

Benutzermodelle, Information Retrieval und Visualisierung

Swantje Willms
University of Pittsburgh, Pennsylvania, USA
swillms@mail.sis.pitt.edu

Abstract: Wir stellen eine zweischichtige Visualisierung vor, die StudentInnen helfen soll, Dokumente zu identifizieren, die für ihr Studium relevant sind. Während die erste Visualisierung eine Übersicht bietet, zeigt die zweite eine adaptierte Perspektive: Wir visualisieren die Beziehung des Benutzermodells zu den Dokumenten, die für die Aufgabe der BenutzerIn relevant sind; jede Benutzermodellkomponente ist durch einen Referenzpunkt repräsentiert.

1 Einleitung

Wir möchten Information Retrieval (IR) und Benutzermodellierung kombinieren, indem wir das Benutzermodell in Bezug auf die Dokumente, die für die Aufgabe der BenutzerIn relevant sind, darstellen. Zur Zeit verwenden wir diese Visualisierung in einer Lernumgebung, wo der Zweck der Visualisierung darin besteht, den StudentInnen zu helfen, Dokumente zu identifizieren, die für ihr Studium relevant sind. Die StudentInnen haben neben den Vorlesungsunterlagen auch Zugang zu weiteren externen Materialien. Oft greifen die StudentInnen aber nicht auf diese externen Unterlagen zu, da sie deren Relevanz nur schwer einschätzen können. Unsere adaptive Visualisierung in WebVIBE, die auf Referenzpunkten basiert, soll ihnen dabei helfen.

Das Vektormodell [SWY75] bietet eine konsistente Repräsentation aller Komponenten des Systems als Vektoren, d.h. für die Repräsentation der Dokumente, der Benutzermodellkomponenten und anderer Referenzpunkte. Ein Referenzpunkt ist ein beliebiger Punkt innerhalb des multidimensionalen Informationsraums, der durch die Indexbegriffe beschrieben werden kann. Es ist ein Punkt, mit dem man die Dokumente vergleichen kann.

2 Verwandte Arbeiten

Oftmals ist die Anfrage oder der Informationsbedarf einer BenutzerIn der einzige Aspekt, der von Information-Retrieval-Systemen als Benutzerstandpunkt berücksichtigt wird.

Inzwischen hat man aber den Bedarf erkannt, Benutzermodellierungstechniken für Visualisierungssysteme zu verwenden [Gra01]. Es gibt ein paar experimentelle Systeme, die Benutzermodelle visualisieren, z.B. PeerGlass [Kli95] und VIUM [UK03]. Diese set-

zen jedoch die Visualisierung des Modells selbst als Fokus, während unser System die Beziehung zwischen dem Benutzermodell und den Dokumenten darstellt. Es gibt einige Forschung über räumliche Visualisierung für IR-Aufgaben, aber die Referenzpunkte waren hauptsächlich Anfragen und ähnliches. Obwohl das Benutzerprofil von Korfhage [Kor97, S.163] als Beispiel für einen Referenzpunkt genannt wird, wurde der Aspekt der Berücksichtigung von Benutzerprofilen in der Forschung bisher kaum betrachtet.

3 Materialien

Die StudentInnen haben Zugang zu den Vorlesungsunterlagen und weiteren, externen Dokumenten. Die Vorlesungsunterlagen (d.h. interne Dokumente) für die Vorlesung “Introduction to Programming” (“Einführung in die Programmierung”) bestehen aus den Folien, die der Professor während der Vorlesung verwendet. Außer zu den Vorlesungsunterlagen haben die StudentInnen Zugang zu weiteren Dokumenten über das Web. In der Umgebung, die wir zur Zeit untersuchen, gibt es insbesondere einige Tutorials als externe Dokumente. Diese Tutorials können nicht nur als ganzes, sondern auch Seite für Seite angesprochen werden.

4 Das Benutzermodell

Ein Benutzermodell besteht aus der Menge der Vorstellungen, die das System über die individuelle BenutzerIn hat. In unserem System besteht das Benutzermodell aus verschiedenen Komponenten: dem Vorlesungsthema als **Langzeit-Ziel**, den **Interessen** und dem **Wissen** der StudentInnen. In der Visualisierung wird jede dieser Komponenten als separater Referenzpunkt behandelt.

