

Projekt meiNetz: Automatische Klassifikation und Benutzermodellierung

Ingo Frost, Robert Freund, Harjo Korte, Sebastian Bitzer, Stephan Weller

Institute of Cognitive Science, Universität Osnabrück

Zusammenfassung

Projekt meiNetz liefert ein Modell, dass die tägliche Informationsflut für den Einzelnen bündelt und einfacher zugänglich macht. Dabei werden soziale Komponenten (Gruppensoziologie) mit nichthierarchischen Informationsmodellen (bidirektional gewichtete Graphen) kombiniert. Auf diese Weise wird ein Rahmenwerk geschaffen, in dem zum Beispiel ein personalisiertes und adaptives Informationsportal für das Internet realisiert werden kann. Erste Implementationen werden bereits an der Universität Osnabrück in Zusammenarbeit der Fachgruppe Kognitionswissenschaft realisiert und evaluiert.

1 Einleitung

Eine der Herausforderungen im Informationszeitalter besteht darin, mit der Flut von Informationen zurechtzukommen. An dieser Schnittstelle zwischen Benutzer und Information lässt sich aus zwei Richtungen arbeiten:

- Betrachtet man den Benutzer, so lassen sich die Benutzerinteressen durch ein Benutzermodell modellieren und es kann eine Vorauswahl der Informationen, die speziell für den Benutzer potentiell interessant sein könnten, stattfinden.
- Aus Sicht der Informationswelt ist eine Klassifikation erforderlich, die Informationen Themen zuordnen kann und die Themen untereinander strukturiert.

Die meiNetz-Software versucht, beide Ansätze zu vereinen und findet z.B. Anwendung im Internet. Dort soll sie Benutzern behilflich sein, die für sie relevanten Informationen zu finden. Dabei steht die Adaptivität des Benutzermodells an den Benutzer und die ebenfalls adaptive Klassifikation von Informationen an zentraler Stelle.

2 Die Struktur von meiNetz

2.1 Benutzermodell

Anstatt die Benutzer unabhängig voneinander zu modellieren, versucht meiNetz mit Hilfe von Benutzergruppen, ähnliche Interessen gemeinsam zu modellieren. Diese Technik basiert auf der Gruppensoziologie, welche Individualität durch die unterschiedlichen Zugehörigkeiten zu sozialen Kreisen beschreibt [SIMMEL].

Ein sozialer Kreis verbindet Menschen mit einem zentralen Interesse. Das Individuum ist also Teilmenge aus verschiedenen Kreisen (hier Gruppen).

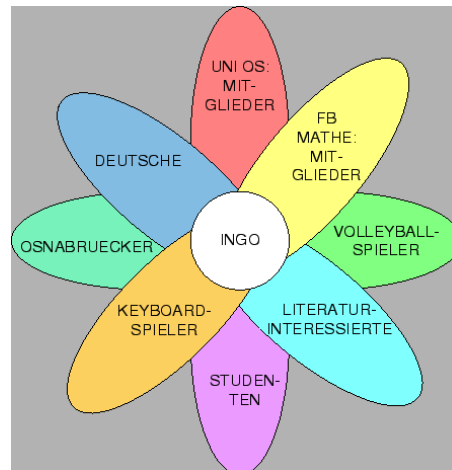


Abbildung 1 beschreibt diese soziale Struktur aus Sicht eines Einzelnen. Ingo ist Teil verschiedener sozialer Kreise, die Aussagen über seine Individualität machen können. So ist er beispielsweise Student, also Mitglied der sozialen Gruppe Studenten. Individuelle Eigenschaften, die sich nicht als soziale Gruppenbeziehungen darstellen lassen (soweit diese überhaupt existieren), bezeichne ich als Innere Individualität, die hier als weiß ausgefüllter Kreis dargestellt ist.

