

Bachelorarbeit
Themennetzwerke und Emergenz von Struktur
am Beispiel vom NIO/meiNetz-Projekt

Universität Osnabrück
FB Humanwissenschaften
Studiengang Cognitive Science

Ingo Frost
5. Fachsemester
Matrikelnr.: 874029
ingo.frost@uni-osnabrueck.de

Erstgutachter
Dr. habil. Helmar Gust

Zweitgutachter
Dr. habil. Kai-Christoph Hamborg

19. Oktober 2004

Zusammenfassung

Große Verzeichnisse werden meist durch hierarchisch gegliederte Kategorien überschaubar gemacht. Die Kategorienhierarchie kann a priori nicht durch objektive Maßstäbe zustandekommen, was unverweigerlich zu Schwierigkeiten führt. Nachdem diese analysiert werden, wird am Beispiel des Linkverzeichnisses des NIO-Portals (<http://www.nio.uos.de>) ein Themennetzwerk als Alternative vorgestellt, dessen Themenrelationen emergent aus dem Verhalten der Benutzer innerhalb des Portals hervorgehen sollen. Dabei konzentriere ich mich auf den Übergang von einem rein assoziationsartigen Netzwerk zu einem stärker strukturierten Netzwerk, bei dem neben Verknüpfungsstärke zwischen zwei Gruppen auch Verknüpfungsarten hervorgehen: ist-Untergruppe, ist-Obergruppe und ist-assoziativ. Ein in PHP implementierter Algorithmus wird dazu vorgestellt.

Inhaltsverzeichnis

1 Die Problematik	3
1.1 Definitionen	3
1.2 Relationen	5
1.3 Traditioneller hierarchischer Ansatz	6
1.4 Netzwerkansatz	8
1.5 Reduktion von Komplexität	10
1.6 Agenda (A1 bis A6)	11
2 NIO: Eine benutzeradaptive Informationsplattform	13
2.1 Grundgedanke	13
2.2 Aufbau von NIO	13
2.2.1 Themennetzwerke	14
2.2.2 Soziale Netzwerke	14
2.2.3 NIO-Gruppen	16
2.2.4 Datenstruktur	17
2.3 Darstellung und Dynamik	18
2.3.1 Umgang mit Gewichten	18
2.3.2 Portaloberfläche	22
2.3.3 Benutzermodell	25
3 Erweiterungsansätze zur Typisierung von Relationen	27
3.1 Problembeschreibung des existierenden Systems	27
3.2 Lösungsvorschläge	28
3.3 Ansatz nach Benutzung der Verknüpfungen	29
3.3.1 Programmskizze	29
3.3.2 Schritt I: Wie Klicks zu Gewichten werden	31
3.3.3 Schritt II: Vergleich der beiden Gewichte pro Richtung	32
3.3.4 Schritt III: Gliederungsanalyse der Teilgruppen	33
3.3.5 Implementation: subnet.class.php	34
3.4 Konzept zur Visualisierung	36
3.4.1 Ansatz auf Textbasis	36
3.4.2 Grafischer Ansatz	37
4 Ausblick	38
Literatur	40
Anhang: Programmquelltext	41

1 Die Problematik

In vielen Bereichen des Alltags in der Informationsgesellschaft sammeln sich enorme Mengen von Daten an. Eine der wichtigsten Aufgaben und Herausforderungen besteht darin, einen Überblick über sie zu behalten. Eine Grundlage dabei bildet das meiNetz-Projekt[FROST 2001b], das versucht, Internetlinks innerhalb einer Benutzergemeinschaft zu verwalten und besser zugänglich zu machen. In diesem Kapitel stelle ich die theoretischen Hintergedanken vor, die bei der Gliederung von Informationen in Form von Verzeichnissen auftreten und die auf alle informationsverarbeitende Systeme zutreffen. Nicht zufällig ergeben sich dabei Parallelen zu Modellen menschlicher Kognition oder Arbeitstechniken wie das Mindmapping[BUZAN . BUZAN 2002] als alternativer Ansatz Assoziationen räumlich anzuordnen und so zu einer Gliederung zu gelangen. Unser Gehirn hat Fähigkeiten entwickelt, mit denen es z. B. bei der Wahrnehmung umfangreiche und komplexe Signale unserer Sinnesorgane vereinfachen, komprimieren und schließlich *erkennen* kann. Ich werde versuchen Ideen innerhalb dieser Analogie auf computerbasierte Informationssysteme zu übertragen. Gleichzeitig werden viele Kritikpunkte zu den momentan vorhandenen hierarchischen Informationssystemen (siehe 1.3) deutlich, und es lässt sich erklären, warum bei deren Handhabung Schwierigkeiten auftreten.

1.1 Definitionen

Zuerst unterscheide ich zwischen den Dingen, die geordnet werden sollen und den Ordnungen in Form von Themengruppen (im Zusammenhang mit dem Netzwerkansatz), oder Kategorien (im Zusammenhang mit dem hierarchischen Ansatz), die beide allgemeiner als Mengen bezeichnet werden. Im Weiteren werde ich deshalb von Elementen und Mengen in Analogie zur Mengenlehre sprechen. Im Verlauf der Arbeit wird von Verweisen (auf wohlunterscheidbare Objekte, die Elemente) und (Themen-)Gruppen/Kategorien (verschiedene Mengen der Elemente in Form von Verweisen) die Rede sein.

Am Anfang steht also die Definition für ein Informationssystem:

- Was sind die Elemente?
- Was repräsentieren die Mengen der Elemente?

Zuerst die Unterscheidung zwischen Element und Verweis. Da die meisten Elemente sich nicht physisch zu Mengen anordnen lassen oder dies einfach nur denkbar unpraktisch wäre, werden stellvertretend Verweise auf diese Elemente angeordnet, die eindeutig auf ein Element *verweisen*. Diese Verweise können z.B. in Form von Adressen, Namen oder Seriennummern vorliegen. Hier beziehe ich mich ausschließlich auf Elemente gleicher Art, mit dem Ziel einer strukturierten Linkebene (NIO¹ wäre formell auch in der Lage verschiedene solcher Verweisebenen mit denselben (Themen-)Gruppen zu modellieren). Es handelt sich also um klassische Verzeichnisse für Personen, Webseiten, Vorlesungsmaterial, Bücher etc., die dann die Rolle von Elementen einnehmen, die zu verschiedenen Mengen zusammengestellt werden können.

¹NIO: *N*etzwerk zur *M*odellierung von *I*ndividualität und *O*rientierung in der *I*nformationswelt

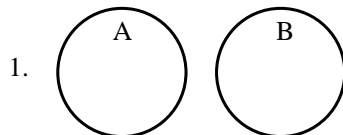
Das Zusammenschließen von Elementen zu Mengen kann nach einzelnen formalen Kriterien, also z. B. in Form von statischen Eigenschaftslisten, die auf alle Elemente pro Menge zutreffen, aufgestellt werden. Danach würden erst die Kriterien festgelegt, die die Mitgliedschaft von Elementen in einer Menge definieren und dann würden nach und nach neue Elemente in die Menge aufgenommen werden. Hier entsteht das Problem, dass die Elemente und ihre Eigenschaften noch nicht genau bekannt sind und sich somit keine Kriterien für eine solche Menge aufstellen lassen. Deshalb wäre es besser diese Mitgliedschaftskriterien flexibel zu lassen und nicht am Anfang zu definieren, denn möglicherweise werden später andere Kriterien wichtig oder sogar die Kombination verschiedener Kriterien. Ähnlich wie bei den menschlichen spontanen Kategorien ('things-to-sell-in-a-garage-sell')[EYSENCK . KEAN 2000] ist also eine Zuordnung zu einer Gruppe, wie eine Eigenschaft, die mehrere Elemente gemeinsam haben, anzusehen. Welche Eigenschaften wichtig sein könnten, hängt sowohl von den Elementen als auch von dem Kontext ab. Die Forderung besteht also darin, pro Element mehrere Eigenschaften zuzulassen, die durch mehrere Zuordnungen zu verschiedenen Gruppen modelliert wird und sich ändern kann.

1.2 Relationen

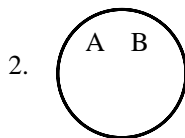
Generell lassen sich hier zwei Arten von Relationen unterscheiden: Die Zuordnungen der Elemente zu den Mengen und, sobald viele solcher Mengen in Form von Gruppen vorhanden sind, Beziehungen zwischen den Gruppen. Auch an dieser Stelle sollte nicht zu viel eingeschränkt werden und eine freie Art der Zuordnungen ermöglicht bleiben. So stellen sich „many-to-many“-Beziehungen für die Verweise und Themengruppen als optimal heraus: Ein Verweis kann beliebig vielen Gruppen zugeordnet werden, jede Gruppe kann beliebig viele Verweise aufnehmen. Bei den Beziehungen zwischen den Gruppen wird es komplizierter. Ich tendiere dazu nur einzelne beidseitige Beziehungen zu benutzen, diese aber mit verschiedenen Attributen pro Richtung auszustatten. Bisher habe ich keinen Fall gefunden, der zeigt, dass nur eine Richtung notwendig ist. Ich beleuchte die Art der Relationen zwischen Gruppen mit Hilfe der Mengenlehre:

In welchen Relationen können zwei Mengen zueinander stehen?

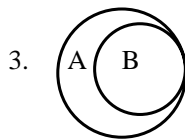
- Fall 1: A und B haben die leere Menge als Schnitt, also keine gemeinsamen Elemente.



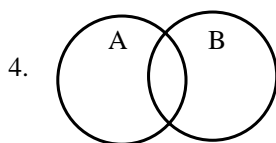
- Fall 2: Identität, also zwei Namen für ein und dieselbe Menge, die Elemente von A entsprechen denen von B.



- Fall 3: B ist Teilmenge von A ($B \subset A$): B kann wiederum Teilmengen haben und ist somit transitiv. Die Relationen sind gerichtet ab- bzw. aufsteigend.



- Fall 4: A und B haben eine Schnittmenge ($A \cap B$): Die Relationen sind ungerichtet und ein Ausdruck von Ähnlichkeit.



Normalerweise spielen alle vier Fälle eine Rolle, es sei denn Informationen sind künstlich nach speziellen Kriterien gegliedert. In diesem Falle wird die vierte Relation ausgeschlossen. Wenn man z. B. versucht Lebewesen zu klassifizieren, dann lässt sich durch Analyse der Gene eindeutig bestimmen, welches aus welchem hervorgegangen ist. So wird der Faktor Zeit² als wichtigstes Kriterium festgelegt und dient als lineare Skala. Es entsteht eine baumartige³ Struktur, aufgebaut aus Relationen von Fall 3. Wichtig dabei ist, dass nur ein Kriterium zum Aufbau dieser Systematik benutzt wird. Wollte man in einer solchen Systematik zusätzlich die morphologische Ähnlichkeit aufnehmen, so wären ganz verschiedene Entwicklungszweige in manchen Lebewesen plötzlich sehr nahe beieinander: Nun ist eine Netzstruktur notwendig, die verschiedene Relationstypen (genetisch und morphologisch ähnlich) und die ein Ähnlichkeitsmaß durch Relationen nach Struktur 4 modellieren kann.

So wurde beispielsweise erst durch die DNA-DNA-Hybridisierung herausgefunden, dass zwei morphologisch ähnliche Pandaarten zu verschiedenen Familien gehören. Der große Panda (/Ailuropoda melanoleuca/) ist tatsächlich ein Vertreter der Großbären und nicht, wie morphologisch angenommen, ein Kleinbär. Der kleine Panda (/Ailurus fulgens/) gehört jedoch zu einer anderen Familie, nämlich der der Kleinbären[CAMPBELL 2000].

Bei Datenstrukturen auf der Dateiebene werden auch rein hierarchische (Fall 3) Relationen benutzt⁴. Eine Datei ist dabei ein Element, das sich genau in einem Verzeichnis befindet und jedes Verzeichnis kann Unterverzeichnisse haben. Die Wahl einer solchen Struktur hatte physikalische Gründe: Auf den Datenbändern wurden die Dateien und Verzeichnisse sequentiell abgelegt. So entspricht die physische Adresse einer Datei auf einem solchen Band der logischen Position im Dateisystem. Wollte man also eine Datei von einem Verzeichnis in ein anderes verschieben, müsste die Datei ihren Speicherort wechseln.

Dieser Ansatz ist analog zu einer Bibliothek: Wollte man die Klassifikation eines Buchs ändern, müsste man seinen Standort wechseln.

1.3 Traditioneller hierarchischer Ansatz

Ich werde diesen Ansatz mit Hilfe des Bibliotheksproblems beschreiben, da es sich hier um ein typisches Problem handelt, dessen Struktur sich gut auf andere Bereiche übertragen lässt: Die Bücher als Elemente müssen zugänglich gemacht werden, indem sie z. B. Kategorien (Mengen) aus der Systematik zugeordnet werden.

In Bibliotheken wird eine Systematik benutzt, die monohierarchisch aus einer Vielzahl von verschachtelten Kategorien aufgebaut ist. Elementar ist dabei, dass jedes Buch nur einmal vorhanden ist und sich durch seinen Standort klassifiziert. Jedes Buch hat also *genau einen* Standort und somit *genau eine* Zuordnung in der Systematik. Damit die Bücher auch nach anderen Kriterien, wie z. B. nach Autoren auffindbar werden, wird pro Buch ein Zettel, der auf den Standort des

²Im KI-Bereich der Suche findet man z. B. die Transformation von einem netzwerkartigen Graphen zu einem Baum: Der Graph repräsentiert den Suchraum und durch den Faktor Zeit kann dieser in einen Suchbaum transformiert werden. Die fortlaufenden Zeitschritte entsprechen dann den Ebenen im Suchbaum.

³Auch ein Baum ist elementar an den Faktor Zeit gebunden, was durch die immer weitergehende Verästelung der Verästelung beim Wachsen des Baumes strukturell deutlich wird.

⁴Ausnahme bilden dabei Links, die in manchen Dateisystemen ermöglichen verschiedene Teiläste eines Verzeichnisbaums zu verbinden.

Buchs verweist, angefertigt und nach einem weiteren Kriterium sortiert. Hätte man alle Bücher doppelt, könnte man die einen Bücher thematisch-systematisch, die anderen nach Autoren ordnen. Da diese Alternative nicht praktikabel ist, werden Zettelkästen benutzt.