Das jeweilige Vorlesungsthema bildet das **Langzeit-Ziel**, evtl. über mehrere Vorlesungen hinweg. Es wird durch Begriffe repräsentiert, die aus den entsprechenden Vorlesungsunterlagen extrahiert werden. Wenn eine Vorlesung mehrere Themen behandelt, kann sie in mehrere Themen unterteilt werden. Dies Ziel bleibt unverändert, bis die StudentIn ein neues Ziel wählt.

Das **Interesse** der StudentInnen wird initialisiert, indem sie am Anfang einer Sitzung ihr Ziel für die Sitzung angeben. Hieraus extrahierte Begriffe bilden den ersten Interessenvektor. Danach drücken die StudentInnen ihr **Interesse** aus, indem sie die Relevanz von Dokumenten bewerten, die für sie von Interesse sind. Der Prozess ist ähnlich zu dem, der für WebMate verwendet wird [CS98]: Zuerst werden die positiv bewerteten Dokumente vorbereitet, indem Stoppwörter gelöscht werden und die Wörter auf ihren Wortstamm reduziert werden. Dann wird der Frequenzvektor für das Dokument extrahiert. Dieser wird entweder zu der Menge von Vektoren addiert, die das Interesse der StudentIn darstellen, oder seine Werte werden zu dem Vektor addiert, der ihm am ähnlichsten ist.

Das **Wissen** der BenutzerIn wird repräsentiert, indem Wörter einbezogen werden, die sich

auf Testergebnisse beziehen. Die StudentInnen absolvieren jeweils einen Test, nachdem ein Vorlesungsthema behandelt worden ist. Jede Frage in diesem Test stellt den StudentInnen mehrere Antworten zur Auswahl. Jede dieser Fragen ist assoziiert mit einigen Wörtern oder Konzepten, die sich auf die Frage beziehen. Wenn die StudentIn die Frage richtig beantwortet, wird angenommen, dass sie das Konzept gelernt hat, das durch diese Wörter repräsentiert wird, und das Wort (der Wortstamm) wird in den Wissensvektor aufgenommen. Das Gewicht des Wortes im Wissensvektor wird erhöht, wenn die StudentIn weitere Fragen über dasselbe Konzept richtig beantwortet.

5 Adaptive Referenzpunkt-basierte Visualisierung

Der Sinn der Visualisierung der Benutzerprofilkomponenten als Referenzpunkte ist, den StudentInnen zu helfen, Dokumente zu identifizieren, die für ihre Vorlesung relevant sind und ihnen beim Studium helfen können. In der Vergangenheit hat man solche relevanten Dokumente lediglich als Listen dargestellt. Eine mehrdimensionale Darstellung könnte aber besser verständlich sein. Sie ist flexibler, denn sie kann zeigen, wie die Dokumente zu den Referenzpunkten in Beziehung stehen.

In einer räumlichen Visualisierung wird jedes Dokument als ein Punkt im Dokumentenraum dargestellt. Die Platzierung der gefundenen Dokumente erfolgt in Bezug auf die Ähnlichkeit zu den Referenzpunkten, die durch das Benutzermodell vorgegeben sind. Allgemeine Trends über die Verteilung einer Menge von Dokumenten können durch die räumliche Darstellung sichtbar werden. Räumliche Visualisierungen stellen explizit die unterschiedlichen Ähnlichkeitsgrade dar.

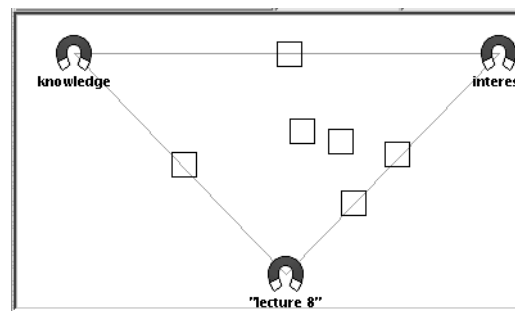


Abbildung 1: Eine beispielhafte WebVIBE-Darstellung

Wir verwenden WebVIBE [Web] (Abbildung 1), die Java-Version von VIBE [OKS⁺93] als unsere Benutzerschnittstelle zur räumlichen Visualisierung. VIBE (Visual Information Browsing Environment) präsentiert jedes Dokument als eine Entität. VIBE benutzt für die Darstellung die relativen Ähnlichkeiten eines Dokuments zu mehreren Referenzpunkten. Ein gewichtetes Zentroidmodell platziert die Dokument-Icons (Rechtecke) je nach Stärke ihrer Beziehung zu den Referenzpunkten (Magneten). Zu Beginn befinden sich

