

Lernendes Assoziatives Ranking für Wiki Webs

12.02.05

Jan Friderici

jan@friderici.net

Seminararbeit im Seminar

NOMADS – Networks of Mobile Adaptive Dependable Systems
bei

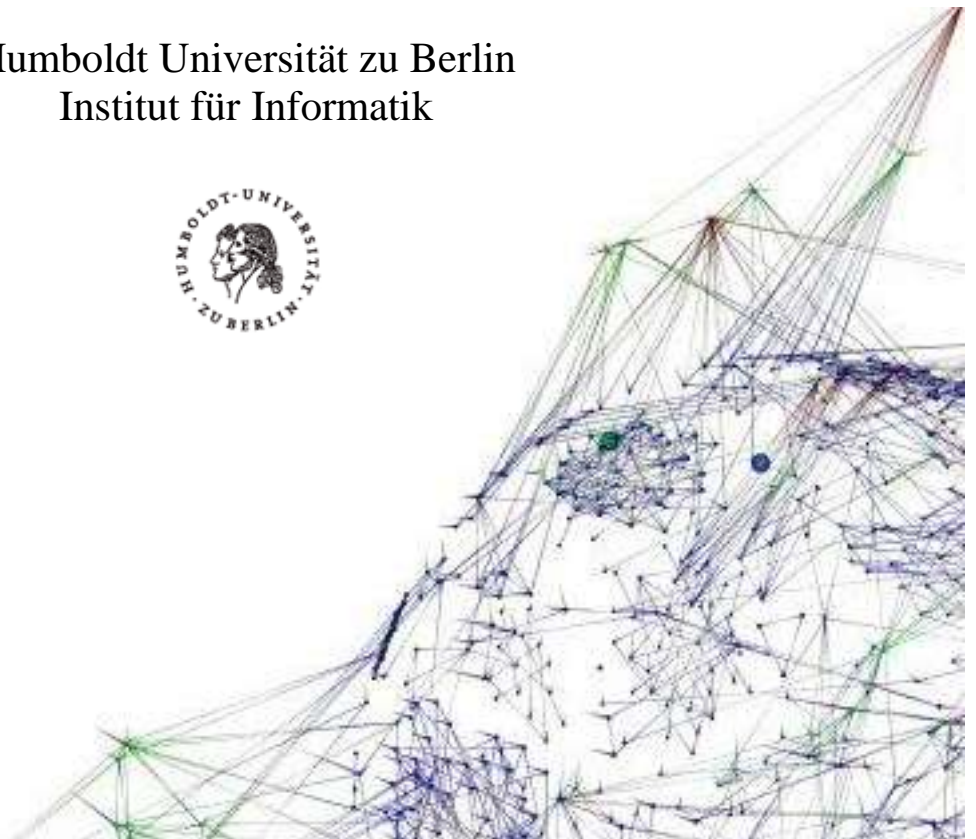
Prof. Mirsolaw Malek

und

Dipl. Inf. Peter Ibach

Humboldt Universität zu Berlin

Institut für Informatik



Abstract

Es werden verschiedene Verfahren untersucht, welche Benutzeraktionen bei Bearbeitung und Abruf von Webseiten in WikiWebs analysieren, um sowohl Seiten als auch Verbindungen zwischen Seiten durch ein Ranking zu bewerten. Die Analyse des Benutzerverhaltens erlaubt es zusätzlich, auf inhaltliche Zusammenhänge zwischen den Seiten zu schließen und so neue Verbindungen einzufügen und ebenfalls zu bewerten. Die Verfahren wurden unter anderem von Modellen des menschlichen Gehirns inspiriert. Über die Visualisierung als Graph mit gewichteten Knoten und Kanten lassen sich Suche und Navigation in Wissensnetzen verbessern.