- **Kennzeichen**

Der Standort entspricht der Klassifikation: Sobald das Buch in die Systematik eingeordnet worden ist, ergibt sich daraus ein eindeutiger Standort.

- **Stärken**

Sobald man den Aufbau der Systematik kennt, lassen sich Bücher relativ einfach finden. Eine weitere Stärke zeigt sich bei Büchern, die sich thematisch recht eindeutig einsortieren lassen, sie kann man dann relativ einfach finden.

An dieser Stelle sei auch auf die Ökonomie hierarchisch organisierter Systeme innerhalb von Computermodellen hingewiesen. Durch die Verschachtelung von Themen wird eine eindeutige Zuordnung erreicht. Im Hinblick auf die Anzahl der nötigen Zuordnungen, den Platzbedarf und den Zugriff ist dieser Ansatz sowohl effizient, als auch elegant und wird häufig in der Informatik eingesetzt.

- **Schwächen**

Die Schwächen dieses Systems ergeben sich aus der Tatsache, dass die Klassifikation relativ statisch ist: Die hohe Effizienz geht auf Kosten von Mehrdeutigkeit⁵ und Dynamik. So ist es problematisch ein Buch in eine andere Kategorie zu sortieren, in der es möglicherweise momentan sehr relevant ist. Dieses Problem besteht auch für die Struktur der Kategorien, die sich in manchen Bereichen ändern kann. Man denke z. B. an das Aufkommen der Disziplin Cognitive Science, die möglicherweise anfangs in einer Bibliothek der Psychologie untergeordnet war und dann zu einer eigenständigen Disziplin wurde.

Der zweite Problembereich hängt damit zusammen, dass jedes Buch nur genau einer Kategorie zugeordnet ist. Da diese Zuordnung nicht völlig objektiv⁶ sein kann, werden manche Bibliotheksbesucher zwangsläufig ein Buch in einer *falschen* Kategorie suchen. Je nach Kontext macht jedoch eine Zuordnung zu mehreren Kategorien Sinn: Ein Psychologiestudent, der sich im Rahmen einer Psychologievorlesung mit Cognitive Science befasst, wird Bücher in den Psychologiekategorien suchen, während jemand, der Cognitive Science grundständig studiert, einen eigenen Bereich mit Büchern zur Kognitionswissenschaft erwartet. Es wird normalerweise aus verschiedenen Kontexten heraus gesucht, aber nur die Suche in einem - willkürlich festgelegten - Kontext führt zum gewünschten Resultat.

Es führt also nur ein Weg durch das hierarchische System zu einem Buch, obwohl es logischer wäre, wenn man eine Kategorie auf verschiedenen Wegen erreichen könnte.

⁵Manche Bücher können z. B. verschiedene Themen kombinieren und sind in diesem Sinne thematisch mehrdeutig, können jedoch nur in eine Kategorie, die nur ein Thema repräsentiert, eingeordnet werden.

⁶Ein Bibliothekar entscheidet eigenständig in welche Kategorie ein neues Buch aufgenommen wird. Dazu gibt es keine objektiven Maßstäbe: Es wäre durchaus denkbar, dass Bibliotheksbesucher dieses Buch an anderen Stellen eingeordnet hätten und so in einer anderen Kategorie suchen. Das Problem ist also die subjektive Entscheidung des Bibliothekars.

Zusätzlich besteht das Problem, dass die Kriterien für die Einordnung der Bücher vorher festgelegt werden. Entscheidet man beispielsweise beim Aufbau einer Bibliothek Bücher thematisch zu ordnen, berücksichtigt man nur ein Kriterium. Durch die hierarchische Ordnung ergibt sich immer nur ein Weg zum Ziel, nämlich der des Abstiegs durch die Hierarchieebenen, die ebenfalls nicht objektiv⁷ festgelegt werden können.

- **Einsatzbereiche**

Eine Vielzahl von Informationssammlungen ist auf diese Weise geordnet (siehe z.B. [INSTITUTION 1985]). Dazu gehören Bibliotheken mit Medien aller Arten und hierarchische Webverzeichnisse wie Yahoo[YAHOO]⁸ oder Dmoz[dmoz], sowie Informationssysteme wie [WÄTJEN . 1998],

- **Suchen und finden**

Die Stichwortsuche wurde früher durch Zettelkästen mit alphabetisch sortierten Stichwörtern ermöglicht, die heute meist durch Software ersetzt sind, aber prinzipiell genauso funktionieren⁹.

Desweiteren ist ein *Stöbern* möglich, indem man in einem Regal nachsehen kann, welche Bücher zu einem Thema vorhanden sind und an Hand des Angebots ein Buch auswählen kann. Diese Art der Suche ist manchmal schwierig, da man möglicherweise in einer falschen Kategorie stöbert und sich thematisch ähnliche Kategorien an völlig anderen Orten in der Bibliothek befinden.

- **Umgang mit Neuem**

Die Entscheidung, wo ein neues Buch einsortiert wird, ist weder objektiv noch nach der Meinung der Mehrheit der Bibliotheksbesucher organisiert, sondern wird von speziell dafür ausgebildetem Bibliothekspersonal erledigt. Genau dieses Problem ergibt sich auch bei der Erstellung neuer Kategorien und Unterkategorien. Es gibt keine objektive Methode um festzustellen, wann eine neue Kategorie eingeführt werden sollte und wie sie in der Hierarchieebene anzusiedeln ist.

1.4 Netzwerkansatz

Da die Bibliothek durch das einmalige Vorhandensein jedes Buchs, das an einen festen Standort gebunden ist, physisch beschränkt ist, werde ich den Netzwerkansatz anhand einer digitalen Bibliothek vorstellen, in der jedes Buch in digitaler Form vorliegt.

Analog zum Fall 2 (siehe Kapitel 1.2 Relationen), bei dem *eine* Menge *zwei* Namen hat, also durch zwei verschiedene Bezeichner referierbar ist, lassen sich auch bei einem digitalen Buch relativ einfach mehrere Verweise auf seinen Inhalt erstellen. Die Mengen werden nun nicht mehr als Kategorien, sondern als Themengruppen bezeichnet.

⁷ Auch hier haben Bibliothekare subjektiv, entschieden welche Kategorie als Unterkategorie einer anderen zu sehen ist. Die Bibliotheksbesucher können durchaus eine andere Vorstellung von der Ordnung der Dinge haben.

⁸ Dieses Problem ist Yahoo selbstverständlich bewusst und so ist Yahoo heute nicht mehr rein hierarchisch strukturiert, sondern benutzt auch Querverbindungen zwischen Kategorien in verschiedenen Ästen.

⁹ Ausnahmen sind Bibliothekssysteme wie OSIRIS[RONTHALER 1998] an der Universität Osnabrück, die weit komplexer sind und viele weitere Funktionen zur Verfügung stellen.

- **Kennzeichen**
Standort und Klassifikation sind unabhängig: beliebig viele Verweise zeigen auf die Bücher.
- **Stärken**
Dadurch dass ein Buch mit Hilfe seines Verweises und dessen Zuordnungen zu Themengruppen mehreren Themen zugewiesen werden kann, ergibt sich die Möglichkeit dieses auch aus verschiedenen Themengruppen, somit auch verschiedenen Kontexten, heraus aufzufinden. Neue Zuordnungen von Verweisen zu Themengruppen als auch neue Themengruppen können ohne das bisher Vorhandene anpassen zu müssen, hinzugefügt werden. Durch die Gewichtung dieser Zuordnungen ist ein wesentlich differenzierteres Verzeichnis mit Änderungen dieser Gewichte möglich. Daraus ergibt sich eine hohe Dynamik. Im hierarchischen Modell muss man bei einem neuen Buch *genau eine* passende Kategorie finden. In einem Netzwerk können *verschiedene* Auffassungen über die passende Themengruppe nebeneinander existieren. Dies gilt auch für die Verknüpfung der Themengruppen untereinander, sodass nun eine Themengruppe durch verschiedene Wege im Netzwerk erreichbar ist.
- **Schwächen**
Wer gewohnt ist mit der Systematik zu arbeiten und diese genau kennt, wird Schwierigkeiten haben, sich auf eine andere sich verändernde Netzwerkstruktur umzustellen. Dieser vermeintliche Schwachpunkt kann jedoch durch den Netzwerkansatz vermieden werden, denn selbst verschiedene Systematiken können zusätzlich in einer Netzstruktur dargestellt werden, indem in einem Bereich (für eine Systematik) die Relationen von Fall 4 vermieden werden. Das Netzwerk kann also beliebige Hierarchien aufnehmen, da es von den Modellierungsmöglichkeiten her umfangreicher ist.
- **Einsatzbereiche**
Die Verwendung von Netzwerken in Masseninformati­onssystemen ist noch nicht sehr stark etabliert. Neben einigen Vorreitern von Webseitenklassifikatoren [WAECHTER 2002, WebBrain] sind Netzwerke oft nur im allgemeineren Sinne im Einsatz: das Internet und seine Netzstruktur durch Hyperlinks, aber auch die Weiterentwicklung des Internets durch XML und insbesondere durch Topicmaps[BIEZUNSKI . 1999]. Abstrahiert man diesen Ansatz auf den Verzicht von hierarchischen Strukturen zu einer Anordnung basierend auf Zusammenhängen, so muß auch das WebSom-Projekt[KOHONEN . 2000] genannt werden, das versucht, diese Struktur durch eine Self-Organizing-Map (SOM) in zweidimensionalen Karten abzubilden.
- **Suchen und finden**
Die Stichwortsuche ist sogar erweitert möglich: die Suchergebnisse können mit Hilfe der Themengruppe inhaltlich gegliedert werden (siehe Kapitel 2.3.2: Abschnitt Suche). Durch die zusätzlichen Informationen in Form von Mehrfachzuordnungen zu Themengruppen und die inhaltliche Verknüpfung von Themen kann das Suchergebnis weiter aufgearbeitet werden (siehe [BITZER 2002]) und wird nicht nur in einer Liste mit einem Ranking

präsentiert wie das bei gängigen Suchmaschinen, z.B. Google[NIE 2002, BRIN . PAGE 1998], der Fall ist.

Stöbern ist durch ein Browsen durch das Themennetzwerk möglich und wird so optimal modelliert, da das Folgen von Verbindungen zu anderen Themengruppen dem Assoziieren sehr ähnlich ist.

- **Umgang mit Neuem**

Neue Themengruppen lassen sich besser hinzufügen, da sie beliebig an ein vorhandenes Gruppennetz angeschlossen werden können, ohne dass die vorhandene Struktur geändert werden muss. Es liegt nahe neue Bücher verschiedenen Themengruppen zuzuordnen: Dieses Verfahren ist einfacher und schneller, als genau eine Kategorie auswählen zu müssen. Wenn Themengruppen zu viele Bücher enthalten und diese nicht ausreichend voneinander differenziert werden können, dann ist es möglich analog zu dem evolutionären Vorgang der Spezialisierung des Wissens (siehe auch Ira Goldsteins Ansatz[GOLDSTEIN 1979]) vorzugehen und speziellere oder thematisch angrenzende Gruppen einzubinden, um so eine stärkere Differenzierung zu ermöglichen. Hätte man eine Themengruppe „Bäume“, die mit mehr und mehr Büchern über diverse Bäume gefüllt würde, wäre ein evolutionärer Schritt notwendig, indem neue speziellere Themengruppen wie z. B. „Laubbäume“ und „Nadelbäume“ eingeführt werden, wobei dann „Nadelbäume“ auch mit der Themengruppe „Weihnachten“ verknüpft sein kann, ohne dass die Gruppe „Bäume“ mit Weihnachten verknüpft sein muss.

1.5 Reduktion von Komplexität

Ein weiterer wichtiger Punkt bei Informationssystemen lässt sich nur schwer mit dem bisher gewählten Ansatz der Mengen und ihren Relationen zueinander erklären. Dieser Punkt hängt eng damit zusammen, dass Mengen mit sehr vielen Elementen für uns leicht unüberschaubar werden. Wir werden täglich mit einer großen Menge an Signalen, empfangen durch unsere Sinnesorgane, konfrontiert mit der wir umgehen müssen. Der größte Teil ist zum Zeitpunkt des Empfangs für uns irrelevant, nur ein kleiner Teil muss isoliert und interpretiert und erkannt werden, also mit dem in Zusammenhang gebracht werden, was uns schon bekannt ist. Dabei ist folgende Entscheidung sehr wichtig: Welche Informationen sind in der aktuellen Situation relevant und welche nicht. Die Forderung an diese Entscheidung ist, dass sie schnell und zuverlässig funktionieren muss, damit wir überleben können¹⁰(siehe z. B. [GAZZANIGA . 1998a]).

Dieses Problem findet sich auch bei der Arbeit mit Suchmaschinen wieder: Bei Eingabe eines Suchwortes in eine Suchmaschine erhalten wir normalerweise sehr viele Treffer. In der Liste der Suchergebnisse befindet sich eine kleine Teilmenge, die für uns von Interesse ist. Es ist problematisch diese Auswahl maschinell zu treffen. Ein solcher „Filter“ funktioniert wie eine Zensur: Es besteht die Gefahr dass wichtige Elemente nicht als solche erkannt werden und folglich auch ungenannt bleiben. Als Alternative bietet sich die Möglichkeit, im Sucher-

¹⁰Selbst wenn wir noch so beschäftigt sind und andere Informationen verarbeiten, so meldet uns unser Wahrnehmungssystem sofort wenn eine Gefahr besteht, beispielsweise wenn sich Rauch in der Luft befindet, obwohl es normalerweise gar nicht auffällt, da wir unbewusst ständig Gerüche wahrnehmen.

gebnis Elemente zu gruppieren und so hervorzuheben, anstatt sie zu filtern und somit einige zu eliminieren. Dies geschieht z. B. durch ein Clusterverfahren, das den Elementen der Liste eine Struktur, in Form von Gruppierungen, auferlegt.