Bei dem Versuch, Gruppen untereinander in Beziehung zu setzen, kann man rein analytisch vorgehen: Sobald eine Schnittmenge, also Menschen, die beide Interessen haben, vorhanden ist, besteht auch ein Zusammenhang zwischen den beiden Gruppen. Dieser Zusammenhang lässt sich weiter differenzieren, indem man die Größe der Schnittmenge mit den Größen der beiden Gruppen in Bezug setzt.

Diese Anordnung der Gruppen kann man sich visualisiert als Netz vorstellen. Gruppen entsprechen Knoten, die mit jeder beliebigen anderen Gruppe und beliebig vielen Gruppen verbunden sein können. Diese Datenstruktur kann man mit Hilfe eines Entity-Relationship-Modells darstellen, das folgende Aussagen macht:

- 1 Als Objekttyp (Entität) lassen sich Person und Gruppe definieren.
- ⊗ Eine Person kann somit mit mehreren Gruppen verbunden sein (an Gruppen teilnehmen).
- ⊗ Umgekehrt kann eine Gruppe mit mehreren Personen verbunden sein (Gruppenmitglieder haben).
- ⊗ Jede Gruppe kann mit mehreren anderen Gruppen verbunden sein (Gruppennetz).

2.2 Informationsmodell

Um große Mengen von Informationen zugänglich und überschaubar zu machen, werden oft hierarchische Klassifikationen [UDC] (im Internet: [OySTER]) benutzt. Eine hierarchische Klassen-

anordnung in Form eines Baums bietet nur einen Weg zu einer Information. Offensichtlich bestimmen Individuum und Kontext die Relevanz. Eine durch eine Person erstellte Einwegklassifikation ist also subjektiv und löst sich vom jeweiligen Zusammenhang, was das Auffinden einer Information schwieriger macht. Weiter ist dieses System sehr starr, denn Modifikationen in unteren und mittleren Ebenen sind problematisch, da übergeordnete Ebenen neu angeordnet werden müssen und gewohnte Pfade sich ändern.

- ⑩ Zwei Objekttypen werden benötigt: Verweis zu Information, der zum Beispiel ein Weblinkobjekt sein kann und Gruppen, die einem Thema oder Cluster entsprechen.
- ⑩ Jeder Verweis kann mit beliebig vielen Gruppen verbunden sein (Informationsverweise spielen in verschiedenen Zusammenhängen, also Themengruppen, eine Rolle).
- ⑩ Umgekehrt kann ein Thema mit mehreren Verweisen zu Informationen verbunden sein (Themengruppe enthält mehrere Informationsverweise).
- ⑩ Jede Themengruppe kann mit beliebigen anderen Themengruppen verbunden sein. (Gruppen-netz).

Damit eine automatische Zuordnung von Verweisen zu Gruppen möglich ist, sind zusätzlich Schlüsselwörter notwendig, die einen Teil der Verweiseigenschaften, als auch der Gruppeneigenschaften repräsentieren.

2.3 Kombination zu NIO

Die beiden Entity-Relationship-Diagramme sind strukturell sehr ähnlich und werden durch das Netzwerk zur Identifikation des Benutzers und Orientierung in der Informationswelt (NIO) zu einem Modell.

In dem Benutzermodell wurden soziale Kreise durch ein übergeordnetes Interesse definiert. Auf dieser Basis können nun die Cluster bestimmt und bezeichnet werden, so dass Gruppen soziale Kreise und gleichzeitig Themen zur Gliederung von Informationen sind. Dies ergibt auch aus anderen Perspektiven Sinn, denn oft produzieren soziale Kreise Informationen für den eigenen sozialen Kreis (Rollenverhalten). So fallen also soziale Kreise und Themengruppen zu einer abstrahierten Form Gruppe zusammen. Die Beziehungen zwischen den Entitäten Schlüssel (für Schlüsselwörter), Gruppe und Verweis sind many-to-many-Beziehungen, die über die Tabellen `G_S`, `S_V`, `G_V` und `G_G` im Datenmodell realisiert werden. Diese Verknüpfungstabellen können zusätzlich Rangpositionen aufnehmen. So kann beispielsweise ein Verweis in einer Gruppe einen hohen Rang und in einer anderen Gruppe einen niedrigen Rang haben. Nur so kann die Qualität der Klassifikation gewährleistet werden, denn nur Zuordnungen mit hohen Rängen machen wirklich treffende Aussagen über die Gruppe.