Inhaltsverzeichnis

1 Einleitung	4
1.1 Verwandte Anwendungen.....	5
1.2 Ranking Verfahren.....	6
2 Nutzen	6
3 Berechnung des Linkranks	7
3.1 Probleme von Rankingverfahren.....	7
3.2 Modelle und Begriffe.....	8
3.3 Annahmen und Definitionen.....	8
3.4 Das Cooccurrence-Verfahren.....	9
3.5 Das Lernende Netz.....	9
3.6 Probleme.....	10
4 Anwendungen	11
4.1 Bewertung von Suchergebnissen, Page-Link-Rank.....	11
4.2 Ansichten nach Gruppen.....	11
4.3 Visualisierung als Graph.....	11
5 Fazit und Ausblick	12
6 Literatur und Verweise	12

1 Einleitung

Eine große Wissensbasis ohne mächtige Such- und Navigations-Funktionen erfüllt ihren Zweck – schnellen und einfachen Zugriff auf Wissen – schlecht oder je nach Umfang gar nicht. Ein gutes Beispiel ist das World Wide Web welches eine ungeheure Menge an Daten beinhaltet. Diese riesigen Datenmengen sind aber schlecht zu recherchieren. Zwar kann man über Suchmaschinen nach Begriffen suchen, aber die Qualität der gefundenen Seiten im Hinblick auf die Suchanfrage ist oft unbefriedigend und läßt sich erst nach Betrachten der Seiten beurteilen. Des weiteren kann die Formulierung der Suchanfrage problematisch sein, da die Suche rein syntaktisch ist und oftmals wichtige Seiten nicht mittels der eingegebenen Suchanfrage aufgefunden werden. Folgende Verbesserungen sollen die nachfolgend untersuchten Verfahren ermöglichen:


- Fehlendes oder unzureichendes Ranking:
Aktuelle Suchmaschinen benutzen zwar zumeist ein Ranking, doch hängt dieses in der Regel nur von den Inhalten der Seiten und den Verknüpfungen zwischen den Seiten ab. Zugriffe der Benutzer auf diese Seiten werden dagegen nicht berücksichtigt.
- Verwandte Themen suchen:
Wer nach Dokumenten über Drachen sucht möchte möglicherweise auch welche über Lindwürmer finden, aber keine über Drachenschiffe. Die Intentionen der Benutzer stochastisch abzubilden, ist über eine Traffikanalyse möglich.
- Ungenaue Suchabfragen formulieren:
Der Benutzer weiß nur ungefähr wonach er sucht und möchte die Suche Schritt für Schritt verfeinern, um so das Ziel „einzukreisen“. Das sind assoziative Verbindungen nötig, die den Benutzer zu verwandten Themen führen.

Um diese Probleme anzugehen, sind Verfahren zum Bewerten von Seiten und Links zwischen ihnen interessant. Im Gegensatz zum Page Rank wird beim Link Rank nicht ein Dokument, sondern die Verbindung – der Assoziationsgrad – zwischen Dokumenten bewertet. Oft wird versucht, den Page Rank und den Link Rank über inhaltliche Analysen zu erstellen oder eine manuelle Bewertungen durch die Autoren, durch die Leser oder durch spezielle Rezensenten vornehmen zu lassen. Solche Verfahren haben ihre Berechtigung, sind in vielen Fällen aber zu träge, aufwändig oder aus anderen Gründen ungeeignet. In dieser Arbeit geht es darum, aus dem Verhalten der Benutzer – dem Aufrufen und Betrachten von Dokumenten – das „Gewicht“ von Dokumenten und die „Stärke“ der Verbindung zwischen ihnen zu ermitteln. Dabei wird davon ausgegangen, dass zeitnah oder gleichzeitig betrachtete Dokumente Gemeinsamkeiten haben.

1.1 Verwandte Anwendungen

Aus dem zeitnahen Betrachten mehrere Dokumente Schlüsse zu ziehen ist als solches nichts Neues. So berechnen Amazon [3] oder iTunes [4]. Produkte welche den Kunden interessieren könnten anhand der Produkte, welche er und andere Kunden mit dem Aktuellen zusammen gekauft haben. Die Online Community Everything2 [5] benutzt ein etwas komplexeres Verfahren um so genannte „Softlinks“ zu berechnen. Von einer Seite ausgehende Softlinks verweisen auf die am häufigsten aufgerufenen folgenden Seiten.

Kunden, die diesen Artikel gekauft haben,



Concrete Mathematics
von Ronald L. Graham, u. a.
Durchschnittliche Kundenbewertung: ★★★★★
Versandfertig bei Amazon in 24 Stunden.

From Book News, Inc.
Graduate level textbook. Major topics include: sums; recurrences; integer functions; elementary number theory; binomial coefficients; generating functions; discrete probability; asymptotic methods. Annotation copyright Book News, Inc. Portland, Or.