Unser Gehirn schafft dies schnell und effizient dadurch, dass der Kontext (siehe z. B. [KOKINOV . YOVEVA 1996]) berücksichtigt wird: In welchem Zustand befindet sich das Gehirn? Welches Ziel wird in dem Moment verfolgt? Welche Erwartungen bestehen zur Zeit? Auf diese Weise werden von Anfang an nur die Cluster berücksichtigt, die sowieso durch den Kontext aktiv sind: Sehen wir nur einen Schwanz hinter der Tür der Nachbarin verschwinden, wissen wir sofort, dass es sich um eine (ihre) Katze handelt, ohne einen Großteil der notwendigen Eigenschaften des Konzepts *Katze* wahrzunehmen.

Bei der Suchmaschine ergeben sich also analog zwei Lösungsmöglichkeiten:

- Clusterung nach allen möglichen Kontexten (= Gliederung der Suchergebnisse, siehe z. B. Vivisimo[Vivisimo]) und/oder
- Berücksichtigung der für den Benutzer wichtigen Kontexte (z. B. durch Beachtung seiner Interessen[MÜLLER 2000] und die sich daraus ergebende Umbewertung der Relevanz der Suchergebnisse).

1.6 Agenda (A1 bis A6)

Die folgende Agenda listet die entscheidenden Punkte, Fragen und Forderungen auf, die von Informationssystemen angegangen und beantwortet werden müssen.

A1 Definition der Entitäten

Was ist in dem Modell das Element (Verweis) und was die Menge (Themen-gruppe)?

A2 Relationen

Was repräsentieren Element-Menge- und Menge-Menge-Zuordnungen?

In wie weit sollten diese Zuordnungen durch Verknüpfungsstärken gewichtet werden?

Macht es Sinn verschiedene Typen für die Relationen einzuführen (siehe Typisierung der Fälle im Kapitel 1.2 Relationen)?

A3 Kontext und Relevanz

Handelt es sich um Elemente, die ihre Zuordnung dynamisch erweitern oder verändern? Dann muss neben der semantischen Kontextrelevanz auch ein weiterer Faktor integriert werden, der die sich über die Zeit verändernde Kontextrelevanz abdeckt.

A4 Zugriff

Neben der verbreiteten Stichwortsuche muss auch eine thematische Suche (Hypertextansatz siehe auch [KOHONEN . 2000]) Zugriff auf das Verzeichnis gewährleisten, da es immer den Fall geben kann, dass sich erst während des Recherchierens herausstellt, wonach genau gesucht wird.

A5 Benutzeradaptivität

Als wünschenswert gelten zusätzlich Mechanismen, die sich den persönlichen Interessen des Benutzers anpassen können (OySTER[MÜLLER 2000] ist hier einer der Vorreiter) und gleichzeitig in der Lage sind, den gesamten Benutzerkreis, z. B. in Form von sozialen Gruppen[SIMMEL 1890, FROST 2001b], zu modellieren und bei dem Zustandekommen der Gewichte zu berücksichtigen.

2 NIO: Eine benutzeradaptive Informationsplattform

Die Plattform NIO ist aus dem Webverzeichnisprojekt *meiNetz*[FROST 2001b] erwachsen und stellt eine Plattform dar, mit der webbasiert verschiedenste Verzeichnisse modelliert und über ein Webportal zum Einsatz kommen können. NIO steht dabei für *Netzwerk zur Modellierung von Individualität und Orientierung in der Informationswelt* und stellt neben dem Webverzeichnis als ein Modul mit einer Spezifikation innerhalb von NIO auch ein Desktopmodul sowie viele weitere Module zur Verfügung. Während das Webverzeichnis ein und dieselbe Sicht auf die Linkdatenbank für alle Benutzer anzeigt, ist der Desktop jedes Benutzers anders und stellt dessen individuelle Interessen in den Vordergrund. Dieses Zusammenspiel entspricht dem, zwischen einer öffentlichen Bibliothek und dem eigenen Bücherregal zu Hause.

2.1 Grundgedanke

Benutzer-Information, Information-Benutzer

Die Entscheidung, beim Zugriff auf Weblinks nicht ausschließlich eine Stichwortsuchmaschine und die Bookmarkverwaltung im Browser zu nutzen, basiert auf folgender Überlegung:

Der Zugriff auf Informationen lässt sich in zwei Richtungen lesen:

1. Ein Benutzer interessiert sich für viele unterschiedliche Informationen,
2. eine spezielle Information ist für viele verschiedene Benutzer interessant.

Der zweite Punkt legt also nahe, Benutzer mit einem gemeinsamen Interesse zu gruppieren und die Informationen, die dann für diese Gruppe interessant sind, auch als solche anzuordnen. Dies geht nur durch eine Modellierung der Benutzerinteressen, z. B. innerhalb eines Webportals bei dem viele Benutzer zusammenkommen und für das System nicht anonym sind.

2.2 Aufbau von NIO

In dieser Arbeit möchte ich mich auf den Aufbau von NIO in Bezug auf die Modellierung eines Webverzeichnisses, realisiert durch die drei Module Desktop, Webverzeichnis und Suche, beschränken.

zu A1: Definition der Entitäten bei NIO

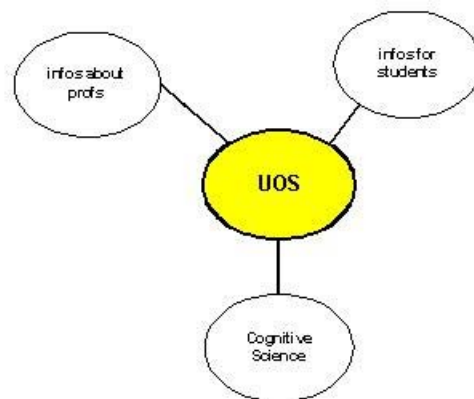
In dem Verzeichnis werden Links zu Internetseiten verwaltet. Jeder Link hat neben der Adresse (URL) einen Titel und eine Kurzbeschreibung. Was genau eine Gruppierung mehrerer Links ausmacht, ist nicht vorab festgelegt. In den folgenden Kapiteln wird das Konzept der NIO-Gruppe entwickelt.

2.2.1 Themennetzwerke

Wie in Kapitel 1 beschrieben, sollen Themennetzwerke zur Modellierung des Webverzeichnisses eingesetzt werden. Dabei entspricht ein Link, mit Titel und Beschreibung über den Inhalt der Seite, einem Verweis. Die Verweise können dann zu Themen gruppiert werden. So kann jeder Verweis beliebig vielen Themengruppen zugeordnet werden.

Ordnet man einen Teil der Themen rund um die Universität Osnabrück (UOS), so könnte sich ein solcher Graph ergeben.

Information model: UniOs



An dem großen Themenbereich der Universität Osnabrück grenzen beispielsweise folgende Themenbereichen an:

- Informationen rund um den Studiengang Cognitive Science,
- allgemeine Informationen über Lehrende an der Universität oder
- Informationen die potenziell für Studenten von Interesse sind.

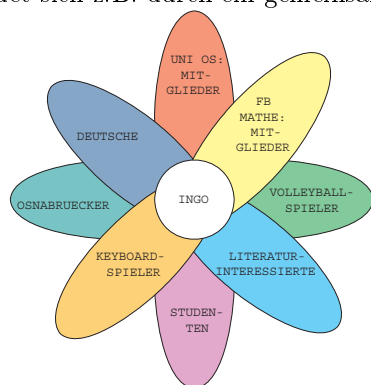
2.2.2 Soziale Netzwerke

Ein gutes Benutzermodell muss in der Lage sein, pro Benutzer zwei Fragen zu beantworten:

- Wie lassen sich die Interessen eines Benutzers beschreiben: Welche Eigenschaften müsste ein Infobroker von mir wissen, um eigenständig für mich arbeiten zu können?
- In wie weit haben zwei Benutzer ähnliche Interessen: Wie unterscheide ich mich von anderen?

Um diese Fragen über Individualität zu beantworten, sind eine Vielzahl von Modellen entwickelt worden. Für das meiNetz-Projekt wurde ein sehr einfacher aber

in diesem Zusammenhang zweckvoller Ansatz aus der Soziologie benutzt, der Individualität im Kontext zu anderen modelliert. Dazu wird auf ein Modell aus der Gruppensoziologie von Georg Simmel[SIMMEL 1890] zurückgegriffen. Demnach lässt sich Individualität durch die Zugehörigkeit zu sozialen Kreisen beschreiben¹¹. Ein sozialer Kreis wie er hier gemeint ist, ist *inhaltlich* gebunden und bildet sich z.B. durch ein gemeinsames Interesse.



So könnte z. B. eine Beschreibung der Interessen der Person *Ingo* wie in der Illustration dargestellt aussehen: Dadurch, dass Ingo in Osnabrück studiert, ist er zwangsläufig ein Osnabrücker, außerdem gehört er z.B. dem Kreis derjenigen an, die sich für Volleyball interessieren.

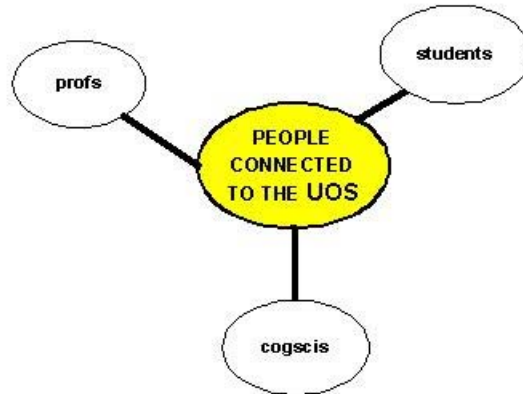
Individualität wird hier also durch die Mitgliedschaften zu sozialen Kreisen modelliert. Wenn sich die Individualität ausbildet oder verändert, ändern sich auch die Mitgliedschaften in sozialen Gruppen¹². Ein Informationssystem sollte ebenfalls nicht nur in der Lage sein einen Ist-Zustand abzubilden, sondern auch die Veränderung von Interessen modellieren können.

Wollte man eine ganze Gruppe von Menschen, also soziale Strukturen als soziale Kreise und die Relation untereinander, modellieren, so ergäbe sich folgendes vereinfachtes Bild der Universität Osnabrück.

¹¹In Simmels Worten: „Die Gruppen, zu denen der Einzelne gehört, bilden gleichsam ein Koordinatensystem, derart, daß jede neu hinzukommende ihn genauer und unzweideutiger bestimmt. Die Zugehörigkeit zu je einer derselben läßt der Individualität noch einen weiten Spielraum; aber je mehr es werden, desto unwahrscheinlicher ist es, daß noch andere Personen die gleiche Gruppenkombination aufweisen werden, daß diese vielen Kreise sich noch einmal in einem Punkte schneiden.“[SIMMEL 1890]

¹²Zu diesem Punkt ein Beispiel von Simmel: „Eins der einfachsten Beispiele ist das angeführte, daß der ursprüngliche Zusammenhang des Familienkreises dadurch modifiziert wird, daß die Individualität des Einzelnen diesen in anderweitige Kreise einreihet;“[SIMMEL 1890]

User model: People at the UniOs



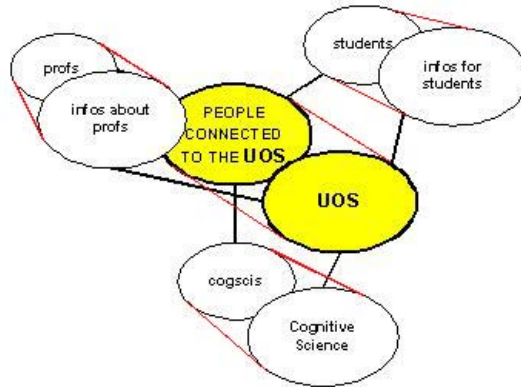
Sieht man die Universität als Organisation im soziologischen Sinne, lassen sich z.B. folgende Gruppen innerhalb dieser Organisation differenzieren: Lehrende, Studenten oder Personen die mit Cognitive Science zu tun haben.

2.2.3 NIO-Gruppen

Die Gruppen innerhalb von NIO versuchen diese beiden Modelle in sich zu vereinen: Wird eine soziale Gruppe aus einem gemeinsamen Interesse heraus definiert, so interessiert sich naturgemäß die Gruppe für dieses Thema. Und genau wie verschiedene Organisationsstrukturen in der Gesellschaft miteinander zusammenhängen, lassen sich auch Themen in Zusammenhang zueinander betrachten. Diese Vorgehensweise ergibt auch aus einem anderen Grund Sinn, denn oft ist es so, dass Mitglieder eines sozialen Kreises als Produzenten von Informationen anzusehen sind, dessen Konsumenten sich eben genau wieder in diesem sozialen Kreis befinden. So sammeln beispielsweise Kognitionswissenschaftsstudenten an der Universität Osnabrück auf der Fachschaftsseite Materialien rund um ihr Studium, die natürlich von besonderem Interesse für alle Studenten dieses Fachbereichs sind.

Die beiden Beispielgraphen, das Themennetzwerk und die sozialen Kreise, würden also zu folgender Struktur verschmelzen, in der eine soziale Gruppe mit je einem Thema vereint wird.

NIO: Uni Os

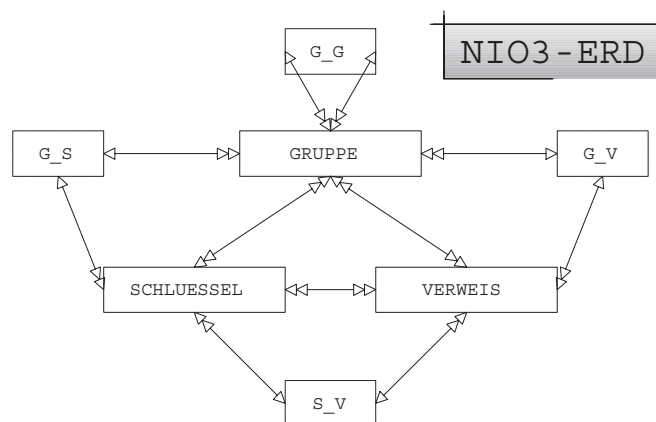


2.2.4 Datenstruktur

Die NIO Datenstruktur besteht aus drei Entitätstypen: Verweis, Gruppe und Schlüssel. Die Idee dahinter besteht darin, dass diese drei Typen gut geeignet sind, um jede Art von Verzeichnissen und Netzen darzustellen. Die Tabelle Verweis dient dabei für den eigentlichen Eintrag, der meist aus einer Adresse auf ein Objekt besteht. In einer Bibliothek wäre dies eine Karte pro Buch auf dem Titel, Autor und Adresse - also in diesem Falle der Standort des Buchs - verzeichnet sind.

zu A2: Relationen bei NIO

Im Falle des NIO-Webverzeichnisses handelt es sich um einen Hyperlink mit Titel und Beschreibung der Webseite. Mit Hilfe der Tabelle G_G (siehe Abb. NIO3-ERD) lassen sich ähnliche Verweise frei zu einer Gruppe zusammenfassen. Eine Gruppe hat einen Titel und eine Beschreibung und wird weiter dadurch definiert welchen anderen Gruppen, Verweisen und Schlüsseln sie zugeordnet ist. Die Schlüssel sind die semantisch kleinsten Einheiten, die Gruppen beschreiben und so einen Sinnzusammenhang bilden. Sie beschreiben auch Verweise durch entsprechende Zuordnungen.



