

BAND: Benutzeradaptiver Netzinformationsdienst

R. Möller, K. Hidde, R. Joswig, Th. Mantay, B. Neumann

Labor für Künstliche Intelligenz, Fachbereich Informatik, Universität Hamburg
„Lavielle“ EDV Systemberatung, GmbH & Co., Hamburg

Zusammenfassung In dem Projekt BAND wurden Grundlagen zur Entwicklung benutzeradaptiver Netz-Informationendienste am Beispiel eines „TV-Assistenten“ entwickelt. Zugriffs und Präsentationstechniken für Netzinformation werden durch wissensbasierte Techniken dynamisch an die Bedürfnisse eines Benutzers angepaßt. Die erreichten Ergebnisse wurden anhand einer prototypischen Implementation des TV-Assistenten evaluiert.

1 Einleitung

Information entwickelt sich in den letzten Jahren immer mehr zu einem zentralen Bedarfsgut in der Wirtschaft, wie auch im privaten Alltag der Menschen. Zur Distribution und Akquisition von Information ist die Kommunikationstechnologie elementares Element der Investitionen der Unternehmen. Aber nicht nur die Kommunikation zwischen Anbieter und Kunde wird verbessert, sondern auch der betriebsinterne Informationsfluß. Insbesondere die Ausstattung der Unternehmen (und im vermehrten Maße auch die der privaten Haushalte) mit Computern sowie deren innerbetriebliche Vernetzung („Intranet“) bzw. Anbindung an überregionale Netzanbieter („Carrier“, „Provider“) hat einen plattformunabhängigen Hypertext-Standard gefördert: Das World Wide Web, kurz: WWW oder auch Web.

Im Kontext des WWW ist die graphische Gestaltung von Oberflächen enorm wichtig geworden. Inzwischen gibt es Firmen, die sich speziell mit dem Design von Web-Anwendungen, der Einbindung von Werbemöglichkeiten u.a. auch mit Animationen usw. beschäftigen. Dennoch sind fast alle heute im WWW verfügbaren Informationsdienste ausschließlich Adaptionen von klassischen, maskenorientierten, meist datenbankbasierten Lösungen. In einem graphischen Kontext werden als Resultat einer mehr oder weniger komplexen Anfrage meist Listen oder Tabellen ausgegeben. Die genaue Spezifikation einer Anfrage ist für die erreichbaren Resultate sehr bedeutsam. Die Anfragespezifikation stellt bei der angebotenen Informationsvielfalt im Web insbesondere für ungeübte Benutzer einen besonderen Schwierigkeitsgrad dar. Da viele Web-Anwendungen sich an die breite Masse von Nicht-Experten wendet, müssen Techniken zur Anfragespezifikation eingesetzt werden, die es gestatten, eine Anwendung ohne Kenntnisse über formale Spezifikationstechniken zu bedienen. Ein Ziel ist es daher, Interaktionsformen zu finden, bei denen möglichst wenig spezifiziert werden muß. Die klassischen Techniken der Anfragespezifikation verwenden u.a. natürlichsprachliche Schlüsselworte und Kategorie-Bezeichner aus einem vorgegebenen Vorrat. Bei Mehrfachnennung von verschiedenen Schlüsselworten oder Kategoriebezeichnern ist die *Vereinigung* der Anfrageergebnisse die standardmäßige Verknüpfungstechnik in Web-Anwendungen.

Es wird also deutlich, daß in den meisten Anwendungen eher zuviel Information präsentiert wird. Der Benutzer hat Schwierigkeiten sich zurechtzufinden, kann aber aufgrund mangelnder Erfahrung auch nicht eine komplexe Anfragesprache erlernen, die in einem solchen Falle eventuell Abhilfe schaffen könnte. Ein weiteres Problem in diesem Zusammenhang ist auch, daß zur Präzisierung von Anfragen auch ein Kenntnis der Datenmodellierung notwendig scheint (vgl. die „Erklärungs- bzw. Verfeinerungsangaben“ des Altavista-Informationsdienstes). Für die breite Masse der Anwender scheinen diese Techniken zu komplex zu sein.