Amazon-Preis: EUR 52,50

[Einkaufswagen](#)


[Auf meinen Wunschzettel](#)

Keine Kreditkarte? [Zahlen Sie einfach auf Rechnung!](#)

haben auch diese Artikel gekauft [Mehr Resultate](#)

Artikel anzeigen von:

- ▶ **Alle Produkte**
- ▶ Bücher (18)
- ▶ DVD (1)












 The Art of Computer Programming, 3 Vol. von Donald E. Knuth Mehr davon	 Computers and Intractability von Michael R. Garey, u. a. Mehr davon	 Introduction to the Theory of Computation von Michael Sipser Mehr davon
 Fifty Challenging Problems in Probability with Solutions von Frederick Mosteller Mehr davon	 A Course of Pure Mathematics von G. H. Hardy Mehr davon	 Computational Complexity. von Christos H. Papadimitriou Mehr davon
 Introduction to Algorithms, Second Edition von Thomas H. Cormen, Charles E. Leiserson Mehr davon	 Combinatorial Optimization von Christos H. Papadimitriou, Kenneth Steiglitz Mehr davon	 Challenging Mathematical Problems with Elementary Solutions, Vol. 1 (Challenging Mathematical Problems with Elementary Solutions) von A. M. Yaglom, u. a. Mehr davon

Abbildung 1 Folgende Bücher schlägt Amazon Interessenten von "Concrete Mathematics" vor

1.2 Ranking Verfahren

Es gibt eine Menge unterschiedlichster Ranking-Verfahren welche sich meist auf Seiten beziehen und nicht auf Links. Da man sie aber mit dem hier vorgestellten Verfahren gut kombinieren kann sollen sie hier kurz aufgezählt werden.

- **Page Ranking** (à la Google):
Der Google Page-Rank Algorithmus interpretiert einen Link von Seite *A* nach *B* als Stimme von *A* für *B*. Zusätzlich wird aber das Gewicht der Seite *A* in die Bewertung mit einbezogen (analog zum echten Leben, wo die Meinung eines bekannten Menschen meist auch mehr Aufmerksamkeit findet). Bewertet wird die Summe aller eingehenden Links einer Seite, gewichtet nach der Wichtigkeit der verweisenden Seiten.
- **Hit Counter:**
Hier wird die Anzahl der Hits, Aufrufe der Seite, als Qualitätsmaß interpretiert. Ein Hit Counter ist zwar trivial zu implementieren, allerdings besteht keine direkte Korrelation zwischen Qualität und Traffic, da ein Benutzer erst nach Aufruf der Seite weiß ob sie das Gesuchte bietet.
- **Voting Systeme:**
Bei einem Voting-System stimmen die Benutzer explizit für oder gegen eine Seite. Die Bewertung einer Seite ist dann die Summe aller Stimmen. Ein Voting-System hat den Vorteil, dass die Stimmen recht verlässlich sind, wobei die Stimmbeteiligung wichtig ist. Beispiele sind die Seiten von Slashdot und die Online-Community Everything2.
- **Autoren Ranking:**
Hier wird davon ausgegangen das ein „guter“ Autor auch in Zukunft „gute“ Artikel verfasst. Das Problem ist festzustellen ob ein Autor „gut“ ist. Dies ließe sich über die „Qualität“ seiner Beiträge bestimmen, welche aber über eine andere Methode berechnet werden muss oder über Voting.
- **Klick-Through Ranking:**
Beim Klick-Through-Ranking wird die Anzahl der benutzten Links auf einer Seite gezählt, unter Umständen normiert und die Summe dann als Qualitätsmaß betrachtet.
- **Edit-Rank (In Wikis):**
Hier wird bestimmt wie häufig eine Seite editiert wurde um z.B. Aussagen über Ihre Aktualität zu ermöglichen. Allerdings gibt es auch Themen die nicht altern (der Satz des Pythagoras z.B.).
- **Wiederkehrrate:**
Hier wird die Anzahl der wiederkehrenden Benutzer als Qualitätsmaß betrachtet. Dies basiert auf der Annahme, dass eine regelmäßig besuchte Seite nicht schlecht sein kann.
- **Word Counter:**
Die Anzahl des gesuchten Wortes in der gefundenen Seite wird hier als Qualitätsmaß interpretiert.