Eine konkrete Beschreibung des Datenmodells und der Berechnung der Ränge folgen im nächsten Kapitel.

3 Organisation neuer Daten

Neue Daten können nur dann automatisch aufgenommen und klassifiziert werden, wenn bereits ein kleines Netz aus Gruppen und Schlüsselwörtern vorhanden ist. Wir gehen also davon aus, dass bereits eine einfache Struktur besteht und es zu einem Dazulernen kommen kann.

Momentan ist vorgesehen, dass Teilnehmer einer Gruppe Verweise selbst hinzufügen können, die sie in dem Themenzusammenhang nutzen. Auf diese Weise können Rechercheergebnisse vieler Einzelner zusammengeführt werden. Sobald eine Gruppe von mehreren Teilnehmern genutzt wird, ist auch der Einsatz eines Web-Agenten denkbar, der dann den Gruppenmitgliedern neue Informationsblöcke vorschlägt.

Diese Vorgehensweise erweist sich als fehlertolerant, da fehlerhafte oder nicht mehr zeitgemäße Zuordnungen von Verweisen in Gruppen gar nicht mehr oder weniger genutzt werden und so im Rang fallen. Fehlende Zuordnungen werden durch eine automatische Probezuordnung in verwandte Gruppen gefunden.

Die Zuordnungen von Verweisen zu Gruppen entsprechen also Vorschlägen an eine Gruppe, die eher angenommen oder eher angelehnt werden können. So wirkt der tägliche Informationskonsum der Gruppenmitglieder rückkoppelnd zu der Einordnung der Verweise in die Gruppen.

Jede Zuordnung eines Verweises zu einer Gruppe ist mit einem Rangwert versehen, so dass die Verweise einer Gruppe feste Positionen haben. Diese Positionen verändern sich durch das Verhalten der Gruppenmitglieder. Oft benutzte Verweise steigen auf, in der letzten Zeit weniger oft verwendete Verweise steigen ab. In einer ersten Realisierung von *meiNetz* für Internetinformationsangebote über Osnabrück (<http://www.meinetz.de>) werden dabei nur die Klicks pro Gruppe gezählt. Spielt ein Informationsangebot in einer Gruppe eine wichtige Rolle, so wird diese im Alltagsgebrauch von *meiNetz* häufig benutzt und erhält so einen hohen Rang. Voraussetzung ist, dass *meiNetz* im Alltag als Internetportal und nicht nur als gelegentliche Recherchemöglichkeit genutzt wird.

3.1 Neue Gruppen

Die ersten Gruppen werden manuell angelegt, manuell mit anderen Gruppen verknüpft und manuell durch Schlüsselwortzuordnungen beschrieben. Alle drei Vorgänge können, sobald eine kritische Menge von NIO-Daten erreicht ist, automatisiert werden. Teilweise geschieht dies schon: So kann, nachdem ein Verweis, bei dem keine Schlüsselwörter vorliegen, manuell Gruppen zugeordnet worden ist, NIO die Schlüsselwörter der zugeordneten Gruppen als Schlüsselwörter für den Verweis vorschlagen. Umgekehrt ist diese Technik auch bereits im Einsatz: Sind Schlüsselwörter vorhanden, aber keine Gruppenzuordnungen, so können an Hand der Schlüsselwörter passende Gruppen ermittelt werden.