Die derzeit am Markt vorherrschenden Informationssysteme haben vielfach den Nachteil, daß der regelmäßige Benutzer vom System nicht in adäquater Weise unterstützt wird: Es verhält sich immer wie beim ersten Anmelden des Benutzers oder der Benutzer muß von sich aus oder nach Aufforderung spezielle Angaben machen. Die klassischen Techniken der Fragebogenausfüllung haben sich nach unseren Kenntnissen bei den allermeisten Benutzergruppen allerdings als wenig beliebt herausgestellt. Bei dieser Art der Gewinnung von Informationen über die Interessen der Benutzer fühlen sich viele Benutzer „ausgefragt“ und verweigern die Kooperation. Anwender haben weiterhin die Aufgabe, sich zuerst mit den Fragebogenmechanismen des Systems vertraut zu machen, bevor sie es effektiv zu nutzen verstehen. Nach einigen Versuchen wird die umständliche Handhabung dieser sogenannten Fragebogen-orientierten Informationserhebung zu einer wirklichen Hürde.

Aus der Analyse der Defizite von bestehenden Web-Informationssystemen ergeben sich folgende Ziele. Im Projekt BAND sollen die grundlegenden Techniken untersucht und entwickelt werden, mit denen es möglich ist, die notwendige Informationsfilterangaben durch Beispiele von relevanten Objekten zu bestimmen. Grundlage für die Beschreibung der Beispiele sollen die von dem Benutzer manipulierten Objekte der Anwendung sein. Wir gehen von der These aus, daß der Benutzer konkrete Objekte eher benennen kann als abstrakte Spezifikationen. Eine Menge von Objekten der Anwendung soll als Menge von Beispielen für zu findende Objekte dienen, d.h. zu den Beispielen sollen automatisch passende Anwendungsobjekte ermittelt und entsprechend dargestellt werden. Zu beachten ist dabei, daß eine Informationsfilterung abhängig von der letztendlich gewählten speziellen Darstellungsform generell auf mehrere Arten erfolgen kann. Es kann zum einen absolut gefiltert werden, d.h. herausgefilterte Informationen werden dem Benutzer vorenthalten. Zum anderen kann durch Techniken der graphischen Hervorhebung die Aufmerksamkeit des Benutzers gezielt auf bestimmte Elemente eines durch klassische Techniken der Informationsrecherche bestimmten Grundbestandes von Informationseinheiten gelenkt werden. Diese Art der perzeptiven Filterung wurde im Projekt BAND insbesondere untersucht. Das Projektziel von BAND war die Entwicklung der theoretischen und praktischen Grundlagen für diese spezielle Art der fallbasierten Informationsrecherche.

Im folgenden beschreiben wir eine Leitanwendung, die in besonderem Maße die innovativen Anwendungen charakterisiert, die durch die im Projekt BAND erarbeiteten Lösungen für die oben skizzierte Problemstellung ermöglicht werden.

2 Die Leitanwendung TV-Assistent

Im Projekt BAND wurde eine Leitanwendung gewählt, in der bislang vielfach die oben genannten klassischen Techniken der Informationsrecherche zur Anwendung kommen: ein TV-Auskunftssystem, auch TV-Assistent genannt. Wenn man sich oberflächlich mit der Materie beschäftigt hat, könnte man zu dem Schluß kommen, daß es schon marktreife Systeme gibt, die sich in der Praxis bewähren. Gerade bei den in diesem Kontext eingesetzten Systemen ist jedoch die zum Teil willkürliche Schlüsselwort- und meist komplizierte Fragebogentechnik zur Informationsfilterung stark verbreitet.

Für die zu entwickelnden Techniken des beispielorientierten Informationszugriffs ließen sich weitere Anwendungskontexte finden. Es sind allerdings gerade die großen TV-Zeitschriftenverlage, dicht gefolgt von den Werbe- und PR-Agenturen von dem Sog der „Neuen Medien“ im Kontext des Webs erfaßt und in dieser Thematik aktiv. Hauptmotivation für diese Unternehmen ist es in erster Linie, Präsenz zu zeigen und mit dem Medium Erfahrung zu gewinnen. In zweiter Hinsicht ist die Kundenbindung und der Imagegewinn Antrieb für den Betrieb solcher Server. Langfristig werden sich die interaktiven Medien ihren Stamplatz neben den klassischen „unidirektionalen“ Medien (Print, TV) vor allem durch den gesteigerten Nutzwert sichern. Diese Überlegungen führen zu einer Orientierung der Objekte als „Marke“: Der Kunde soll mit dem „guten Namen“ des Objektes hochwertige Information verbinden. Aus diesem Grund wurde das Projekt in dem Bereich TV angesiedelt. Außerdem besteht in diesem Bereich ein harter Wettbewerb. Innovative Ideen, die mehr Kunden anziehen, werden in diesem Bereich eher akzeptiert. Die Entwicklung der bestehenden Informationssysteme kann als auf einem lokalen Maximum befindlich bewertet werden. Die Inhalte werden derzeit nur wenig flexibel präsentiert: Nutzer können nach einem ihnen bekannten Titel suchen, sie können nach Sendungstypen suchen (Spielfilm, Dokumentation, News etc.) oder Filme können nach Anfangszeit aufgelistet werden. Über gefundene Sendungen können ggf. weitere Informationen durch einen einfachen Klick abgerufen werden. Damit sind die Möglichkeiten bereits erschöpft.