2 Nutzen

Wer schon einmal die tausenden Ergebnisse einer Google Suche durchforstet hat weiß das ein gutes Ranking sehr viel Zeit und Ärger sparen kann. Der wichtigste aber nicht einzige Nutzen ist also die Verbesserung der Suche und der Navigation. Weitere Nutzen wären

- **Experten finden:** Durch eine automatische Bewertung der Seiten und Seiten-Kategorien ist es möglich Experten für ein Thema zu finden. In einem Hilfesystem lassen sich so Experten zu bestimmten Themengebieten finden.
- **Implizites Wissen:** Mit dem aufgezeigten Verfahren lässt sich implizites Wissen bzw. Beziehungen zwischen Dokumenten aufzeigen. Durch den Algorithmus bilden sich Cluster von zusammenhängenden Themen. Visualisiert man diese gewichteten Zusammenhänge – z.B. in einem

Graph - so sind diese sehr leicht zu erkennen. Eine Navigation auf Basis des Graphen wäre ebenfalls denkbar.

- **Seiten Navigation:** Der Betreiber einer Seite kann durch die Auswertung sein Seiten-Layout so umgestalten das die Navigation erleichtert wird. Beliebtere Seiten können so direkt und nicht über Umwege erreicht werden. Bei der Deutschen Bank ist es z.B. recht schwer die Bankleitzahl einer Filiale zu ermitteln. Glücklicherweise wird auf der Seite eine Top-Ten der meist gesuchten Begriffe geführt. Diese wird seit Wochen von „Filiale“ und „Bankleitzahl“ angeführt.
- Nicht nur die Suche, auch das „Surfen“ wird verbessert. Möchte man einfach nur „mal so“ sehen was es an Informationen gibt, ist ein funktionierendes Assoziationsnetz recht angenehm.
- Ähnliche Pfade lassen auf gemeinsame Interessen schließen. Wertet man diese auf Gemeinsamkeiten aus, lassen sich **Interessen-Gemeinschaften** innerhalb einer Community erkennen.

3 Berechnung der Link Ranks

Da wir den Link Rank ohne aktiven Eingriff der Benutzer berechnen wollen, bieten sich Verfahren an die aus dem Verhalten der sie benutzenden Menschen lernen. Als Inspirationsquelle sollen die Hebbsche¹ Regel [6] und das Verhalten sozialer Insekten dienen. Die Hebbsche -Regel ist eine Lernregel für ein Neuronales Netz, in dem Verbindungen gestärkt werden, wenn sie die dazugehörenden Neuronen gleichzeitig aktiviert werden. Ameisen sind in der Lage als Kollektiv sehr komplexe Probleme zu lösen ohne das den Individuen klar ist, was sie tun. Die Lösung emergiert aus der einfachen Interaktion der Individuen mit Ihrer Umwelt und ihresgleichen. So finden Ameisen **Pfade** zu Futterquellen und sind in der Lage das „Traveling Saleman Problem“ recht gut zu lösen². Analog zu den Ameisen möchten wir Pfade erzeugen, nur das diese Pfade Relationen oder Assoziationen zwischen Dokumenten beschreiben und nicht den Weg zu einer Nahrungsquelle. Bewegt sich ein Benutzer von Seite A nach B so wird der Pfad markiert, existiert er schon so wird er verstärkt. Je mehr Benutzer nun einem Pfad folgen umso stärker wird er. Werden alle Pfade nach jedem Schritt normiert so werden selten benutzte Pfade mit der Zeit schwächer. Interpretiert man die Pfade als Assoziationen zwischen Themen so ergibt sich ein assoziatives Netz welches verwandte Themen verbindet.

Bevor die Verfahren weiter erläutert werden soll noch kurz auf einige Probleme hingewiesen werden.

3.1 Probleme von Rankingverfahren

Einige Probleme stellen sich bei fast jedem automatischem Ranking-Verfahren.

- **Die selbst erfüllende Prophezeiung:**
Dieses Problem tritt auf wenn aus verschiedenen Gründen eine Seite hoch bewertet und auf Grund dessen häufig besucht wird. Dies führt wiederum zu einem besseren Ranking. So kann der Titel einer Seite sehr viel versprechend sein obwohl der Inhalt sehr schlecht ist.