So ergeben sich also die Tabellen `nio_verweis`, `nio_gruppe` und `nio_schluesssel`, sowie die dargestellten Verknüpfungstabellen um eine n:m Zuordnung zwischen allen Haupttypen zu ermöglichen. Diese Datenstruktur zeichnet sich als besonders flexibel aus und kann auch andere Daten aufnehmen als die hier beschriebenen. So ist beispielsweise das Benutzermodell durch eine spezielle Gruppe realisiert, die mit den Verweisen verknüpft wird, die der Benutzer verwendet und die Gruppenmitgliedschaft wird durch Relationen zu anderen Gruppen abgebildet.

Auffällig sind die fehlenden selbstreferentiellen Verknüpfungen zwischen Verweisen bzw. zwischen den Schlüsseln. Die Ähnlichkeit zwischen Verweisen soll nicht direkt zwischen den Verweisen selbst modelliert werden, sondern durch die Gruppenmitgliedschaften, die dann als ein Kontext dienen. So können zwei Verweise in einem Kontext ähnlich sein und zusammengehören und in einem anderen Kontext unterschiedlich sein¹³. Die gleiche Überlegung trifft auch auf die Schlüssel zu: Ihre Ähnlichkeit wird auch nur über die Gruppenverknüpfungen beschrieben.

2.3 Darstellung und Dynamik

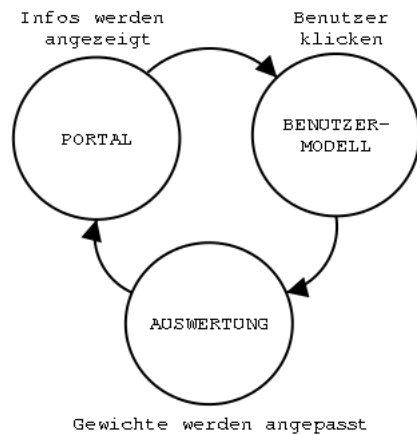
2.3.1 Umgang mit Gewichten

Der NIO-Kernel stellt verschiedene Arten von Gewichten bereit: knotenbezogene und verknüpfungsbezogene, die sich teilweise in Abhängigkeit des Benutzerverhaltens anpassen. Dieser Mechanismus und die einzelnen Veränderungen der Gewichte werden in diesem Kapitel behandelt.

zu A3: Kontext und Relevanz bei NIO

Die Relationen und die durch sie mögliche relative Gewichtung bilden so die Grundlage, um verschiedene Kontexte je mit ihrer Relevanz abzubilden. Dies wird durch einen Kreislauf erreicht:

¹³Angelehnt an die Synonymdefinition aus der Linguistik. Ein ähnliches Modell wird bei Wordnet durch die Synsets verwendet. Siehe dazu <http://www.cogsci.princeton.edu/~wn/>.



Die Daten werden an Hand ihrer Gewichtung in eine Reihenfolge gebracht und in dem PORTAL dargestellt. Durch die Klicks der Benutzer werden Events ausgelöst, die in das BENUTZERMODELL einfließen und schließlich durch die AUSWERTUNG neue Gewichte erhalten, was dann zu einer anderen Darstellung führen kann.

Im Zusammenhang der Gewichte werde ich zuerst auf die Auswertung eingehen, dann das Portal und seine Oberfläche erklären und mit dem Benutzermodell den Kreis schließen. Dabei beziehe ich mich auf die „Zähler“-Spalten der Tabellen des Datenmodells, die konzipiert sind, um die Benutzungshäufigkeit abzubilden (VERWEIS.ZAEHLER, GRUPPE.ZAEHLER), bzw. eine Gewichtung der Relationen zu ermöglichen (GRUPPE_VERWEIS.ZAEHLER, GRUPPE_GRUPPE.ZAEHLER).

1. Knotenbezogene Gewichte (absolut)

- WEBLINK: VERWEIS.ZAEHLER entspricht einem absoluten Ranking über alle Weblinks und ähnelt vom Prinzip her dem Google Pagerank [BRIN . PAGE 1998, NIE 2002]. Als ein eindeutiger Wert pro Weblink kann er die Benutzung dieses Links innerhalb von NIO, also über alle Gruppen und Benutzer hinweg abbilden.

Einsatz bei NIO: Dieses Feature wird zum aktuellen Stand der Implementation nicht genutzt, doch ein Ranking über alle Links wird dadurch erreicht, dass jeder Link sich in der Gruppe „Öffentlich“ befindet. Diese Art von Gruppe stammt noch aus der Zeit, in der innerhalb des Verknüpfungssystems nicht nur Themen sondern auch Zugriffsrechte modelliert wurden. Momentan entspricht also die Position in der Gruppe „Öffentlich“ dem absoluten Gewicht. Weiteres Einsatzgebiet könnte die Schlüsselwortsuche sein, die zwei Kriterien verwenden könnte: (1) wie gut passt die Suchphrase zu dem Link und (2) wie stark ist der Link absolut bewertet. Bei der Bewertung der Suchergebnisse würde man also zusätzlich davon ausgehen, dass die häufiger benutzten Links den seltener benutzten Links den Vorzug geben, wenn das erste Kriterium einen gleichen Wert liefert.

- THEMENGRUPPE: GRUPPE.ZAEHLER entspricht ebenfalls einem Ranking über alle Einheiten unabhängig von Verbindungen zu anderen. Es wird also die Benutzung abgebildet und so ein Ranking ermittelt, dass die Benutzungshäufigkeit der Themengruppen abbildet.

Einsatz bei NIO: Vor Beginn dieser Arbeit wurde dieses Gewicht benutzt,

um die verknüpften Gruppen der aktuell fokussierten abzubilden (siehe auch Kapitel 3.1 zur Problembeschreibung des existierenden Systems). Da die Schlüsselwortsuche auch passende Themengruppen findet, würde sich auch hier ein Vorgehen wie bei den Weblinkgewichten anbieten, doch momentan werden aus technischen Gründen die von den Schlüsseln ausgehenden Verknüpfungen nicht gewichtet.

2. Verknüpfungsbezogene Gewichte (relativ)

- THEMENGRUPPE - WEBLINK: GRUPPE_VERWEIS.ZAEHLER. Diese Verknüpfungsgewichte bilden die Benutzung der Weblinks innerhalb einer Themengruppe ab. Da normalerweise mehrere Links pro Gruppe vorhanden sind, interpretiert man die Verknüpfungsstärken innerhalb einer Gruppe miteinander. Es ergibt sich also eine Menge von Linkverknüpfungen pro Gruppe und somit auch ein Ranking pro Gruppe und nicht wie bei den knotenbezogenen Gewichten ein absolutes Gewicht pro Verweis unabhängig von den Gruppenmitgliedschaften.

Einsatz bei NIO: Dieses Gewicht wird eingesetzt, um Kontextrelevanz abzubilden. Eine Themengruppe steht für den Kontext und in diesem Kontext wird ein Link geklickt und erhält so sein Gewicht. Ein Link zum Institut für Kognitionswissenschaft in Osnabrück ist auf jeden Fall der Themengruppe „Uni-Osnabrück“ als auch der Themengruppe „Studiengang Cognitive Science“ zugeordnet. Im Kontext „Uni Osnabrück“ hat der Link eine geringere Relevanz (weil ein Institut von vielen) als im Kontext des Studiengangs, was durch verschiedene Gewichte zu den beiden Themengruppen modelliert wird.

- THEMENGRUPPE - THEMENGRUPPE: GRUPPE_GRUPPE.ZAEHLER. Mit dieser Verknüpfung wird das Themennetzwerk modelliert. Die Verknüpfungen sind dabei einseitig gewichtet: Eine „Subjektgruppe“ ist mit einer „Objektgruppe“ verknüpft (hier im grammatikalischen Sinne: a (Subjekt) ist mit b (Objekt) verbunden). Im Themennetzwerk existieren immer zwei solcher entgegengesetzter Verknüpfungen zwischen zwei Gruppen, so dass die verschiedenen Richtungen eine unterschiedliche Verknüpfungsstärke haben können. Hinter dieser Vorgehensweise steht die Überlegung, dass grundsätzlich Verknüpfungen einen semantischen Bezug haben, der fast immer gerichtet ist: a zu b steht in einem anderen Verhältnis als b zu a. Unterschiedliche Gewichte können eine solche Gerichtetheit ausdrücken.

Einsatz bei NIO: Durch die Auswertung der beiden Verknüpfungsgewichte kann Struktur emergent hervorgehen. Dieser Prozess wird ausführlich in dem 3. Kapitel vorgestellt.

3. Aufwertung eines Gewichts

In diesem Einsatzbereich des Netzwerkes lassen sich somit folgende Charakteristika beschreiben: Die Gruppen untereinander bilden ein flexibles Netzwerk, der Rang der Links pro Gruppe ist ebenfalls nicht festgelegt. Diese beiden Verknüpfungen zwischen Gruppen, und Gruppen und Links, sollen durch das Verhalten der Benutzer ständig adaptiv trainiert werden. Auf dieser Abstraktionsebene ist es analog zu einem künstlichen neuronalen Netz und es liegt sehr nahe, lokale Lernmechanismen, die bei künstlichen neuronalen Netzen benutzt werden, hier auch einzusetzen. Insbesondere bietet sich hier die hebbische Regel an, da ein Linkklick in einer Gruppe einer gleichzeitigen Aktivierung dieser Gruppe mit diesem Link

entspricht und dann das Gewicht zwischen diesen beiden Knoten verstärkt werden kann. Ähnlich kann man auch mit den Gruppen vorgehen und sich vorstellen, dass der Wechsel von Themengruppe A zur Themengruppe B einer zeitlich sehr nahe beieinander liegenden Aktivierung entspricht und so das Gewicht von A nach B stärkt. Wir ([FROST . 2003a]) gehen also davon aus, dass das Verhalten der Benutzer innerhalb dieses Netzwerkes nach sinnvollen Kriterien abläuft, die nach und nach in multisubjektiver Art von dem Netzwerk aufgenommen werden (siehe dazu auch Kapitel 3.3.2: Wie Klicks zu Gewichten werden).

Diese Analogie zwischen der NIO-Webverzeichnisdatenstruktur und neuronalen Netzen ist erst vor relativ kurzer Zeit in das Projekt aufgenommen worden. Entscheidender Vorteil gegenüber mehreren anderen Techniken, die vorher benutzt und getestet wurden, besteht in der Autonomie der Gewichte: Die Aufwertung wird nur an Hand des aktuellen Gewichtes¹⁴ vorgenommen. Dies ist sehr effizient und schnell gegenüber anderen Techniken, die zentral von einem Programm gesteuert wurden, das normalerweise über Nacht alle Logdaten ausgewertet und gegebenenfalls die Rangfolge der Links pro Gruppe verändert hat.

Ursprünglich wurde dieser Ansatz aus der Idee gewonnen, dass eine Gemeinschaft von Benutzern implizit ein *Socialclustering* und *Socialranking* über die Verweise und Gruppen durch die Benutzung des Systems macht. Diese Idee steht im Gegensatz zu Methoden, die versuchen, nur aus den Informationen selbst ein Ranking und Clustering zu erstellen und wurde bereits in anderen Feldern erfolgreich eingesetzt [SHARDANAND . MAES 1995] [BASU . 1998].

4. Abwertung eines Gewichtes

Als wichtigen Unterschied zu einem neuronalen Netz, das eine Trainingsphase hat und so in einen Zustand gerät, der die erkannten Regelmäßigkeiten der Eingangsmuster registriert hat, besteht darin, dass die Trainingsphase dauerhaft ist und sich inhaltlich verändert. Die Gewichte konvergieren also gegen die aktuell optimale Gewichtevertelung, erreichen diese jedoch nie, da auch sie sich verändert. Die Relevanz von Links innerhalb eines Kontextes verändert sich über die Zeit. Auch die Beziehungen unter den Themengruppen sind von dynamischer Natur und ändern sich ständig. Das Netz muss also in der Lage sein „zu vergessen“. Alle Gewichte sollten also den Drang haben sich abzuwerten, wenn sie nicht mehr benutzt werden (siehe dazu auch Kapitel 3.3.2: Wie Klicks zu Gewichten werden). Dabei ergibt sich ein Problem: Die Benutzung der Gruppen ist sehr unterschiedlich stark. Somit sind die höchsten Gewichte innerhalb der Gruppen unterschiedlich und ein einheitlicher Abwertungsmechanismus führt dazu, dass die Gewichte in selten benutzten Gruppen schnell gegen Null gehen. Deshalb wurde während der Erstellung dieser Arbeit eine Abwertungsfunktion der Gewichte, abhängig von der Nutzungsintensität jeder Gruppe, eingeführt und wird zur Zeit erprobt.