Am Beispiel eines Anfragesystems zu TV-Programmen verdeutlicht sich leicht, daß mittels einer einfachen Abfrage (z.B. „Welche Sendungen laufen zur Zeit?“) die Aufnahmefähigkeit des Nutzers weit überschritten werden kann. Soll die Ausgabe des Systems eingeschränkt werden, so kann ein strukturiertes Datenbanksystem nur die Parameter heranziehen, die von vornherein explizit vorgesehen wurden. Die klassischen Auswahlparameter (z.B. Uhrzeit, Dauer, Sendungstyp, Format, Mitwirkende etc.) sind im Alltag der Konsumenten kaum ausreichend, um benutzeradaptiv relevante Sendungen besonders herauszufiltern. Es mußte also ein Informationssystem geschaffen werden, das in der Lage ist, sowohl inhaltliche als auch benutzerabhängige Parameter bei der Auswahl von Information und bei der Auswahl der passenden Präsentationsform in Bezug auf eine absolute oder perzeptive Filterung hinzuzuziehen. In dem Projekt BAND wurden die Grundlagen erarbeitet, mit denen es möglich ist, Informationssysteme dieser Art mit weniger Aufwand zu entwickeln.

Alle großen Programmzeitschriften haben das WWW im Internet für sich als Instrument entdeckt, junges, kaufkräftiges Publikum mit ihrem Objekt bekannt zu machen bzw. an ihr Objekt zu binden. Die Qualität der Präsentation in optischer Hinsicht spielt dabei eine große Rolle. Eine Studie der existierenden Auskunftssysteme verdeutlichte, daß in einem speziellen graphischen Kontext bei vielen Auskunftssystemen Sendungen in Listenform oder in speziell gestalteten „Programmtabellen“ angeboten werden. Die Gestaltung der Programmtabellen ist dabei eng an die Gestaltung des jeweils zugeordneten Print-Mediums gebunden. Obwohl aus Sicht der Standardisierung von Bedieneinheiten zur Erleichterung der Computerbenutzung sicher zweifelhaft, werden z.B. Designer sogar mit der Gestaltung von speziell gestalteten Schaltflächen (buttons) usw. beauftragt. Im Sinne der Schaffung einer „Corporate Identity“ für das jeweilige Informationssystem als Produkt wird die Motivation für dieses Vorgehen deutlich. Ein Auskunftssystem mit erweiterten Möglichkeiten muß sich in diesen Kontext einfügen. In diesem Zusammenhang ist zu beachten, daß unter dem Schlagwort „benutzeradaptive Darstellung“ nicht die automatische Berechnung von neuen Darstellungsformen an sich verstanden werden soll. Diese Idee ist bei der speziellen, firmenspezifischen Gestaltung von Programmtabellen nur in dem Maße anwendbar als daß in vorbereiteten graphischen Musterdokumenten (templates) Lücken gefüllt werden können. Vielmehr wird deutlich, daß – wie oben schon angedeutet – besonders im Kontext einer speziellen Programmtabellendarstellung eine perzeptive Informationsfilterung durch Hervorhebungen von Filmen, die für den jeweiligen Benutzer interessant sind, erweiterte Möglichkeiten bietet, die auch im praktischen Umfeld akzeptiert werden.

3 Geleistete Arbeiten und erzielte Ergebnisse

Damit die hier skizzierte Kernidee der fallbasierten Informationsrecherche in einer speziellen Leitanwendung in die Tat umgesetzt werden konnte, wurden zu folgenden Problempunkten Lösungen erarbeitet:

- Adäquate Beschreibungs- und Modellierungsformen für Anwendungsobjekte,
- Abstraktionsoperatoren, die es gestatten, aus einer Menge von Beispielen das Gemeinsame zu ermitteln,
- Automatische Modellakquisition zur Reduzierung der Kosten für die mit der BAND-Technologie zusätzlich möglichen Dienste.

Die benutzeradaptive Informationspräsentation durch perzeptive Filterung, die durch die Ergebnisse bei den oben aufgeführten Punkten möglich wird, läßt sich am besten durch Betrachtung einer beispielhaften Benutzung des erstellten Prototypen verdeutlichen.