¹ Eine gute Zusammenfassung des Verfahrens findet sich hier: <http://home.zait.uni-bremen.de/~mn/old/glossar/g/node118.html>

² **Informationen über Ameisenpfade:**

http://www.techfak.uni-bielefeld.de/ags/wbski/lehre/digiSA/WS0304/IntAlg/Folien/Swarm_Intelligence.pdf

- **Startups :**
Eine neue Seite hat noch keine Besucher und damit auch noch kein hohes Ranking. Egal wie „gut“ die neue Seite ist, wird sie auf Grund ihres niedrigen Rankings wahrscheinlich unbekannt bleiben. Ein Lösungsansatz ist neue Seiten getrennt anzuzeigen wie Wikis dies tun und dann z.B. nach den Autoren zu Bewerten.
- **Spoofing:**
Der Autor einer „schlechten“ Seite kann Eigenwerbung betreiben. Entweder macht er dies selbst, was recht unwahrscheinlich ist, oder er lässt die Seite durch einen Roboter regelmäßig besuchen. Ist die Wissenssammlung für jeden zugänglich und sind die Besuche auf mehrere IP-Adressen verteilt so ist das Spoofing kaum zu verhindern.
- **Divergente Interessen:**
Was eine Gruppe von Benutzern für relevant erachtet kann von einer Anderen als irrelevant angesehen werden. Dieses „Problem“ kann aber auch als Vorteil genutzt werden. Z.B lassen sich damit ähnliche „Geschmäcker“ finden.

3.2 Modelle und Begriffe

Wir werden uns im Folgenden auf Ranking Verfahren für Wikis beschränken. Wikis haben einige interessante besondere Eigenschaften. Die meisten Wikis sind von jedem jederzeit editierbar, d.h. es gibt wenig Beschränkungen der Rechte. Aus diesem Grund sind Wikis im Gegensatz zu herkömmlichen Seiten recht **dynamisch**. Ein Vorteil für unser Ranking verfahren ist das die meisten Wikis auf einen Themenbereich **beschränkt** sind und es somit sehr wahrscheinlich ist das es viele verwandete Themen zu finden gibt. Des weiteren wird auf einer Wiki Seite meist nur ein Thema dargestellt.

3.3 Annahmen und Definitionen

- Unter **Dokumente** verstehen wir alles was sich mit einer URI Adressieren lässt, meist allerdings HTML-Dokumente.
- **Assoziationen** sind Verbindungen zwischen Dokumenten.
- Ein Wiki, oder allgemeiner eine Menge von verknüpften Dokumenten, kann als (un)gerichteter **Graph** betrachtet werden. Ob der Graph gerichtet oder ungerichtet sein soll hängt vom Verfahren und der Anwendung ab. Die Knoten werden als Dokumente, die Kanten als Assoziationen bezeichnet.
- Ein **Pfad** ist eine chronologisch sortierte Folge von besuchten Dokumenten.
- Assoziationen zwischen Themen sind meist Symmetrisch und Transitiv.
- Durch Bewertung der Links (Kanten) wird es möglich dem Graph eine Gewichtsfunktion zuzuordnen, wobei die Gewichte der Kanten die Stärke der Assoziationen darstellen.
- Im Allgemeinen wird sich jeder Benutzer „nur“ für einen Teilgraph interessieren
- Wird in einer Seite ein Link genutzt so kann davon ausgegangen werden, dass die Seite für brauchbar erachtet wurde. Durch benutzen eines Links auf einer Seite wird für diese „gestimmt“.
- Bei allen Verfahren ist das „Feedback“ durch die Benutzer wichtig.
- Der **Link Rank** einer **Assoziation** wird durch die folgenden beiden Verfahren direkt berechnet.
- Der **Link Rank** eines **Dokumentes** ist die – unter Umständen gewichtete oder normierte - Summe aller **Assoziationen** von und zu der Seite.

Nach den Definitionen sollen nun die beiden von Bollen und Heylighen beschriebenen Verfahren vorgestellt werden.