die Referenzpunkte am Rande des Darstellungsfensters, können aber von der BenutzerIn umplatziert werden. Die Ähnlichkeit zwischen den Referenzpunkten und den Dokumenten wird mit Hilfe des Cosinus-Maßes berechnet, das im Information Retrieval beliebt ist [SM83].

6 Anwendung

In der Lernumgebung, die wir untersuchen, ist bisher eine Visualisierung verwendet worden, die auf einer kartenartigen Übersicht basiert: KnowledgeSea [BR02] visualisiert alle Konzepte, die sich auf die Vorlesung über das ganze Semester hinweg und all die dazugehörigen Dokumente beziehen. Sie zeigt eine zweidimensionale Karte der Lerndokumente mit einer Liste von Wörtern in jeder Zelle, die die Konzepte in dieser Zelle beschreiben (Abbildung 2). Dokumente, die semantisch verwandt sind, befinden sich auf der Karte nahe beieinander. Wenn sie sich in derselben Zelle befinden, werden sie als sehr ähnlich angesehen.

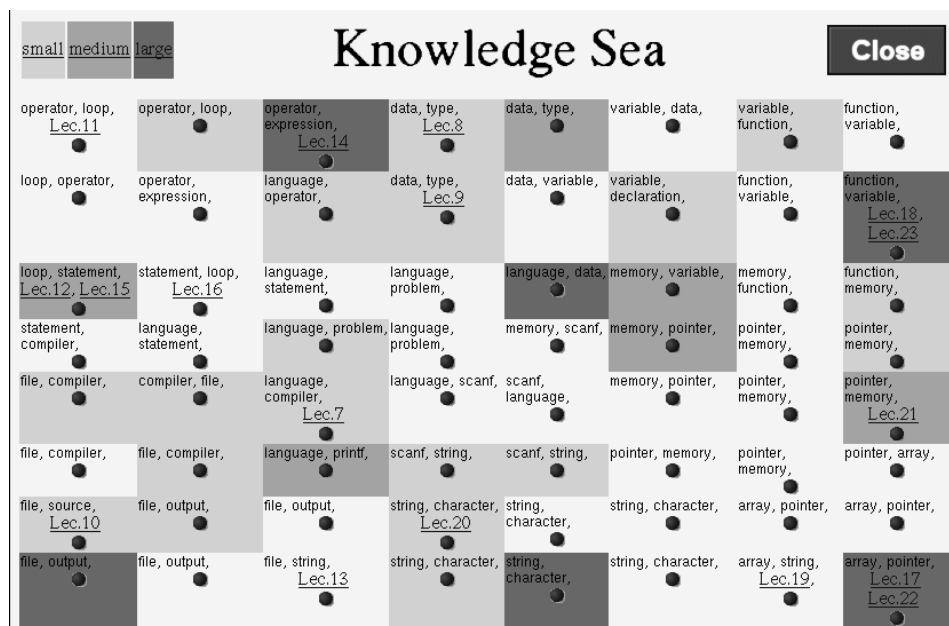


Abbildung 2: KnowledgeSea-Karte für die Vorlesung “Einführung in die Programmierung” (“Introduction to Programming”)

Wir verwenden die WebVIBE-Visualisierung im Zusammenhang mit der KnowledgeSea-Visualisierung. Bisher wurden die BenutzerInnen von jeder Zelle von KnowledgeSea aus mit einer Liste von Dokumenten verbunden, die sich auf die Konzepte dieser Zelle bezogen. Stattdessen werden wir die BenutzerInnen nun mit einem WebVIBE-Fenster (Abbildung

1) verbinden, so dass sie sehen können, wie die Dokumente miteinander und mit den Referenzpunkten in Beziehung stehen.