3.1 Beispielorientierte perzeptive Informationsfilterung im Rahmen des TV-Assistenten

In Abbildung 1 wird ein Schnappschuß des im Projekt erstellten TV-Assistenten gezeigt. Mit den im linken Teil des Fensters gezeigten Interaktionswerkzeugen kann der Benutzer

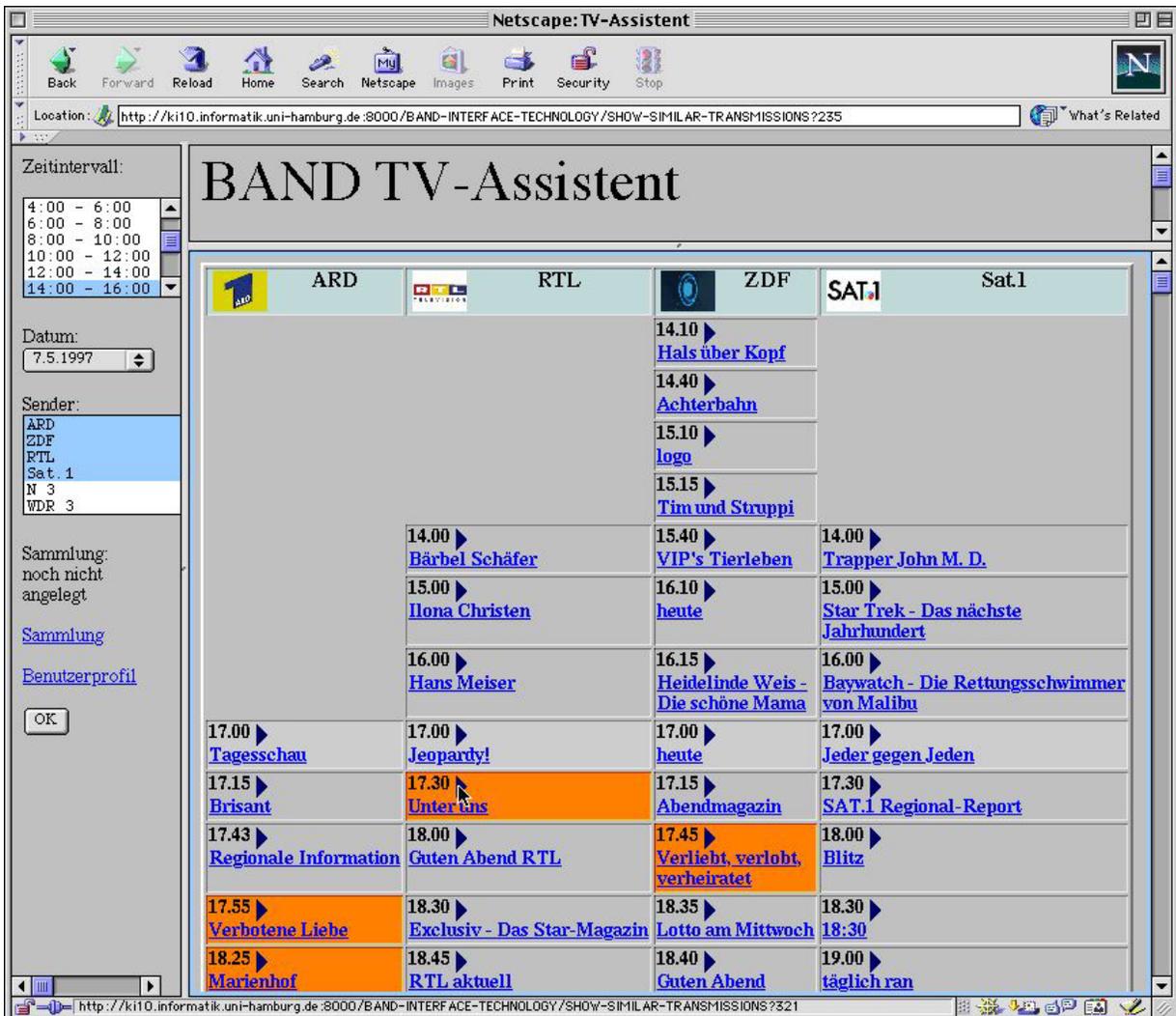


Abb. 1. BAND TV-Assistent mit Hervorhebungen von ähnlichen Sendungen zu „Unter uns“ (RTL, 17:30).