3.4 Das Cooccurrence-Verfahren

Folgende Verfahren wurden 1996 [1] und 1999 [3] von Bollen & Heylighen vorgestellt. Dabei wird der Zusammenhang zwischen zwei Dokumenten wie folgt definiert: Zwei Dokumente sind miteinander verwandt, wenn sie gleichzeitig bzw. in kurzem Abstand aufgerufen werden. Je häufiger zwei Dokumente „gleichzeitig“ betrachtet werden desto stärker hängen sie zusammen. Der Link Rank der Assoziation berechnet sich dann wie folgt:

$$LR(A, B) := P(A|B) = \frac{Hits(A \wedge B)}{hits(A)},$$

wobei $Hits(A)$ die Summe aller Aufrufe des Dokumentes A und $Hits(A \wedge B)$ die Summe aller zeitnahen Aufrufe von A und B bezeichnet.

Die Definition von *Cooccurrence* setzt zwar Parallelität voraus – die Verbindung wird verstärkt wenn zwei Dokumente gleichzeitig präsentiert werden - der Mensch kann aber nur ein Dokument auf einmal lesen. Und der Zugriff im Netz erfolgt meist sequentiell. Solange das Zeitintervall zwischen den Aufrufen von A und B klein genug ist lässt sich auch aus sequentiell Zugriff auf den Zusammenhang von Themen schließen.

Die Cooccurrence lässt sich sehr leicht berechnen. Ein Nachteil der *Cooccurrence* ist, dass weder der Grad der Symmetrie, noch der der Transitivität getrennt bestimmt werden kann. Um Symmetrie und Transitivität beeinflussen zu können bietet sich das Lernende Netz Verfahren an. Welches nun erläutert wird.

3.5 Das Lernende Netz

Das „lernende Netz“ lernt, wie der Name andeutet, aus dem Verhalten der Benutzer. Zu Beginn wird das Gewicht aller Kanten auf 0 gesetzt. Der Algorithmus wendet nun in jedem Schritt folgende Regeln an. Navigiert ein Benutzer entlang eines Pfades $\dots a \rightarrow b \rightarrow c \dots$ so werden die direkten Verbindungen $(a \rightarrow b)$ und $(b \rightarrow c)$ um einen festen Wert f erhöht (Frequenzregel). Die symmetrischen Verbindungen $(b \rightarrow a)$ und $(c \rightarrow b)$ werden um den Wert $s < f$ erhöht (Symmetrieregeln). Als letztes werden die transitiven Verbindungen $(a \rightarrow c)$ um den Wert $t < f$ erhöht (Transitivitätsregel). Die Symmetrie- und Transitivitätsregel sind für die „Kreativität“ des Algorithmus wichtig, da sie bestimmen wie schnell neue Links erzeugt werden.

Werden die Bewertungen nach jedem Schritt normiert so werden selten benutzte Pfade im Vergleich zu den stärker frequentierten Pfaden abgeschwächt. Die Normierung kann z.B. durch einfaches Dividieren des Link-Ranks durch die Anzahl der gesamten Hits realisiert werden.

Als Erläuterung dient ein kleines Szenario in dem ein Benutzer die Seiten A , dann B und danach C besucht. Die Werte für die Frequenz-, Symmetrie- und Transitivitätsregel seien $f=1; s=0.7; t=0.25$. Alle Link-Ranks sind mit 0 initialisiert.

1. Schritt. Die Seite A wird aufgerufen und der Hit-Count von A um 1 erhöht.

	A	B	C
A	1	0	0
B	0	0	0
C	0	0	0

2. Schritt. Nach A wird B aufgerufen. Der Hit-Count von B wird um 1 erhöht, die Frequenzregel wird auf (A,B) und die Symmetrieregelle auf (B,A) angewendet

	A	B	C
A	1	1	0
B	0.7	1	0
C	0	0	0

3. Schritt. Als letztes wird die Seite C aufgerufen. Ihr Hit-Count wird um 1 erhöht und sowohl Frequenzregel, Symmetrieregelle als auch die Transitivitätsregel werden angewendet. In der Haupt-Diagonalen stehen nun die Hit-Counts der jeweiligen Seite und in den Elementen i,j der jeweilige Link-Rank (i,j)