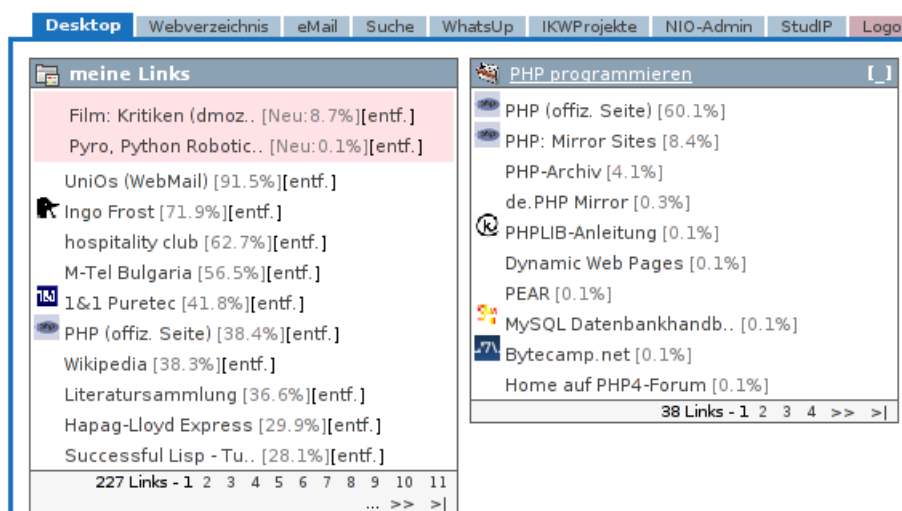
¹⁴Mit Berücksichtigung der Refraktärphase (vgl. Kapitel 3.3.2: Wie Klicks zu Gewichten werden).

2.3.2 Portaloberfläche

Innerhalb der drei Module Desktop, Webverzeichnis und Suche können verschiedene Events ausgelöst werden, die dann das Benutzermodell anpassen und Gewichte verändern.

1. Desktop:

Der Desktop ermöglicht einen schnellen, dem Benutzer angepassten, Zugriff auf wichtige Weblinks und Themengruppen. Dies wird durch die Startgruppe und durch aus dem Webverzeichnis importierte Themengruppen erreicht. In der Startgruppe, die in dem Moment den Status „fokussiert“ hat, befinden sich alle Links, die der Benutzer aus dem Webverzeichnis bisher geklickt hat. Ein Klick hier löst ein KLICK-LINK-STARTGRUPPE-Event aus.



Die für den Benutzer wichtigsten Themengruppen werden in konzentrierter Form¹⁵ angezeigt und ermöglichen den direkten Zugriff auf die wichtigsten Links der Gruppen: Ein Klick auf einen Link in solchen Gruppen löst ein KLICK-LINK-THEMENGGRUPPE-Event aus. Es ist auch möglich durch Klicken auf den Namen der Themengruppe oder auf die verkürzt dargestellten ikonisierten¹⁶ Themengruppen in das Webverzeichnis zu wechseln und dort die Gruppe zu öffnen. Dann wird im Webverzeichnis ein entsprechendes GRUPPE-FOKUSSIERT-Event ausgelöst.

2. Webverzeichnis als Gruppennetzwerk

Nach der Anwahl des Webverzeichnisses muss zunächst eine Gruppe fokussiert werden, um das Browsen durch das Gruppennetz zu starten. Dazu stehen zwei Listen von Einstiegsgruppen zur Verfügung: Die Themengruppen, die spezifisch für den Benutzer sind und auch auf dem Desktop angezeigt werden, sowie eine Liste mit allgemeinen Themengruppen.

¹⁵ Dies geschieht, indem Beschreibung und URL versteckt werden und nur der Titel angezeigt wird.

¹⁶ Ist nicht in der Abbildung dargestellt.

Ein Klick auf eine Gruppe führt dazu, dass diese Gruppe fokussiert wird und löst ein GRUPPE-FOKUSSIERT-Event aus (hier Gruppe „Bildung / Beruf“). Ab diesem Punkt lassen sich Links (hier z.B. jobs.de) aus der fokussierten Gruppe klicken. Dadurch kommt ein KLICK-LINK-THEMENGRUPPE-Event zustande. In der linken Spalte sind alle verknüpften Themengruppen (als stärkste hier „Universitäten“ oder „Universität Osnabrück“) angezeigt, in die man wechseln kann und sie so fokussiert. So wird ein Browsen durch das Gruppennetz ermöglicht.

3. Suche

Das Suchmodul[BITZER 2002] ermöglicht eine Stichwortsuche über alle Links und Gruppen und berücksichtigt dabei die Schlüsselwörter, die mit Gruppen und Links verknüpft sind. Die Eingabe eines Suchbegriffs (hier

beispielsweise „buecher“) führt zur Auflistung aller gefundenen Links in der rechten Spalte, in der linken werden alle Gruppen dieser Links angezeigt. Am Anfang ist keine dieser Gruppen aktiviert.

The screenshot shows the NIO Suche web interface. At the top, there are navigation tabs: Desktop, Webverzeichnis, eMail, **Suche**, WhatsUp, IKWProjekte, and NIO-Ad. The main header reads "NIO Suche -- Suchmaschinen, W". Below the header, there is a search bar with the text "Anfrage: buecher". On the left side, under "SUCHERGEBNIS AUS:", there is a list of categories: "Gesamtes Webverzeichnis" (highlighted), Bücher, Literaturrecherche, Einkauf im Web, Literatur, Medien, Computer, HTML, Informatik, Zeitschriften/Zeitungen, and PHP programmieren. On the right side, there is a list of search results, each with a checkbox and a link:

- [Computer-Literatur online lesen](#) - [http://](#) Kostenlose Fachbuecher zu den Themen
- [Buecher.de - wir begeistern mit Medien](#) - Bücher. Riesenauswahl. Bis zu 75% Preis! [LinkEdit]
- [Abebooks.de](#) - [http://www.abebooks.de](#) - [LinkEdit] Marktplatz für antiquarische, vergriffene
- [ZVAB](#) - [http://www.zvab.de/](#) [60.2%] - Zentrales Verzeichnis Antiquarischer Bü [LinkEdit]
- [Missing Link Versandbuchhandlung](#) - [htt](#) Online Buchhandel, spezialisiert auf die L Lieferzeiten!!!) [LinkEdit]

Ein Klick auf eine solche Gruppe (hier „Einkaufen im Web“) zeigt eine Teilmenge der Suchergebnisse an; nämlich genau die Links, die zu dem Suchbegriff passen und sich in der gewählten Gruppe befinden. Auch hier führt ein Klick auf einen Link zu einem KLICK-LINK-THEMENGROUPE-Event.

The screenshot shows the NIO Suche web interface with the "Einkauf im Web" category selected in the left sidebar. The search bar still contains "Anfrage: buecher". The search results on the right are now filtered to show only links within the selected category:

- [Buecher.de - wir begeistern mit Medien](#) - Bücher. Riesenauswahl. Bis zu 75% Preis! [LinkEdit]
- [Abebooks.de](#) - [http://www.abebooks.de](#) - [LinkEdit] Marktplatz für antiquarische, vergriffene
- [Missing Link Versandbuchhandlung](#) - [http](#) Online Buchhandel, spezialisiert auf die B Lieferzeiten!!!) [LinkEdit]
- [Amazon.de](#) - [http://www.amazon.de/..](#) [Onlioneversandhaus fuer Buecher, Musik
- [Buch.de](#) - [http://www.buch.de/](#) [32%] - Internet-Shop für Bücher, CDs, Software,

zu A4: Zugriff auf Links bei NIO

Durch das Webverzeichnis wird also ein thematischer Zugriff, und durch das Suchmodul ein Zugriff mit Suchbegriffen im Sinne der Agenda ermöglicht.

2.3.3 Benutzermodell

Der Begriff Benutzermodell ist nicht eindeutig definiert, denn das Wort Benutzer ist singular wie auch plural. Genauso ist das Benutzermodell auch innerhalb des NIO-Modells zu verstehen, denn es modelliert den einzelnen Benutzer, als auch Benutzergruppen und das Verhalten aller Benutzer. So kommen folgende drei Abstraktionsebenen zustande[FROST 2001a].

Die Mikroebene: Der einzelne Benutzer Die Hauptidee für das Benutzermodell stammt aus dem sozialen Netzwerk (siehe oben): Zu welchen sozialen Gruppen gehört der Benutzer / für welche Themen interessiert er sich. Dies wird durch Zuordnungen zu NIO-Gruppen hauptsächlich aus dem Webverzeichnis modelliert. Die Stärke der Benutzung dieser Gruppen spiegelt die Wichtigkeit der Gruppen für den Benutzer wider. Neben den Gruppenzuordnungen gibt es noch die Startgruppe, in der von dem Benutzer geklickte Links gesammelt und danach angeordnet werden.

Die Mesoebene: Die soziale Gruppe Da eine Gruppe von verschiedenen Benutzern gemeinsam benutzt wird, ist also die Themengruppe als solche ein Abbild von kollektivem Verhalten und Interesse.

So würde z. B. eine Gruppe „Parteien“, die Links zu den Webseiten der einzelnen Parteien enthielte, sich wahrscheinlich so anordnen, dass die großen Volksparteien auch in der Gruppe auf den oberen Rängen wäre. Das hat einen weiteren positiven Nebeneffekt: Wenn Benutzer eine Gruppe das erste Mal nutzen, erhalten sie einen Überblick, wie die Seiten im Vergleich zueinander in diesem Gruppenkontext benutzt werden. Diese Information ist wichtig, aber kann nur von einem solchen System und nicht aus dem Internet an sich gewonnen werden. Beispielsweise könnte sich ein Ausländer, dem möglicherweise die deutschen Parteien noch gar nicht bekannt sind, durch Ansicht dieser Gruppe ein relativ repräsentatives Bild machen¹⁷.

Die Makroebene: Gemeinsamkeit der sozialen Gruppen Nun existiert noch eine abstraktere Adaptionsebene aufgespannt durch die Relationen zwischen den Gruppen, die auf das Verhalten aller Benutzer zurückgeht. Das Netz aus diesen Gruppen ist dann wie eine allgemeine Vorstellung darüber anzusehen, wie Themen zusammengehören. Während innerhalb einer Gruppe hauptsächlich „Spezialisten“ (=Gruppenmitglieder) über den Rang der Links implizit entscheiden, wird das Gruppennetzwerk von allen benutzt und spiegelt ein gemeinsames allgemeines Wissen wider, so dass - im Idealfall - sich jeder dort zurechtfindet.

¹⁷Voraussetzung dabei ist, dass die Benutzer des Portals politisch repräsentativ sind und das Internet auf ähnliche Weise nutzen.

zu A5: Benutzeradaptivität bei NIO

Das Benutzermodell wird durch folgende EVENTS verändert und erweitert, was sich durch die Manipulation der Verknüpfungen äußert (siehe oben). Folgende Events werden durch das Benutzerverhalten auf dem Portal ausgelöst und bearbeitet:

- **KLICK-LINK-STARTGRUPPE:**
Die Gewichte von dem geklickten Link werden in der Startgruppe aufgewertet.
- **KLICK-LINK-THEMENGRUPPE:**
Die Gewicht in der Gruppe „Öffentlich“ werden aufgewertet. Falls gerade eine Gruppe fokussiert ist, wird auch dort das Gewicht aufgewertet. Es wird überprüft, ob sich der Link in der Startgruppe befindet, falls ja, wird sein Gewicht aufgewertet, sonst wird eine neue Verknüpfung angelegt.
- **GRUPPE-FOKUSSIERT:**
Es wird überprüft, ob zuvor eine Gruppe fokussiert wurde, über die die aktuelle Gruppe erreicht wurde. Falls ja, dann wird die Verbindung einseitig gestärkt.

3 Erweiterungsansätze zur Typisierung von Relationen

Die Anforderung an eine Lösung besteht darin, einerseits Themenrelationen mit Relationstypen (nach Kap 1.2 Relationen) aus dem Verhalten der Benutzer emergent hervorgehen zu lassen und andererseits diese Struktur benutzerfreundlich und kognitiv adäquat darzustellen. Das ist besonders wichtig, da diese wieder auf das Benutzerverhalten zurückwirkt. So bildet sich durch das benutzeradaptive Verhalten des NIO-Webverzeichnisses eine Rückkopplung, da die neu ermittelten Verknüpfungstypen und Gewichte sofort in die Darstellung der Daten einfließen.

Dieses Kapitel gliedert sich in folgende Teile: Zuerst beschreibe ich den vorherigen Stand der Implementation im Umgang mit Gruppengewichten innerhalb des NIO-Webverzeichnisses, dann stelle ich mögliche Lösungsansätze einander gegenüber, greife einen heraus und stelle ihn skizzenhaft vor, gehe in den folgenden Absätzen genauer auf die einzelnen Schritte ein und schließe mit Ansätzen und Ideen zur Visualisierung ab. Im Ausblick stelle ich dann Möglichkeiten zur Evaluation vor.

3.1 Problembeschreibung des existierenden Systems

Vor Einführung des hier vorgestellten Ansatzes wurden die Themengruppen unabhängig von ihren Verknüpfungen zu anderen Gruppen bewertet und sortiert. Sobald eine Gruppe fokussiert wurde, das heisst ihre Links angezeigt wurden, erhielt diese eine einfache Aufwertung des ihr zugeordneten absoluten Zählers. Wurde zusätzlich ein Link aus dieser Gruppe geklickt, wurde ihr Zähler erneut aufgewertet. Der Suchvorgang durch Browsen war also in zwei Phasen gegliedert: Erstens das Gruppenbrowsen und sobald eine Gruppe gefunden wurde, zweitens die Erhöhung des Zählers der Gruppe-Linkzuordnung. An Hand der ersten Phase wurden die verknüpften Gruppen der aktuell fokussierten Gruppe in Form einer einfachen Liste angeordnet. Da manche Gruppen mit sehr vielen anderen verknüpft sind, wurden also in manchen Fällen nur die Gruppen mit den stärksten Rängen angezeigt.

Nach einiger Zeit stellten sich mehrere Probleme mit dieser Vorgehensweise heraus, die sich mit der vernachlässigten Kontextrelevanz zusammenfassen lassen: Da jede Gruppe nur für sich, und nicht in Bezug auf Vorgänger und Nachfolger im Suchpfad bewertet wurde, stellte sich unter den Gruppen eine Art Rangfolge der am stärksten benutzten Gruppen heraus. Im Fall eines Suchpfades spielen jedoch andere Gruppen eine wichtigere Rolle, als einfach nur die am meisten benutzten. So ist z. B. jede Gruppe mit der Gruppe „Öffentlich“ verknüpft und da dies die meist benutzte Gruppe (da zu dem Implementationszeitpunkt Einstiegspunkt ins Netzwerk) ist, steht diese an erster Stelle in der Liste der verknüpften Gruppen, spielt jedoch in den seltensten Fällen eine wichtige Rolle, und wird fast nie von dort aus angeklickt.