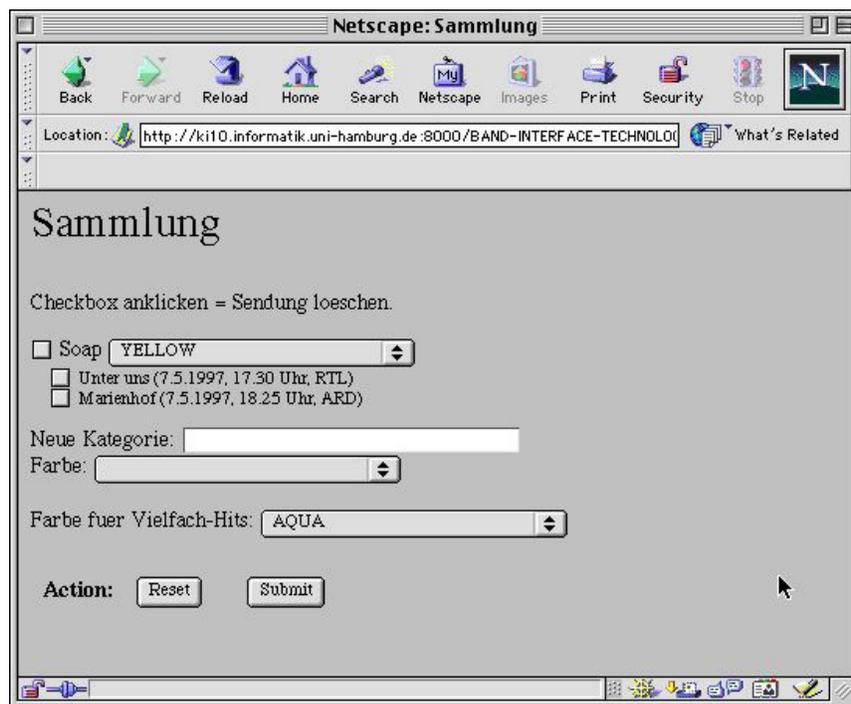


Abb. 2. Fenster zur Verwaltung einer Sammlung von Sendungen.

die Uhrzeit, den Tag und die ihn interessierenden Sender wählen. Loggt er sich ein, erhält er nach Kennworteingabe für die Darstellung der Programmtabelle adäquate Standardwerte (Intervall um die aktuelle Uhrzeit am aktuellen Tag) für die Darstellungsparameter der Programmtabelle (siehe Abbildung 1 für ein Beispiel). Der Benutzer kann den jeder Sendung zugeordneten Pfeil anklicken und erhält eine Hervorhebung der zu der angeklickten Sendung ähnlichen Sendungen im gleichen Ausschnitt der Programmtabelle. Im Beispiel in Abbildung 1 nehmen wir an, der Benutzer habe den Pfeil zu der Sendung „Unter uns“ (RTL, 17:30) angeklickt. Die in der Abbildung gezeigten Hervorhebungen sind plausibel.

In einem separaten Interaktionsfenster kann der Benutzer eine Sammlung von Sendungen verwalten. In dem Beispiel, das in Abbildung 2 gezeigt wird, nehmen wir an, ein Benutzer habe die in Abbildung 2 markierten Sendungen „Marienhof“ und „Unter uns“ zu einer Gruppe „Soap“ zusammengefaßt. Er kann dieser Gruppe eine Farbe zur Hervorhebung in der Programmtabelle zuordnen. Die Gruppe kann nun in jeder gezeigten Programmtabelle zu einer Suche nach ähnlichen Sendungen herangezogen werden. In Abbildung 3 wird die Gruppe „Soap“ aus der Sammlung in der linken Spalte angezeigt. Ein Klick auf den Namen der Gruppe sorgt dafür, daß in der Programmtabelle ähnliche Sendungen durch Techniken der farblichen Hervorhebung herausgefiltert werden (perzeptive Filterung). Das Ergebnis zeigt wiederum die Praxistauglichkeit des im Prototypen implementierten Ansatzes.