	A	B	C
A	1	1	0.25
B	0.7	1	1
C	0	0.7	1

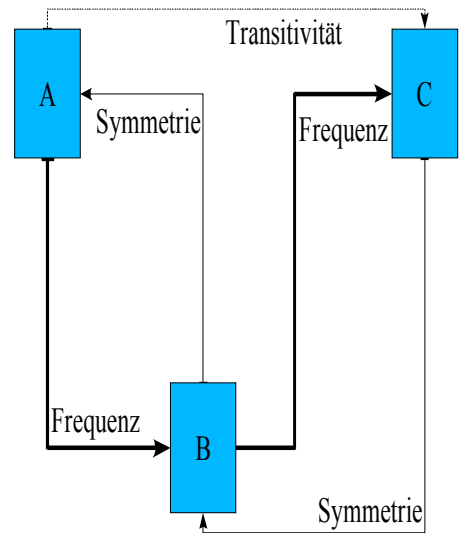


Abbildung 2 Beispiel für einen Pfad

Dieses Verfahren hat gegenüber der Cooccurrence den Vorteil das die Stärke der Symmetrie und Transitivität explizit bestimmt werden können. Dadurch gewinnt der Algorithmus an „Kreativität“ da mehr neue Verbindungen entstehen. Die Kreativität kann direkt durch die Parameter s und t beeinflusst werden.

3.6 Probleme

Auch wenn einige der oben angeführten allgemeinen Probleme abgeschwächt werden ist das vorgestellte Verfahren nicht Perfekt. Folgende Probleme liegen auf der Hand

- **Anfangsproblem:**
Der Algorithmus benötigt mehrere zusammenhängende Seiten und möglichst viele Benutzer. Am Anfang sind alle Links gleichwertig. Eine gute Bewertung lebt davon, dass die Seiten viel und „bewusst“ genutzt werden. Existieren Logfiles zu den Wikis so kann der Algorithmus mit diesen Daten initialisiert werden.
- **„Noise“:**
Durch die Benutzung werden neue Links erzeugt von denen einige sicherlich als „Noise“ einzustufen sind. Da diese aber nicht weiter benutzt werden, werden sie im Vergleich zu den anderen Links schwächer und nicht mehr betrachtet. Das „Signal To Noise Ratio“ hängt sicherlich von den Werten t und s ab.

- **Missbrauch (Spoofing):**
Bei diesem Ranking kann die Bewertung einer Seite verbessert werden, in dem ein Agent wiederholt die Seite und alle von der Seite ausgehenden Links besucht.

4 Anwendungen

Mittels der gewonnen Assoziationen können neue Anwendungen realisiert werden. Als Ausblick sollen hier einige Anwendungsszenarien vorgestellt werden.

4.1 Bewertung von Suchergebnissen

Eine naheliegende Anwendung ist mittels des Link-Ranks der Assoziationen einen Page-Rank zu berechnen. Angenommen ein Benutzer sucht nach dem Wort w . Die Liste der gefundenen Seiten kann wie folgt sortiert werden. Jede Seite A welche das Wort w enthält wird mit

$$Pr(A) = R(A) * (\sum_{i \in assoziationen(A)} Lr(A, i)) / |assoziationen(A)|$$

bewertet, wobei $R(A)$ die Anzahl der gefundenen Suchwörter w in der Seite A ist. Dieses Ranking kann entweder über die Reihenfolge der gefundenen Seiten oder durch einfache Icons, wie es z.B. das Heise-Diskussionsforum macht, dargestellt werden. Dort werden beliebte Diskussionsfäden mit einem grünen, nach rechts wachsenden und unbeliebte Diskussionsfäden mit einem roten, nach links wachsenden Icon dargestellt.

4.2 Ansichten nach Gruppen

Wertet man die Pfade der Benutzer getrennt aus, so lassen sich verschiedene „Ansichten“ auf das Wiki generieren. Diese spiegeln die Wertung der Themen und Assoziationen durch andere Benutzer und ermöglichen neue Perspektiven.

Da eine „Ansicht“ ja recht viel über den „Geschmack“ des Benutzers aussagt, kann diese Information benutzt werden um Gruppenbildung zu unterstützen. Vergleicht man die verschiedenen Gewichtungen der Kanten aller Benutzer, so lassen sich Benutzer mit ähnlichen Interessen finden.