Es ergeben sich also durch die mangelnde Berücksichtigung des Kontexts folgende Probleme:

- Überbewertung der häufig benutzten Gruppen,
- Unterbewertung der speziellen Gruppen,

- Irritation durch Rangfolge beim Suchen,
- Irritation durch ungeordnete Gruppenliste.

Agendaabgleich Mit Hilfe der Agenda lassen sich nun diese Probleme erklären, da hier NIO nicht den Anforderungen der Agenda gerecht wird. Insbesondere die Punkte A2 und A3 der Agenda sind hier ausschlaggebend:

- zu A2, Relationen: Zu viele Verknüpfungen, die pro Gruppe angezeigt werden und ungenügend geordnet und strukturiert sind, verwirren die Benutzer und führen dazu, dass das Netzwerk nicht einfach zu benutzen ist. Der Suchpfad, der normalerweise zu den konkreteren Themen führt und einen Abstieg vom Allgemeinen ins Spezielle darstellt, ist nur schwer bei so vielen Abzweigungen pro Gruppe zu finden. Es zeigt sich also die Notwendigkeit Relationstypen einzuführen.
- zu A3, Kontext und Relevanz: Eine andere Form der Ordnung besteht in der Sortierreihenfolge der verknüpften Gruppen. Hier müssen unbedingt relative Gewichte benutzt werden, um die Gruppe-Gruppe-Verknüpfungen kontextadäquat zu sortieren. Auch dieser Anforderung kam NIO bisher nicht nach.

3.2 Lösungsvorschläge

Generell ergeben sich innerhalb der NIO-Architektur verschiedene Lösungsansätze:

- Ansatz nach Benutzung der Gruppe-Gruppe-Verknüpfungen (direkter Ansatz),
- indirekter Ansatz auf Basis der Gruppe-Link-Verknüpfungen und
- indirekter Ansatz auf Basis der Gruppe-Schlüsselwort-Verknüpfungen.

Der erste Ansatz arbeitet mit einem Algorithmus, der ausschließlich auf den Verknüpfungen zwischen den Gruppen operiert. Die Gewichte dieser Verknüpfungen ändern sich ständig beim Browsen der Benutzer durch das Webverzeichnis. Dieser Ansatz wird hier näher verfolgt.

Weitere Verknüpfungen werden durch die indirekten Ansätze einbezogen. Das Verhältnis zwischen zwei Gruppen würde dann z.B. über die verknüpften Links der beiden Gruppen und dem Verhältnis dieser Links zueinander erschlossen. Hier wären folgende Regeln denkbar:

- Ist Gruppe A mit mehr Links verknüpft als Gruppe B, so könnte Gruppe A allgemeiner - also umfassender - sein.
- Bilden die Links der einen Gruppe mehr oder weniger eine Teilmenge der anderen Gruppe, so könnte die andere Gruppe allgemeiner sein.

Diese Regeln setzen voraus, dass Links nicht nur den speziellsten Gruppen, sondern auch den allgemeineren zugeordnet sind. Das ist leider nicht immer der Fall. Wenn ein Link dennoch mit abstrakteren Gruppen verknüpft ist, so ist diese

Verknüpfung meist sehr schwach und von der Dynamik der Gewichtsänderungen durch das Event KLICK-LINK-THEMENGRUPPE ausgeschlossen.

Die Aussagekraft der Regeln nimmt sehr stark ab, wenn die Schnittmengen gemeinsamer Verweise oder Schlüssel sehr klein sind. Deshalb kommt auch der andere indirekte Ansatz auf Basis der Gruppe-Schlüsselwort-Verknüpfungen nicht infrage. Dieses Problem könnte man jedoch angehen, indem man die Links pro Gruppe, bzw. Schlüsselwörter pro Gruppe expandiert und so noch eine Stufe indirekter wird: Man würde dann z.B. nicht nur die Schlüsselwörter einer Gruppe, sondern auch die Schlüsselwörter der Links dieser Gruppe verwenden. Im ersten Augenblick erscheint dieser Ansatz sehr verlockend, doch muss man sich der Qualität und Art der Verknüpfungen, die man dann zusätzlich benutzt, bewusst werden. Es ergeben sich folgende Faustregeln die dann keine Beachtung fänden. Deshalb beschränke ich mich hier auf den direkten Ansatz.

- **Direkt vor indirekt:** Zieht man Verknüpfungen zwischen anderen Entitäten hinzu, muss man sich bewusst sein, dass diese einer anderen Logik folgen und eine andere Bedeutung haben. Aus ihnen kann nie vollständig eine andere Verknüpfung hervorgehen, allenfalls können Hinweise erschlossen werden, die durch steigende Komplexität (durch Expansion) schwächer werden je indirekter die Verknüpfungen werden.
- **Verknüpfungsdynamik benutzen:** NIO verfügt momentan durch die Benutzeradaptivität über Verknüpfungen, die sich ständig anpassen und so die Dynamik auffangen. Diese Dynamik sollte ausgenutzt werden und bezieht sich hauptsächlich auf benutzte Links innerhalb von Gruppen und Gruppe-Gruppe-Verknüpfungen.
- **Statische Verknüpfungen bergen Fehlerpotenzial:** Alle anderen Verknüpfungen sind statisch, also möglicherweise fehlerhaft oder veraltet. Sie bilden keine gute Grundlage. Selbst die dynamischen Verknüpfungen der Gruppe-Link-Verknüpfungen verändern nur Gewichte. Sie umfassen zur Zeit noch nicht die Option, dass sich alte überhaupt nicht benutzte Verknüpfungen löschen, oder dass sich manche Links automatisch in neuen Gruppen *vorstellen* und so neue Verknüpfungen zustande kommen können.
- **Schlüsselwortproblem:** Die Qualität der ausgelesenen Schlüsselwörter ist noch relativ gering im aktuellen Datenbestand. Neue Schlüsselwörter reorganisieren sich nicht, obwohl dies eigentlich notwendig wäre, um die Konsistenz zu gewährleisten. Das hängt damit zusammen, dass nach und nach neue Schlüsselwörter durch neue Verweise ins System hinzukommen, durch die aber auch die bereits vorhandenen Webseiten und Gruppen näher beschrieben werden könnten, indem die neuen Schlüsselwörtern mit dem vorhandenen Datenbestand verknüpft werden.

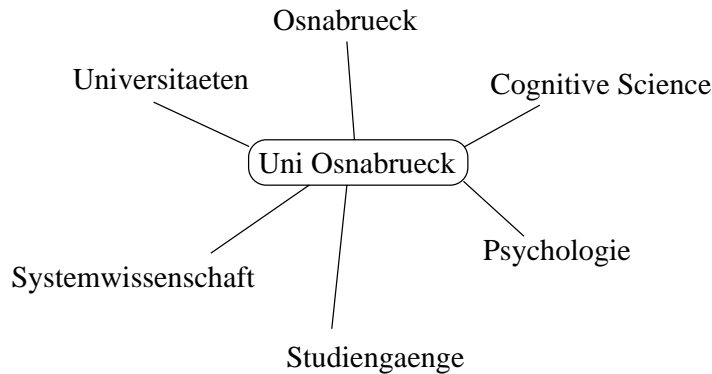
3.3 Ansatz nach Benutzung der Verknüpfungen

3.3.1 Programmskizze

Um einen ersten Überblick über den Algorithmus zu schaffen, stelle ich hier an einem einfachen Fall seine Arbeitsweise vor. Ausgangspunkt sei die Themengruppe „Uni Osnabrueck“, die mit den darum angeordneten Gruppen etwas zu

tun hat (im Sinne der Relationsbeziehung Fall 4 siehe Kap. 1.2). Mehr Aussagen können an dieser Stelle zu den Verknüpfungen und Beziehungen nicht gemacht werden, da die Verknüpfungen eine schwache Affinität zwischen den Gruppen modellieren.

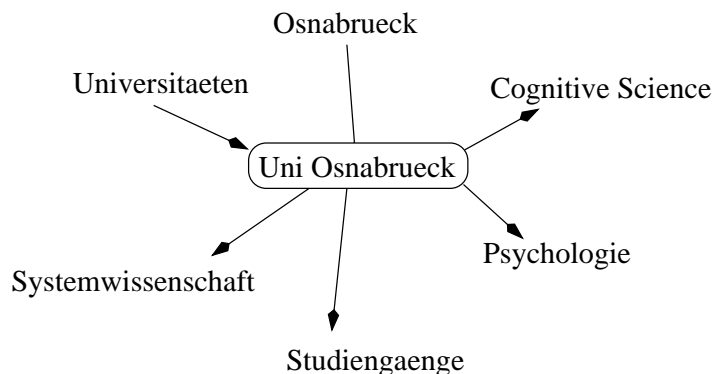
1. **Ausgangssituation:** Gruppe „Uni Osnabrueck“ mit assoziativ verknüpften Gruppen



Beobachtet man nun das Verhalten der Benutzer in dem Internetportal, stellt man z.B. folgenden Suchpfad zur Seite des Studiengangs Kognitionswissenschaft (befindet sich unter den ersten 10 Links in Gruppe Cognitive Science) fest: „Universitaeten“ - „Uni Osnabrueck“ - „Cognitive Science“ - Linkklick. Anhand dieses Suchpfades werden die Gewichte berechnet. Dabei wird berücksichtigt, in welche Richtung der Assoziationspfad benutzt wurde.

Das führt dazu, dass die verknüpften Gruppen nach Verknüpfungsstärke angeordnet werden können und in der Liste in der sie angezeigt werden, einen Rang erhalten.

2. **Einführung von Verknüpfungstypen** durch Vergleich der Verknüpfungsstärken

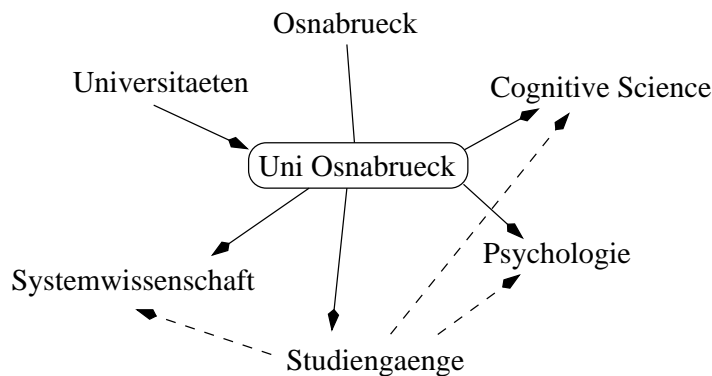


Werden die festgestellten Gewichte pro Richtung zwischen zwei Gruppen verglichen, stellt sich heraus, dass diese oft sehr unterschiedlich sind. So wird die Verknüpfung „Uni Osnabrueck“ - „Cognitive Science“ häufiger

benutzt, als die Verknüpfung in der Rückrichtung (hier illustriert durch einen Pfeil). Das hat damit zu tun, dass bei Suchvorgängen normalerweise, die Suche eingegrenzt wird, also von allgemeinen zu speziellen Gruppen geklickt wird, denn Einstiegsgruppen sind immer von sehr allgemeiner Natur.

Nachdem das Netzwerk ausgiebig benutzt wurde, ergeben sich also drei Verknüpfungstypen aus Sicht einer fokussierten Gruppe: zu einer allgemeineren (hier z.B. „Universitaeten“, da die Universität Osnabrück eine von vielen Universitäten ist), zu einer spezielleren (hier z.B. „Psychologie“ als ein Studiengang an der Universität Osnabrück) sowie zu einer assoziativen Gruppe, bei der die Gewichte cirka gleich stark sind (hier: „Osnabrueck“).

3. Analyse der Teilgruppen führt zur Gliederung



Innerhalb der spezielleren Gruppen (hier 3 Studiengänge sowie die Gruppe „Studiengaenge“) kann nun festgestellt werden, ob gemeinsame Obergruppen existieren. In diesem Fall ist dies die Gruppe „Studiengaenge,“ unter der die anderen zusammengefasst werden, so dass eine gegliederte Ansicht entsteht, die eine Ordnung in die Untergruppenliste hervorbringt.

Meist ist diese Suche nach gemeinsamen Obergruppen nicht so einfach, deshalb müssen neben den unmittelbaren Nachbargruppen wie hier zusätzlich die Nachbarsnachbargruppen der Analyse hinzugezogen werden (das wäre hier der Fall, wenn „Studiengaenge“ nicht direkt mit „Uni-Osnabrueck“ verknüpft wäre).

3.3.2 Schritt I: Wie Klicks zu Gewichten werden

Die Algorithmen rund um die Aktualisierung der Gewichte nach Daten durch die KLIICK-EVENTS ist eine der großen Herausforderungen innerhalb des NIO/meiNetz-Projektes und hat nach vielen Änderungen eine zufriedenstellende Lösung hervorgebracht, die folgende Charakteristika aufweist:

- die *Aufwertung* eines Gewichtes geschieht *in Echtzeit*, also sofort nach dem Linkklick,
- eine *Abwertung* findet nur wenn nötig statt; dies wird *periodisch* geprüft,
- eine *Sensibilität* für neue *seltene Ereignisse* ist durch schnelle Aufwertung,

mit einer langsamen Abwertung der Gewichte seltener Ereignissen gekoppelt,

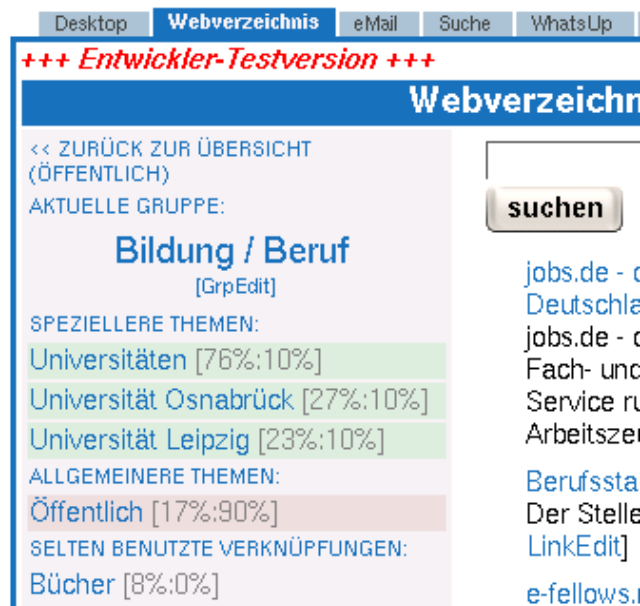
- es findet ein *geringer Zuwachs* bei sehr *häufigen Ereignissen* gekoppelt mit einer schnellen Abwertung ihrer Gewichte statt.