Wir verwenden also eine zweischichtige Visualisierung: (1) Kartenartige Übersicht in KnowledgeSea, (2) Adaptive Referenzpunkt-Visualisierung einer Zelle mit WebVIBE. Wir passen die Implementation von WebVIBE so an, dass wir die adaptiven Referenzpunkte unseres Benutzermodells verwenden können und dass wir es zusammen mit KnowledgeSea verwenden können.

Das WebVIBE-Fenster, das den StudentInnen präsentiert wird, hat mindestens drei Referenzpunkte: ihr **Interesse**, ihr **Wissen** und das Vorlesungsthema als **Langzeit-Ziel**, wie oben beschrieben. Man wird erwarten, dass die StudentInnen über die Konzepte, die weit von ihrem Wissen entfernt sind, aber nahe ihrem Interesse oder dem Vorlesungsthema liegen, etwas lernen möchten.

Wir werden im Laufe des Semesters verfolgen, auf welche Dokumente die StudentInnen zugreifen. Mit diesen Daten können wir herausfinden, ob es wahrscheinlicher ist, dass die StudentInnen auf externe Materialien zugreifen, wenn sie die Gelegenheit haben, auf diese durch die Visualisierung zuzugreifen.

Außerdem planen wir Sitzungen mit StudentInnen, in denen sie die Aufgabe haben, Material zu identifizieren, das ihnen beim Studium helfen kann. In dieser kontrollierten Situation können wir herausfinden, ob die StudentInnen relevante Dokumente schneller finden, wenn sie die Visualisierung verwenden.

7 Zusammenfassung und Ausblick

In diesem Paper haben wir vorgestellt, wie man ein Benutzermodell im Kontext von Information Retrieval in einer Lernumgebung visualisieren kann. Die Visualisierung präsentiert jede Benutzerprofilkomponente als einen Referenzpunkt. Dadurch können die BenutzerInnen die Beziehung zwischen den Dokumenten und dem Benutzermodell beurteilen.

In Zukunft könnte man außerdem untersuchen, was sich ändert, wenn nicht vorgegebene Materialien verwendet werden, sondern im Laufe der Zeit neue Materialien gefunden und eingebunden werden. Außerdem könnte man WebVIBE unabhängig von KnowledgeSea und auch in anderen Umgebungen verwenden, in denen die Visualisierung von einem Benutzermodell und Dokumenten sinnvoll ist.

8 Danksagung

Ich möchte mich bei Peter Brusilovsky und Michael Lewis für die Anregung zu diesem Forschungsthema und ihre Betreuung meiner Arbeit bedanken.

Literatur

- [BR02] Peter Brusilovsky and Riccardo Rizzo. Map-Based Horizontal Navigation in Educational Hypertext. *Journal of Digital Information*, 3(1), 2002.
- [CS98] Liren Chen and Katia Sycara. WebMate: A Personal Agent for Browsing and Searching. In Katia P. Sycara and Michael Wooldridge, editors, *Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98)*, pages 132–139, New York, 9–13, 1998. ACM Press.
- [Gra01] Beate Grawemeyer. User adaptive information visualization. In *HCT-2001: 5th Human Centered Technology Postgraduate Workshop*, University of Sussex, School of Cognitive and Computing Sciences, September 2001.
- [Kli95] Jill Kliger. Model Planes and Totem Poles: Methods for Visualizing User Models. Master's thesis, MIT, 20 Ames St, Cambridge, MA 02139, May 1995.
- [Kor97] Robert Korfhage. *Information Storage and Retrieval*. Wiley, New York, 1997.
- [OKS⁺93] K.A. Olsen, R.R. Korfhage, K.M. Sochats, M.B. Spring, and J.G. Williams. Visualization of a document collection: The VIBE system. *Information Processing and Management*, 29:69–81, 1993.
- [SM83] G. Salton and M.J. McGill. *Introduction to modern information retrieval*. McGraw Hill, 1983.
- [SWY75] G. Salton, A. Wong, and C.S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18:613–620, 1975.
- [UK03] James Uther and Judy Kay. VIUM, A Web-based visualisation of large user models. In *UM2003 User Modeling: Proceedings of the Ninth International Conference*. Springer, 2003.
- [Web] Homepage für WebVIBE. URL: <http://www2.sis.pitt.edu/~webvibe/WebVibe/>.