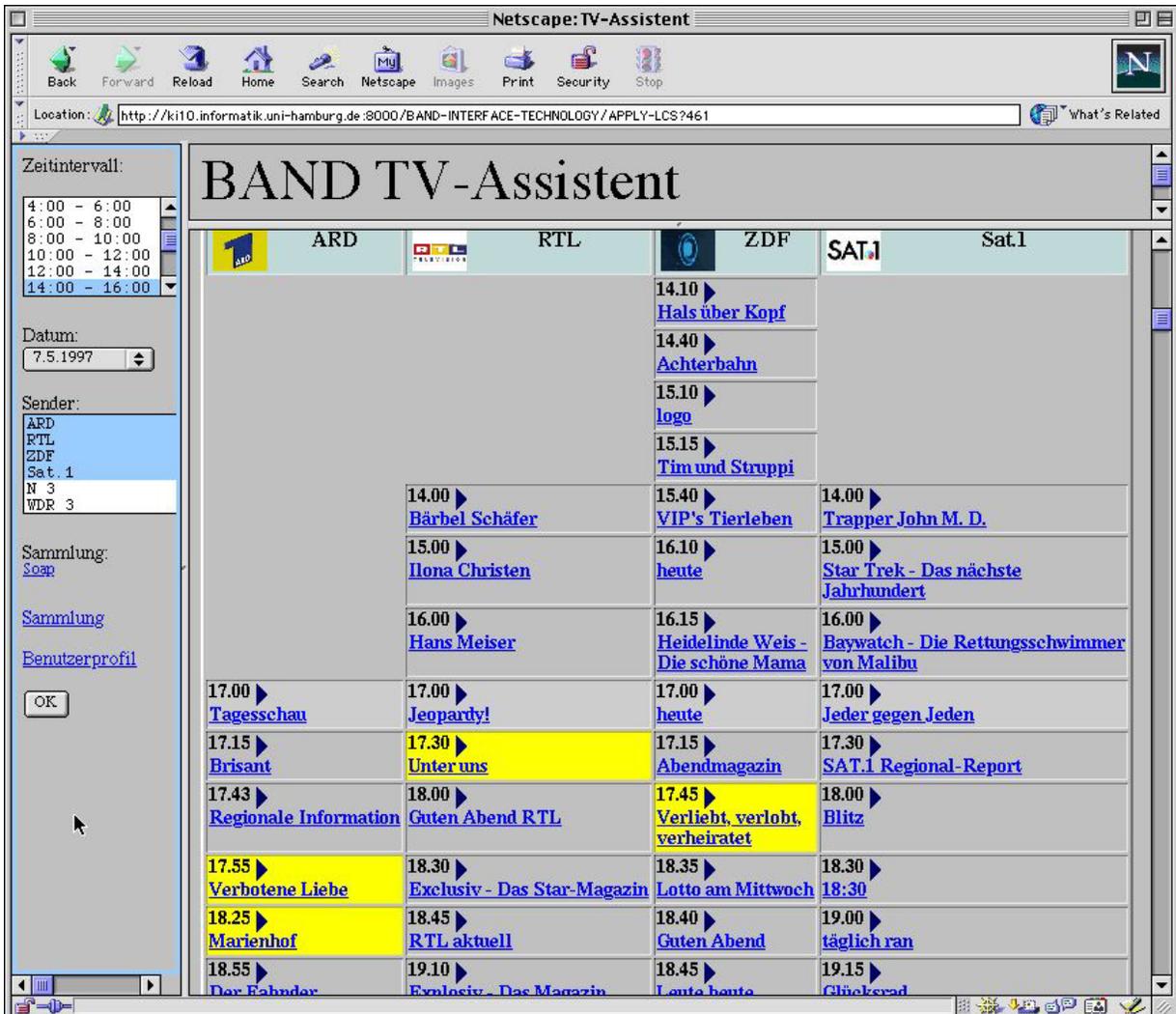


Abb. 3. Hervorhebung auf Basis einer Gruppe von Sendungen aus der angelegten Sammlung.

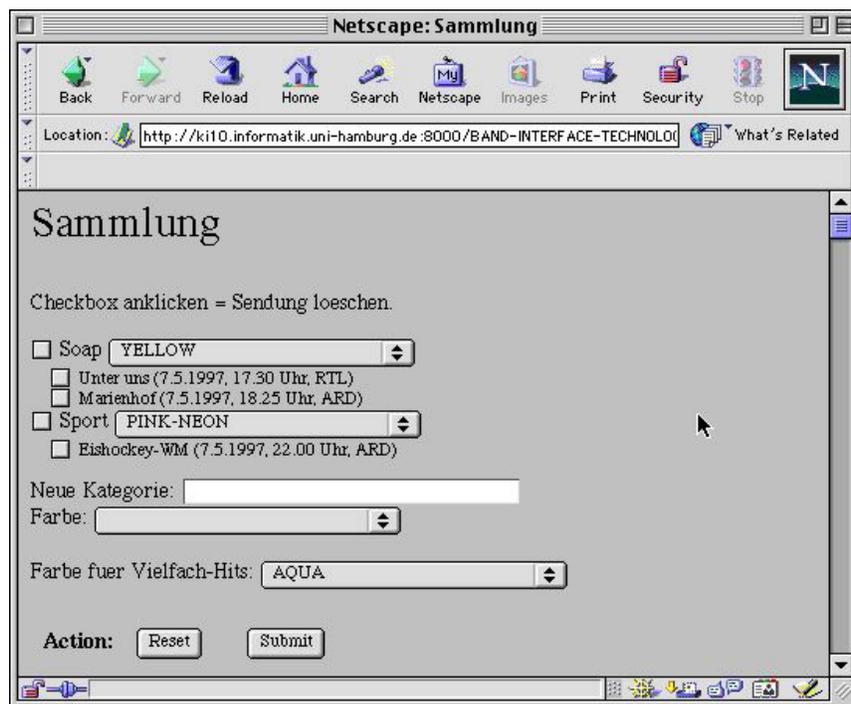


Abb. 4. Einrichtung einer weiteren Gruppe für Sportsendungen.