4.3 Visualisierung als Graph

Ein interessanter Ansatz wäre es die Seiten und die „Assoziationen“ zwischen ihnen als Graph zu visualisieren. Die „Wichtigkeit“ der Themen und deren „Assoziationen“ lassen sich mit Farben, Umfang, oder Entfernung der Kanten bzw. Knoten darstellen. Diese Visualisierung kann benutzt werden um einen Überblick über die Themen und deren „Assoziationen“ zu gewinnen. Eine schrittweise Suche liesse sich damit ebenfalls implementieren. Es gibt schon einige Ansätze die Verbindungen zwischen Seiten – oder allgemeiner zwischen Themen - zu visualisieren. Allerdings fehlt bei allen bisherigen Verfahren die Bewertung der Kanten. Mit dem hier vorgestellten Verfahren hat man aber ein Ranking für Seiten und Links und kann damit die Visualisierung bedeutend verbessern.

Einen ähnlichen aber viel einfacheren Ansatz verfolgt der Wiki Browser (oder Google Browser) von Touchgraph³. Dort werden alle Dokumente und Ihre Verbindungen in einer Graphenansicht angezeigt. Diese Ansicht berücksichtigt aber nur die von den Autoren angelegten Links zwischen den Dokumenten. Der Wiki Browser unterstützt kein Ranking, der Google Browser das von Google berechnete.



Abbildung 3 Touchgraphs Wiki Browser

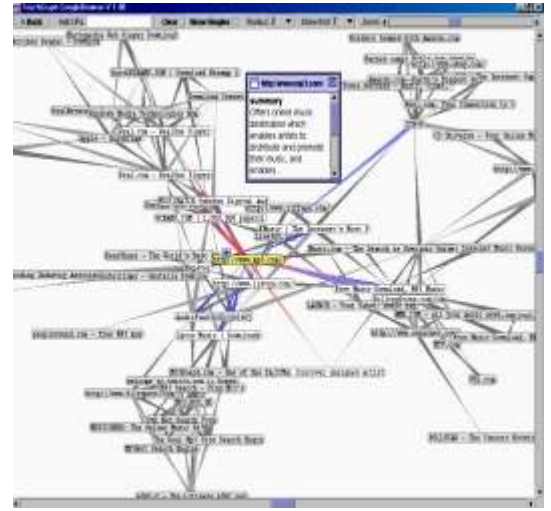


Abbildung 4 Touchgraphs Google Browser

Beide Beispiele zeigen was ohne Link Rank mit Visualisierung alles möglich ist. Eine zusätzliche Information, nämlich die Stärke der Verbindungen zwischen den Knoten, erhöht die Aussagekraft zusätzlich.

5 Fazit und Ausblick

Assoziativ lernende Wikis oder Wissensnetze stellen eine interessante Methode zum Auffinden und Bewerten von „Assoziationen“ zwischen Themen bzw. Seiten dar. Sie ermöglichen eine kollektive „Mind Map“ zu generieren. Diese „Assoziationen“ werden durch das Verhalten der Benutzer und dem Feedback zwischen Benutzer und System erzeugt ohne den Benutzer zu stören. Mit diesen Bewertungen ist es möglich – wie in einigen Beispielen gezeigt - Wikis oder Wissensnetze zu verbessern und mit neuer Funktionalität zu bereichern. Die Wertung der Assoziationen ist natürlich subjektiv, was aber in der Natur der Sache liegt und auch gewollt sein kann, da jede Bewertung nur Subjektiv sein kann.

6 Literatur und Verweise

- [1] Bollen J. & Heylighen F. (1996): [Algorithms for the self-organisation of distributed, multi-user networks. Possible application to the future World Wide Web](#), in: Cybernetics and Systems '96, R. Trappl (ed.), (Austrian Society for Cybernetics), p. 911-916.
- [2] Heylighen F. (1999): "[Collective Intelligence and its Implementation on the Web](#): algorithms to develop a collective mental map", Computational and Mathematical Organization Theory 5 (3), p. 253-280.
- [3] Der Onlinehändler Amazon: www.amazon.de bzw <http://www.amazon.com>
- [4] Der Onlinemusikhandel von Apple, iTunes, ist unter www.itunes.com zu finden.
- [5] Die Online Community Everything2 ist unter www.everything2.com zu finden.
- [6] Donald Hebb (1949): Organization of behavior. Wiley, New York

³ TOUCH-Graph: <http://www.touchgraph.com/>