Außerdem ergibt sich folgendes Problem: Die Annahme, dass die Benutzung eines Links innerhalb einer Gruppe der Benutzung dieser Webseite entspräche oder der Fokuswechsel eine Aussage über die Relation zweier Themen mache, ist nur begrenzt richtig, denn es kann schnell passieren, dass die „Neu laden“ oder „Zurück“ Funktion des Browsers genutzt wird oder aus Versehen ein Fenster geschlossen wird und so die vorherige Ereigniskette wiederholt wird, was zu einer Überbewertung der Ereignisse führen würde. Dieses Problem kann auch in neuronalen Netzen auftreten. Dort wird es durch die Einführung einer Refraktärphase, die ein erneutes, sofortiges, ähnlich starkes Feuern eines Neurons vermeidet, gelöst. Überträgt man diesen Ansatz, so vermeidet man analog eine Aufwertung, wenn kurz zuvor bereits eine stattgefunden hat (zum Prinzip der Refraktärphase siehe u.a. [GAZZANIGA . 1998b], zur Übertragung ins NIO/meiNetz-Projekt siehe [FROST . 2003a]).

3.3.3 Schritt II: Vergleich der beiden Gewichte pro Richtung

Der Programmteil, der beide Gewichte vergleicht, berücksichtigt und löst folgende Probleme:

- Beide Gewichte müssen einen *Schwellenwert* überschreiten, so dass genug Daten einfließen können;
- Der *Vergleich* der beiden Werte muss *relativ* und nicht absolut geschehen;
- Anhand der Verhältnisse der Gewichte zueinander kann eine *Typisierung* der Verknüpfung stattfinden:
 - „<“ Untergruppe,
 - „>“ Obergruppe,
 - „~“ Assoziativgruppe und
 - „u“ unbenutzte Verknüpfung.



Zusammen mit dem neuen Algorithmus ergibt sich also folgendes Bild beim fokussieren der Gruppe „Bildung / Beruf“ (vergleiche alte Version in Kap. 3.2.3, Abschnitt Webverzeichnis als Gruppennetzwerk): Die Verknüpfungen zu den Gruppen „Universitäten“, „Universität Osnabrück“ und „Universität Leipzig“ zeichnen sich dadurch aus, dass sie wesentlich häufiger von „Bildung / Beruf“ aus angewählt werden als umgekehrt. Dies lässt sich an den Werten in den eckigen Klammern hinter den Gruppentiteln ablesen, die jeweils die Gewichte der Hin- und Rückrichtung angeben. Umgekehrt ist das Verhältnis zu der allgemeineren Gruppe „Öffentlich“ von der aus „Bildung / Beruf“ - sozusagen als Einstiegspunkt - angewählt wurde.

3.3.4 Schritt III: Gliederungsanalyse der Teilgruppen

In diesem Schritt sollen die Teilgruppen - insbesondere die spezielleren Themengruppen strukturiert werden. Dabei ergeben sich folgende Aufgaben für den Algorithmus:

- *Extraktion* von einem *Teilnetz* (Nachbarn und Nachbarsnachbarn der fokussierten Gruppe werden extrahiert);
- im Teilnetz werden alle *Verknüpfungen typisiert* (siehe Schritt II) und dann werden
- pro Teilgruppe gemeinsame *Ober- bzw. Untergruppen* gefunden: dazu kommt ein
- „*Spread of Activation*“-Ansatz im Teilnetz zur Anwendung.

suggesting topic structure(s):

Universitäten

[Universität Osnabrück | Universität Leipzig]

The screenshot shows a web directory interface. At the top, there are navigation tabs: Desktop, Webverzeichnis (selected), eMail, Suche, and WhatsUp. Below the tabs, there is a red banner that says '+++ Entwickler-Testversion +++'. The main heading is 'Webverzeichnis'. On the left side, there is a navigation menu with links: '<< ZURÜCK ZUR ÜBERSICHT (ÖFFENTLICH)', 'AKTUELLE GRUPPE:', 'Bildung / Beruf [GrpEdit]', 'SPEZIELLERE THEMEN:', 'Universitäten [76%:10%]', 'Universität Osnabrück [27%:10%]', 'Universität Leipzig [23%:10%]', 'ALLGEMEINERE THEMEN:', 'Öffentlich [17%:90%]', and 'SELTEN BENUTZTE VERKNÜPFUNGEN:', 'Bücher [8%:0%]'. On the right side, there is a search box with a 'suchen' button and a list of search results: 'jobs.de - de', 'Deutschlan', 'jobs.de - die', 'Fach- und F', 'Service run', 'Arbeitszeug', 'Berufsstart', 'Der Stellen', 'LinkEdit]', and 'e-fellows.ne'.

Im Falle der Gruppe „Bildung / Beruf“ berechnet der Algorithmus, dass „Universität Osnabrück“ und „Universität Leipzig“ nicht nur spezieller als „Bildung / Beruf“ sind, sondern auch spezieller als „Universitäten“ und ordnet diese beiden Gruppen „Universitäten“ unter. Auf diese Weise wird die Kopfzeile „suggesting topic structure(s)“ (siehe Abbildung) erzeugt.

3.3.5 Implementation: subnet.class.php

Die Realisierung dieser drei Schritte findet durch die Klasse `subnet.class.php`¹⁸ statt. Im Webverzeichnis wird die Klasse `Subnet` (als `$o_onesubnet`) instanziiert und einige Methoden (`compute_relation_types()`, `mark_unused()`, `find_structure()` und `get_first_neighbours()`) aufgerufen, die schrittweise die Daten holen, bearbeiten und schließlich nach Relationstypen gegliedert ausgeben. Hier stelle ich jeweils diese externen Aufrufe vor und beschreibe wie diese durch die `Subnet`-Klasse abgearbeitet werden.

Der Konstruktor:

`$o_onesubnet = new subnet ('w', 't', $this->http_vars['grp']);` Ein neues Objekt (`$o_onesubnet`) der Klasse `Subnet` wird instanziiert. Dabei wird die Domäne „w“ (=Webverzeichnis) und der Typ „t“ (=Themengruppe) übergeben, um die abstrakten NIO-Datensätze richtig interpretieren zu können¹⁹. Als dritter Parameter wird die Variable übergeben, die die Nummer der aktuell fokussierten Gruppe beinhaltet. Durch den Aufruf des Konstruktors (`subnet.subnet($domain='w', $type='t', $g_idnr=focused_group_id);`) werden einige In-

¹⁸Der Quellcode dieser Klasse befindet sich im Anhang dieser Arbeit.

¹⁹Wie oben erwähnt wird auch das Benutzermodell durch Gruppen modelliert. Durch „w“ und „t“ wird garantiert, dass nur Themengruppen aus dem Webverzeichnis wiedergegeben werden.

stanzvariablen mit übergebenen Werten und Daten aus der Datenbank gesetzt. Dazu werden zwei SQL-Aufrufe ausgeführt: der erste gibt die direkten Nachbarn, der zweite die Nachbarsnachbarn zurück²⁰. Die Rückgabewerte der Datenbank werden umgewandelt und in den Instanzvariablen gespeichert. Ein Array (`$this->a_nodes`) nimmt dabei alle Themengruppen auf, das zweite Array (`$this->a_conns`) nimmt die Verknüpfungen zwischen den Knoten auf. Pro Verknüpfung werden zwei Datensätze angelegt, die je für Hin- und Rückrichtung stehen. Durch diese beiden Arrays werden die Kanten des Graphen, als Ausschnitt des Themennetzwerkes, modelliert.

Berechnung der Relationstypen: `$o_onesubnet->compute_relation_types()`;

Diese Methode analysiert die Verknüpfungen und speichert die Ergebnisse in dem Array „`$this->a_conns`“ pro Verknüpfung im Feld „`type`“. Folgender Funktionsaufruf (`subnet.compute_relation_types($tolerance=10)`;) mit dem Parameter „`$tolerance`“ - standardmäßig mit 10 initialisiert - führt dazu, dass pro Verknüpfung (in `$this->a_conns`) die angegebene Gewichtung mit der der Gegenrichtung verglichen wird: Aus den beiden absoluten Gewichten `w1` und `w2` werden zwei Intervalle (`i1` und `i2`) gebildet, indem von den beiden Gewichten je der Toleranzwert (`$tolerance`) abgezogen und aufaddiert wird. Nun können diese Intervalle folgendermaßen zueinander stehen:

- Keine Überlappung: Alle Werte von `i1` sind *kleiner* als alle Werte von `i2`
Dieser Fall führt zu einer Markierung am Index „`Type`“ in der aktuellen Verbindung (Element aus `$this->a_conns`) mit dem Zeichen „`<`“.
- Keine Überlappung: Alle Werte von `i1` sind *größer* als alle Werte von `i2`
Hier wird das Feld entsprechend mit „`>`“ belegt.
- Überlappung: Die Intervalle haben einige gemeinsame Werte.
In diesem Fall erhält das Array an dem Index „`Type`“ das Zeichen „`~`“.

Die Variable „`$tolerance`“ steuert also die Empfindlichkeit dieser Zuordnung: Ein kleiner Wert setzt nur wenige „`<`“ und „`>`“ Markierungen und viele „`~`“. Bei einem großen Wert kehrt sich dieses Verhältnis um. Dieser Wert ist bisher willkürlich gesetzt; einen bessern Wert kann man erst setzen, wenn einige Erfahrungen mit der Typisierung und dem Verhalten der Verknüpfungsgewichte gemacht worden sind.

Markierung der wenig benutzten Verknüpfungen:

`$o_onesubnet->mark_unused()`; Überschreibt die Variable „`$this->a_conns`“ beim Index „`Type`“ mit „`u`“ falls diese Verknüpfung selten benutzt wird und deshalb als sehr spekulativ gilt. Im Methodenaufruf (`subnet.mark_unused($threshold=2500)`;) werden alle Verknüpfungen, deren Gewichtswert (aus Spalte Zähler) unter dem Schwellenwert liegt, als unbenutzt („`u`“) beim Index „`Type`“ markiert, da erst Aussagen über den Typ der Verknüpfung gemacht werden können, wenn sie intensiv genug benutzt wurden.

²⁰Selbstverständlich ist SQL sehr ausdrucksstark und könnte auch in *einer* Anfrage die benötigten Daten selektieren. Ich habe trotzdem zwei Anfragen gestellt, da das verwendete Datenbanksystem diese schneller bearbeitet als wenn nur eine sehr komplexe Anfrage stattfinden würde.

Gliederung der Gruppen: `$o_onesubnet->find_structure()`; Dieser Methodenaufruf belegt die Instanzvariablen „`$this->a_nodeact`“ mit Werten durch Hilfe eines Spread-of-activation-Vorgangs und gibt die entsprechenden Ergebnisse aus (siehe oben: „suggesting topic structure(s)“). Standardmäßig werden die Untergruppen (durch den Methodenaufruf (**`subnet.find_structure($reltype='>')`**);) gegliedert, die durch den Relationstyp „>“ charakterisiert sind. Die Untergruppen, die direkt mit der aktuell fokussierten Gruppe verknüpft sind, werden durch einen weiteren Methodenaufruf (**`subnet.get_first_neighbours($reltype='>')`**;) bestimmt. Jede dieser Gruppen dient als Quelle für die Aktivierung des Netzwerks und wird mit einer Aktivierung von 1 initialisiert. Dazu dient die Instanzvariable „`$this->a_nodeact`“. Die Aktivierung fließt nun von jedem dieser Knoten über „>“-Relationen des Teilnetzes. Dabei wird die Aktivierung der verbundenen Knoten jeweils um einen Zähler erhöht. Zusätzlich enthält „`$this->a_nodeact`“ pro Knoten mit Aktivierungswert eine Liste mit Knotennummern, die ihn aktiviert haben. Die Aktivierung hat sich also immer nur über eine Verbindung ausgedehnt. Nun werden die am stärksten aktivierten Knoten ermittelt: es sind diejenigen, die mindestens einen Wert über 2 haben, also ihre eigene Aktivierung von 1, plus die Aktivierung von mindestens 2 verbundenen Knoten. Diese „Spitzenwerte“ werden in „`$a_peak_ids`“ mit ihrem Aktivierungswert gespeichert und dann abgearbeitet, so dass der Name des Knoten mit der stärksten Aktivierung ausgegeben wird und darunter erscheinen alle Knoten, von denen dieser Knoten aktiviert wurde (Darstellungsform: siehe oben).

In den nächsten Schritten folgen noch einige Methodenaufrufe, die die Ergebnisse dieser Subnet-Klasse in die Datenstruktur des Webverzeichnisses umformatieren und so integrieren. Auf diese soll hier nicht weiter eingegangen werden.

3.4 Konzept zur Visualisierung

Die Entwicklung eines Lösungsansatz zur Visualisierung würde den Umfang dieser Arbeit sprengen, da dieser Bereich noch sehr am Anfang seiner Entwicklung steht: Bisher haben sich alternative Visualisierungsansätze kaum durchgesetzt. Dies lässt sich z. B. an der breiten Menge unterschiedlicher Ansätze erkennen (siehe [WAECHTER 2002]). Trotzdem sehe ich es als meine Pflicht an, Ansätze aufzuzeigen, die die vorgeschlagene Netzwerkalternative visualisierbar und in einer grafischen Oberfläche nutzbar machen.

Generell kann diese Visualisierung in Textform stattfinden, ähnlich wie ein Dateiverwaltungsprogramm den Verzeichnisbaum darstellt. Alternativ ist auch eine grafische Darstellung denkbar, die die Charakteristika des Netzwerkes und des Bewegens im Netz von Knoten zu Knoten besser abbildet.