Abbildung 4 zeigt eine erweiterte Sammlung. Es wurde eine Gruppe „Sport“ eingeführt. In diese Gruppe wurde eine Sendung „Eishockey-WM“ (ARD, 22.00 Uhr) eingefügt.

Da der Benutzer nun schon zwei Gruppen angelegt hat, stellt sich die Frage, in welcher Weise er bei Einfügung einer neuen Sendung in die Sammlung bei der Auswahl einer Gruppe unterstützt wird. In Abbildung 5 wird gezeigt, welche Sendebeiträge vergleichbar mit der Eishockeysendung aus Abbildung 4 sind. Nehmen wir an, der Benutzer möchte die hervorgehobene Sendung „täglich ran“ zu der Sammlung hinzufügen. Im Fenster für die Sammlungen wird durch graphische Anzeichen (+) und (-) bei den Gruppennamen dem Benutzer vermittelt, daß der Sendebeitrag „täglich ran“ zu „Eishockey“ paßt und weniger gut zu „Soap“. Der helle Querstreifen in den Abbildungen 3 und 5 um 19.00 Uhr symbolisiert – wie bei Programmzeitschriften üblich – eine Uhrzeit zur Ausrichtung der Darstellung von Sendungen. Für den Benutzer wird damit ein Vergleich des Programmes auf verschiedenen Sendern erleichtert.

Zur Illustration der Verallgemeinerung von Beispielen wird ein weiteres Szenario betrachtet. In Abbildung 6 wird eine wiederum erweiterte Sammlung aufgezeigt. Die nachfolgende Abbildung 7 zeigt, wie der BAND TV-Assistent die Krimisendungen „Derrick“ und „Ein Mord für Quandt“ in einer Programmtabelle zur Hervorhebung von „Der Fahnder“ und „Küstenwache“ verwendet wird.