Wird die Menge an Verweisen in einer Gruppe zu groß und unübersichtlich, kann durch ein Klusteringverfahren festgestellt werden, ob Verweise zusätzliche Eigenschaften haben, durch die eine kleinere Gruppe möglich wäre. Beispielsweise könnte die Gruppe Bäume eine neue Gruppe Laubbäume bilden und alle Verweise aufnehmen, die neben dem Schlüsselwort Baum auch das Schlüsselwort Laub oder Blatt haben. Dabei bleibt die vorhandene Gruppe erhalten.

So sollen aber keine Untergruppen in Form von Teilmengen von bereits bestehenden großen Gruppen gebildet werden, da solche hierarchischen Ansätze grundsätzlich subjektiv sind. Aber man kann trotzdem von verfeinerten, spezialisierteren Gruppen sprechen, die auch Eigenschaften von anderen Gruppen erben können, aber dennoch gleichberechtigt und autark bleiben.

Bezogen auf das Baumbeispiel bedeutet das: Würde man eine Gruppe Ahorn anlegen, so könnte man Ahorn mit der Gruppe Kanada verknüpfen. Eine Verknüpfung Kanada-Laubbaum hat jedoch nur wenig Sinn.

Teilen sich zwei Gruppen Schlüsselwörter oder nehmen viele Nutzer an zwei Gruppen teil oder gehören mehrere Verweise zwei Gruppen an, so kann automatisch eine Verknüpfung zwischen diesen beiden Gruppen angelegt werden.

Wenn fast alle Verweise in einer Gruppe Schlüsselwörter enthalten, die jedoch nicht der Gruppe zugeordnet sind, so können diese Schlüsselwörter zu Gruppenschlüsselwörtern werden.

3.2 Neue Informationsblöcke

meiNetz selbst nimmt keine Informationen auf. Es wird vielmehr ein Verweis zu einem Informationsblock im Sinne einer Ansammlung von Informationen zu einem Thema aufgenommen. Ein solcher Verweis kennt jedoch wichtige Eigenschaften des Informationsblocks, die durch Schlüsselwörter abgebildet werden. Ein neuer Verweis (Tabelle VERWEIS) wird in NIO also zuerst mit den bereits existierenden Schlüsselwörtern verbunden (Tabelle S_V). Gegebenenfalls werden dabei neue Schlüsselwörter angelegt (Tabelle SCHLÜSSEL). Wenn nun die schon bekannten Schlüsselwörter mit Gruppen verknüpft sind (Tabelle G_S), wird der neue Verweis ebenfalls mit diesen Gruppen verknüpft (Tabelle G_V).

3.3 Neue Benutzer

Ein neuer Benutzer wird als Gruppe (Metaobjekttyp Benutzergruppe) in NIO angelegt, die mit für den Benutzer wichtigen Gruppen als auch mit den wichtigen Verweisen (auch Innere Individualität) verknüpft werden kann. Am Anfang ist die Benutzergruppe nicht mit anderen Gruppen verknüpft, da dann noch keine Aussagen über den Benutzer gemacht werden können. Gruppen, die von dem Benutzer angesteuert werden und Verweise, die von ihm benutzt werden, können automatisch der Benutzergruppe zugeordnet werden, so dass nach und nach eine individuelle Sicht auf NIO entsteht. Die besonders häufig benutzten Gruppen und Verweise steigen aus Sicht des Benutzers im Rang auf, so dass eine weitere Differenzierung stattfinden kann. Bei der Benutzergruppe entscheidet das Verhalten des einzelnen Benutzers (individuelle Sicht), bei den anderen Gruppen das Verhalten der Gruppenmitglieder (kollektive Sicht).

4 Fazit

Werden die hier vorgestellten Modelle in Form eines Internetdesktops oder Internetprotals vollständig implementiert, und beim täglichen Informationskonsum eingesetzt, ergeben sich folgende Möglichkeiten:

⑩ Intuitives Browsen

Diese bereits im Internet rudimentär entwickelte Technik wird durch bidirektional gewichtete Gruppenverknüpfungen perfektioniert. So wird eine intuitive Suche durch Assoziieren ermöglicht.