3.4.1 Ansatz auf Textbasis

Wie bereits teilweise auf den Bildschirmdarstellungen aus dem NIO-Projekt erkennbar (zum Beispiel im Kapitel 3.3.4) scheint mir folgende Darstellung gut brauchbar.

Speziellere Themen:

```
THEMA #1.1
* Unterthema #1.1.1
* Unterthema #1.1.2
...
```

```
THEMA #1.2
* Unterthema #1.2.1
* Unterthema #1.2.2
...
```

Verwandte Themen:

```
THEMA #2.1
THEMA #2.2
...
```

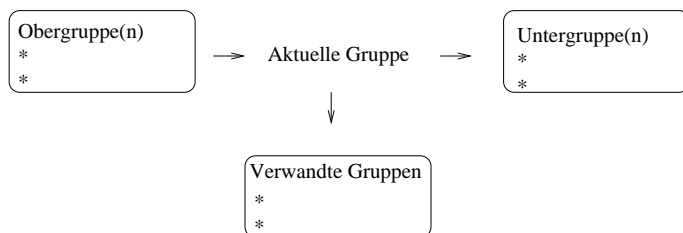
Allgemeinere Themen:

```
THEMA #3.1
THEMA #3.2
...
```

Analog zum Netz und den Relationstypen werden drei verschiedene Gruppen²¹ aus den verknüpften gebildet. Innerhalb der Gruppen können die Einträge nach Verknüpfungsstärke sortiert werden. Im Idealfall stehen dann die wichtigsten Gruppen, die im nächsten Schritt fokussiert werden, ganz oben am Bildschirm. Da normalerweise der Suchvorgang hin zu den spezielleren Gruppen geht, stehen diese am Anfang der Liste und sind näher aufgegliedert. Dann folgen die ähnlichen, verwandten Gruppen und schließlich werden alle Obergruppen am Ende der Liste dargestellt.

3.4.2 Grafischer Ansatz

Die grafische Version benutzt ebenfalls diese Dreiteilung der verknüpften Gruppen. Dabei wird die Bewegung vom Allgemeinen zum Speziellen auf eine räumliche Bewegung von links nach rechts abgebildet. Ein Springen auf ähnlicher Abstraktionsebene zu verwandten Themen ist dann eine Bewegung nach unten:



Zusätzlich kann bei diesem Ansatz noch die zuletzt fokussierte Gruppe hervorgehoben werden, sodass auch die Anzeichen eines Pfades sichtbar werden.

²¹ In NIO wird experimentell zusätzlich der vierte Relationstyp am Ende der Liste angegeben: die unbenutzten Gruppen, die so eine Chance erhalten in eine der ersten drei Gruppen der benutzten aufzusteigen, indem sie geklickt werden.

4 Ausblick

Nach dem Vorschlag eines adaptiven Netzwerkansatzes als Alternative zur hierarchischen Informationsgliederung, folgte die Implementation dieses Ansatzes [FROST 2001b, FROST . 2003b] in der Domäne des Internets und der Weblinks. Im dritten Schritt ist eine Evaluation notwendig. Bisher wurde diese auf einer informellen Ebene durchgeführt, indem die Entwickler und einige Studenten das System intensiv über einige Jahre benutzt, kritisiert und verbessert haben. Es haben sich dabei eine Menge Indizien angesammelt, die die hier vorgestellten Ideen bestätigen: Innerhalb des Themennetzwerkes bilden sich insbesondere bei häufig benutzten Pfaden durch das Themennetzwerk (z.B: Bildung / Beruf - Universitäten - Universität Osnabrück) sinnvolle hierarchische Strukturen heraus. Dies hängt insbesondere damit zusammen, dass eine thematische Suche meist bei sehr allgemeinen Themenfeldern beginnt, die dann im Laufe der Suche eingeschränkt werden. Bei selten benutzten Themengruppen oder selten benutzten Verknüpfungen zwischen Gruppen jedoch lassen sich zur Zeit keine Aussagen mit dieser Methode gewinnen. Es wird also deutlich, dass für diese Methode ein notweniger Grad an Aktivität durch die Benutzer vorhanden sein muss, bevor das System Relationen zwischen Themen ableiten kann.

Es ergeben sich verschiedene Möglichkeiten eine Evaluation durchzuführen:

1. Test unter Laborbedingungen, sowie
2. Blindstudie durch Auswertung von Logdaten der Benutzer.

Ein Test unter kontrollierten Bedingungen ist sehr aufwändig, da es sich um eine Langzeitstudie handeln müsste, die Alltagsverhalten in Bezug auf die Nutzung von Internetdiensten einbezieht. Wenn eine Blindstudie repräsentativ gelten soll, müssen genug Benutzer NIO/meinNetz über einen gewissen Zeitraum einsetzen. NIO sollte dabei so attraktiv sein, dass es freiwillig im Alltag von den Benutzern eingesetzt wird. Eine solche professionelle Evaluation wurde bisher nicht durchgeführt, da sich dabei einige Schwierigkeiten ergeben: NIO gliedert und organisiert seine Verzeichnisse an Hand des Verhaltens der Benutzer. Daraus folgt, dass am Anfang die Verzeichnisse nicht gut organisiert sind. Dies ist eine Hemmschwelle, da es auf neue Benutzer nicht sehr praktisch und eher chaotisch wirkt. Es ist also den von Menschen subjektiv und hierarchisch organisierten Verzeichnissen unterlegen. Würden aber viele Nutzer gezwungenermaßen einige Zeit das System benutzen, dann würden sich auch die Gewichte entsprechend anpassen und die Vorteile von NIO kämen zum Tragen. Wichtig ist dabei, dass die Benutzer ein breites Interessensspektrum abdecken und das Internet häufig benutzen, da sonst die Sichtweise zu einseitig ist: Das System passt sich nur einer kleinen spezifischen Gruppe an und ist keinesfalls repräsentativ über diese Gruppe hinaus.

Eine Möglichkeit - die beide Testarten kombiniert - für eine Evaluation bestünde darin, das System mit Testpersonen zu trainieren, indem bestimmte alltagstypische Aufgaben gestellt werden und so ein intensives Browsen durch die Gruppen erforderlich wird. Dann könnte man beobachten, wie sich das Netz verändert und wie gut die Testpersonen mit der Oberfläche und einer alternativen Strukturierung von Informationen zurechtkommen. Möglicherweise ließe sich dies mit entsprechenden Testläufen in hierarchischen Verzeichnissen (z.B.

[dmoz, YAHOO]) vergleichen. Ein solcher Testlauf könnte z.B. in Form von folgenden Aufgabenstellungen durchgeführt werden:

- *Finden Sie die Webseite der Mensa in Osnabrück!*
Beobachtung: Welcher Weg wird durch das Netzwerk gewählt, wo erwartet der Benutzer den Link zu finden?
In dieser Art von Aufgabenstellung können verschiedene Pfade zum Ziel gewählt werden. In diesem Beispiel: „Deutschland -> Niedersachsen -> Osnabrück: Mensa“ oder „Bildung / Beruf -> Universitäten -> Uni Osnabrück: Mensa“.
- *Versuchen Sie eine Webseite zu finden, über die man einen Flug buchen kann!*
Beobachtung: Welche Kategorien werden als potenzielle Fundstellen angesehen? Was passiert, wenn der Benutzer nie vorher versucht hat eine solche Webseite zu finden? Findet er auf Anhieb die von den anderen Nutzern am häufigsten verwendete?
Hier geht es also darum den kreativen Aspekt zu untersuchen, der wichtig wird, wenn man zum ersten Mal in einem Bereich sucht, den man noch nicht kennt. An dieser Stelle ist der Benutzer gut beraten sich auf die Erfahrung der anderen Benutzer zu verlassen und sich auf (von anderen) häufig benutzten Pfaden zu bewegen.
- *Mit welchen Themen hängt der Studiengang „Cognitive Science“ zusammen?*
Ist dem Benutzer die Netzwerkstruktur und ihre Interpretation bewusst und kann er das Wissen aus dem Netz extrahieren, indem er die Gruppe fokussiert und die verknüpften Themen nennt? Wie praktisch sind allgemein Netzwerkstrukturen im Vergleich zu hierarchischen Verzeichnissen? Welche Faktoren beeinflussen diese Entscheidung?

Ich hoffe, dass in einer späteren Studie das Projekt professionell evaluiert werden wird und somit auch diese Arbeit abgeschlossen werden kann.

Literatur

- [BASU . 1998] BASU, CHUMKI, H. HIRSH . W. W. COHEN (1998). *Recommendation as Classification: Using Social and Content-Based Information in Recommendation.* . AAAI/IAAI, . 714–720.
- [BIEZUNSKI . 1999] BIEZUNSKI, MICHEL, M. BRYAN . S. R. NEWCOMB (1999). *ISO/IEC 13250:2000 Topic Maps: Information Technology.*
- [BITZER 2002] BITZER, SEBASTIAN (2002). *meiNetz-Suchmodul.* Seminar Selected Topics in AI: Modeling User Interests, Hausarbeit + Projekt.
- [BRIN . PAGE 1998] BRIN, SERGEY . L. PAGE (1998). *The anatomy of a large-scale hypertextual Web search engine.* Computer Networks and ISDN Systems, 30(1–7):107–117.
- [BUZAN . BUZAN 2002] BUZAN, TONY . B. BUZAN (2002). *Das Mind-map-Buch: die beste Methode zur Steigerung Ihres geistigen Potenzials.* mvg, Landsberg am Lech.
- [CAMPBELL 2000] CAMPBELL, INGO (2000). *Biologie,* . 513. Spektrum Akademischer Verlag, Heidelberg.
- [dmoz] DMOZ. *dmoz.Org Internetverzeichnis (nicht-kommerziell, kollektiv organisiert): dmoz.org.*
- [EYSENCK . KEAN 2000] EYSENCK, MICHAEL W. . M. KEAN (2000). *Cognitive Psychology,* . 293. Psychology Press, Hove.
- [FROST 2001a] FROST, INGO (2001a). *Über meiNetz: Benutzermodell und Adaptivität.* Seminar Selected Topics in AI: User Adaptive Systems, Hausarbeit.
- [FROST 2001b] FROST, INGO (2001b). *Projekt meiNetz: Automatische Klassifikation und Benutzermodellierung.* . Tagungsband der GI-Workshopwoche: Lernen - Lehren - Wissen - Adaptivität, Dortmund. University of Dortmund, Department of Computer Science.
- [FROST . 2003a] FROST, INGO, R. FREUND . S. BITZER (2003a). *NIO/meiNetz - modelling users with information networks based on group sociology.* (Angenommen im 'Socio-cognitive Grids'-Workshop der 'Tales of the disappearing computer'- Konferenz 2003).
- [FROST . 2003b] FROST, INGO, R. FREUND . S. BITZER (2003b). *NIO/meiNetz - modelling users with information networks based on group sociology.* (Eingereicht zur Eurocogsci 2003).
- [GAZZANIGA . 1998a] GAZZANIGA, MICHAEL S., R. B. IVRY . G. R. MANGUN (1998a). *Cognitive Neuroscience,* . 4, . 121–162. W. W. Norton & Company, New York.
- [GAZZANIGA . 1998b] GAZZANIGA, MICHAEL S., R. B. IVRY . G. R. MANGUN (1998b). *Cognitive Neuroscience,* . 2, . 40. W. W. Norton & Company, New York.

- [GOLDSTEIN 1979] GOLDSTEIN, IRA P. (1979). *The genetic graph: A representation for the evolution of procedural knowledge*. International Journal of Man-Machine Studies, 11:51–77.
- [INSTITUTION 1985] INSTITUTION, BRITISH STANDARDS (1985). *Universal Decimal Classification Index (UDC)*. FID Publications 571.
- [KOHONEN . 2000] KOHONEN, T., S. KASKI, K. LAGUS, J. SALOJÄRVI, J. HONKELA, V. PAATERO . A. SAARELA (2000). *Self Organization of a Massive Document Collection*. IEEE Transactions on Neural Networks, Special Issue on Neural Networks for Data Mining and Knowledge Discovery, 11.
- [KOKINOV . YOVEVA 1996] KOKINOV, B. . M. YOVEVA (1996). *Context Effects on Problem Solving*. . *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, NJ. Erlbaum, Hillsdale.
- [MÜLLER 2000] MÜLLER, MARTIN E. (2000). *OySTER: Web Search as a playground for User Modelling Techniques*. . *Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen*, Osnabrück. University of Osnabrück.
- [NIE 2002] NIE, ANDREAS (2002). *Google*. Seminar Selected Topics in AI: Modeling User Interests, Hausarbeit.
- [RONTALER 1998] RONTALER, MARC (1998). *Osiris: Qualitative Fortschritte bei der Literaturrecherche*. . J. DASSOW, R. KRUSE, .: *Informatik '98 : Informatik zwischen Bild und Sprache*, Berlin u.a. Springer, Informatik aktuell.
- [SHARDANAND . MAES 1995] SHARDANAND, UPENDRA . P. MAES (1995). *Social Information Filtering: Algorithms for Automating "Word of Mouth"*. . *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems*, . 1, . 210–217.
- [SIMMEL 1890] SIMMEL, GEORG (1890). *Über sociale Differenzierung*, . *Über die Kreuzung socialer Kreise*. DigBib.Org, www.digbib.org/Georg_SimmeL1858/Ueber_sociale_Differenzierung.
- [Vivisimo] VIVISIMO. *Vivisimo Metasuchmaschine mit Clusterung der Suchergebnissen*: www.vivisimo.com.
- [WAECHTER 2002] WAECHTER, UDO (2002). *Visualisierung von Netzwerkstrukturen*. Seminar Selected Topics in AI: Modeling User Interests, Hausarbeit.
- [WebBrain] WEBBRAIN. *WebBrain Internetverzeichnis*: www.webbrain.com.
- [WÄTJEN . 1998] WÄTJEN, HANS-JOACHIM, B. DIEKMANN, G. MÖLLER . K.-U. CARSTENSEN (1998). *Bericht zum DFG-Projekt: GERHARD German Harvest Automated Retrieval and Directory*.
- [YAHOO] YAHOO. *YAHOO Internetverzeichnis (kommerziell)*: www.yahoo.com.