit seinen eigenen Wörtern und seinem Ermessen was relevant ist, ist es wiederum eine gute Art seine Informationen zu verwalten. Im Gegensatz zur professionellen Annotation ist die soziale Annotation gleich öffentlich.

Eine soziale Annotation ist aber kein Ersatz für ein formales Klassifizierungssystem, es könnte aber ein ergänzendes Mittel sein um Informationen und Resultate besser zu managen, denn das Suchergebnis ist mit einer Beschreibung, mehreren Tags und einem persönlichen Kommentar versehen.

Somit ist eine soziale Annotation verbunden mit der professionellen bzw. automatischen Annotation die Zukunft des World Wide Webs.

7. Quellenverzeichnis

[1] <http://de.wikipedia.org/wiki/Googeln>

[2] www.del.icio.us

[3] www.flickr.com

[4] <http://de.wikipedia.org/wiki/Del.icio.us>

[5] <http://de.wikipedia.org/wiki/Folksonomy>

[6]

<http://www.heute.de/ZDFheute/inhalt/25/0,3672,2281241,00.html>

[7] <http://www.dlib.org/dlib/april05/hammond/04hammond.html>

[8]

http://blog.namics.com/2005/Folksonomies_Cooperative_Classification.pdf

[9] http://dbs.informatik.uni-halle.de/GvD2006/gvd06_jaeschke.pdf

[10] <http://www.kde.cs.uni->

[kassel.de/hotho/pub/2006/seach2006hotho_eswc.pdf](http://www.kde.cs.uni-kassel.de/hotho/pub/2006/seach2006hotho_eswc.pdf)