⑩ Strukturierte Suchergebnisse

Anonyme Suchanfragen innerhalb von meiNetz können automatisch klassifiziert werden, in dem die Verweis-Gruppenzuordnungen sowie Metainformationen (durch Gruppenverknüpfungen) berücksichtigt werden.

⑩ Personalisierte Schnittstelle

Dem Portal liegt nach einiger Zeit ein Benutzermodell vor. Dadurch können Suchanfragen nochmals optimiert werden: Jemand, der zum Beispiel viele Informationsdienste der Universität Osnabrück benutzt, erhält mit der Suchanfrage „Speiseplan“ als erstes Suchergebnis den Speiseplan der Osnabrücker Mensa.

⑩ Relevanz: gewichtete Mehrfachzuordnung

Durch die Tatsache, dass Verweise mehrfach Gruppen zugeordnet und diese dann dynamisch gewichtet werden, kann Kontextrelevanz als auch zeitsensitive Relevanz abgebildet werden. Verliert ein Verweis seine Aktualität in einer Gruppe, so wird er weniger benutzt und somit fällt seine Relevanz innerhalb dieser Gruppe.

- Automatische fehlertolerante Gruppenzuordnung

Durch ständiges Feedback, dass zum Beispiel durch einfaches Zählen der Verweisklicks innerhalb von Gruppen stattfinden kann, können fehlerhafte Zuordnungen von Verweisen zu Gruppen schnell korrigiert werden.

⑩ Semantisches Netzwerk

Durch die dynamische Datenstruktur und die Metadaten ergibt sich indirekt ein semantisches Netzwerk. Es lässt sich also ohne Probleme feststellen, in wie weit ein Schlüsselwort mit einem anderen zu tun hat (gemeinsame Gruppen bzw. Verweiszuordnungen). Diese Tatsache zeigt, dass das Wissen innerhalb des Netzes zunimmt. Dadurch können neue Daten besser klassifiziert werden, da dies auf der Basis des vorhandenen Wissens passiert.

⑩ Berücksichtigung von sozialen Strukturen

Anonyme Informationsdienste wie Yahoo [Yahoo] oder Google [Google] berücksichtigen keine sozialen Strukturen, die jedoch sehr wichtig sind, um Informationen zu gliedern und um Aussagen über Relevanz zu treffen.

⑩ Networking - Erfahrungen von anderen nutzen

Dadurch kommt man dem psychologischen Konzept des „Networkings“ nahe: Erfahrungen von anderen Personen, die Spezialisten in ihren persönlichen Gruppen sind, können genutzt werden.

5 Aussicht

Wenn meiNetz tatsächlich als Softwareschnittstelle zwischen Benutzer und Internet eingesetzt und diese gut genutzt wird, dann erhält NIO das zur exakten Klassifikation und Verbesserung des Benutzermodells nötige Feedback, so dass ein sich selbst strukturiertes autonomes Informationsnetz wachsen kann. Dabei ist dieses System nicht auf das Internet begrenzt und könnte beispielsweise Bibliothekskataloge besser zugänglich machen. Auch eine Kombination aus verschiedenen Informationstypen ist denkbar.

6 Literatur

[SIMMEL] Georg Simmel, *Über soziale Differenzierung*, Kap. Über die Kreuzung sozialer Kreise
Amsterdam, Liberac N.V. publishers, 1966

[UDC] *Univesal Decimal Classification Index (UDC)* British Standards Institution, 1985; FID
Publication 571

[OySTER] Martin Mueller, *OySTER: Web Search as a playground for User Modeling Techniques*
In: 8. GI Workshop Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen,
(ABIS 2000), 2000

[Yahoo] <http://www.yahoo.com>

[Google] <http://www.google.com>