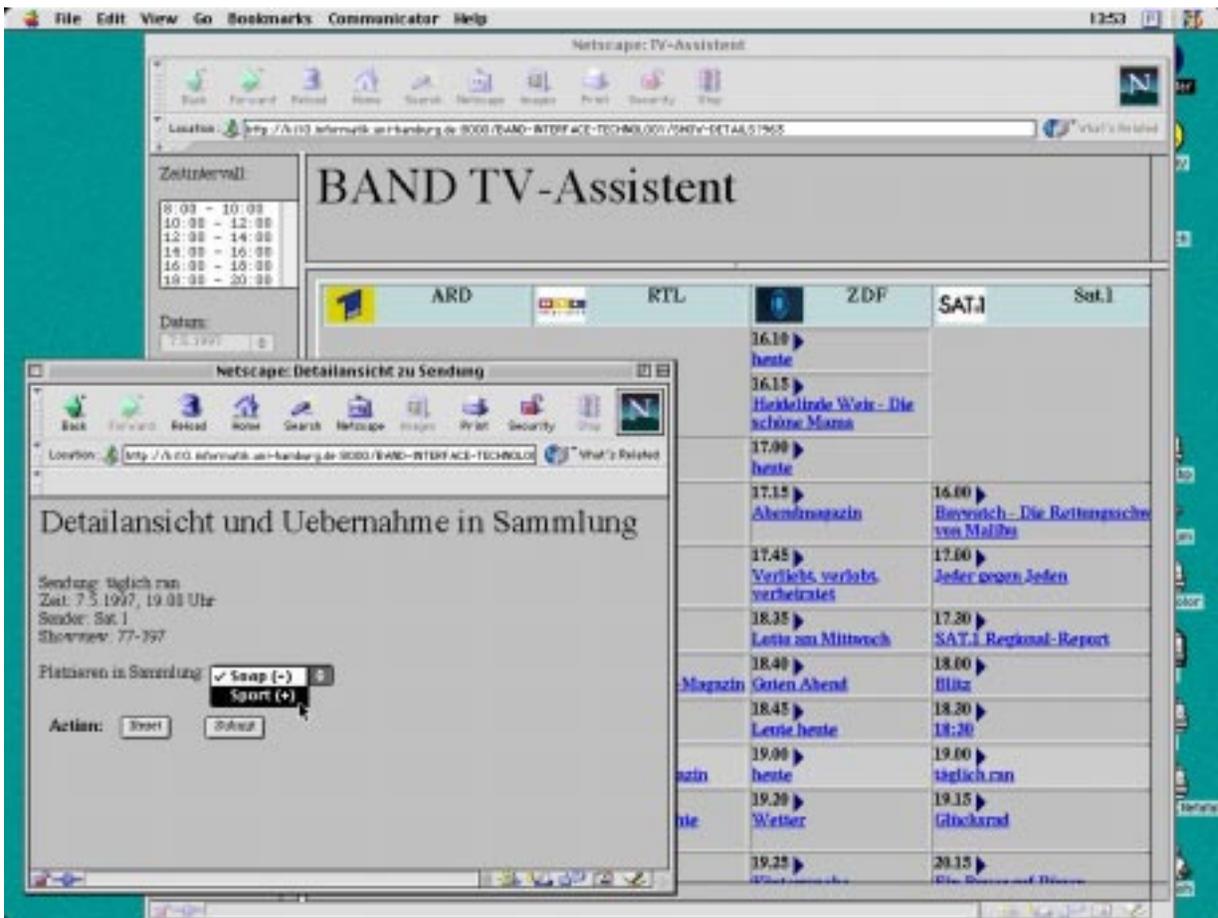


Abb. 5. Ein weiteres Beispiel für die Hervorhebung von Filmen (siehe Text).

Da sich jeder Benutzer eine eigene Sammlung anlegen kann, die durch einen Server-Rechner verwaltet wird, und die Präsentation durch die Sammlungen gesteuert wird (im Prototypen durch Hervorhebung), sprechen wir von einer benutzeradaptiven Darstellung. Jedoch wird durch das Konzept von BAND nicht nur eine perzeptive Filterung durch Hervorhebungen unterstützt. Beim Anmelden beim TV-Assistenten könnte auf einer Vorschaltseite auch eine Liste der Filme zusammengestellt werden, die aufgrund der vorhandenen Sammlungen als besonders interessant für den jeweiligen Benutzer bewertet werden. Für kommerzielle Anbieter mag sich hier eine Möglichkeit zur Plazierung von Werbeflächen usw. bieten.

Die Einrichtung von Sammlungen könnte als ähnlich aufwendig, wie die Ausfüllung von Fragebögen bewertet werden. Wie sich jedoch auch in anderen Anwendungen gezeigt hat, ist der Umgang mit den Objekten der Anwendung für die meisten Benutzer selbstverständlich. Ein Beispiel ist ein Informationssystem über Musik-CDs. Benutzer können sich auch hier Sammlungen anlegen und mit anderen Benutzern über die entsprechenden



Abb. 6. Darstellung der erweiterten Sammlung.

Musikstücke diskutieren. Im Rahmen von BAND kann sich der Benutzer z.B. umfangreiche Zusatzinformationen zu den Sendungen ansehen (siehe z.B. Abbildung 10). Mit Hilfe einer Sammlung kann er im Sinne von „Bookmarks“ auf diese Zusatzinformationen jederzeit zugreifen. Es wird also deutlich, daß der Aufwand zum Einrichten und Verwalten der Sammlung sich für den Benutzer aus verschiedenen Gründen rentiert.

3.2 Integration von Standard-Techniken zum Information-Retrieval

Die Auswahlelemente auf der linken Seite der BAND-Seite mit der Programmtabelle dienen zur Spezifikation einer Anfrage. Die ermittelten Programmdateien werden im Hauptfenster in der Programmtabelle dargestellt. Dieser Grunddatenbestand wird mit Standard-techniken der Informationsrecherche bestimmt. Hierzu gehören z.B. Indizierungstechniken (R-Bäume) zur schnellen Eingrenzung der Menge von Sendungen in einem bestimmten Zeitintervall. Für eine Programmwoche werden ca. 10000 Sendungen mit den entsprechenden Hintergrundinformationen (in Form von Zeichenketten) verwaltet. Die Sendungen enthalten neben den zeitlichen Informationen noch Zusatzinformationen zum Inhalt, zu den



Abb. 7. Hervorhebung von Krimisendungen in einer Programmtabelle.

Schauspielern usw. Für die Generierung der Hervorhebungen sind jedoch die inhaltlichen Beschreibungen der Sendbeiträge maßgebend.

3.3 Computerinterne Beschreibungsformen für Anwendungsobjekte

In diesem Abschnitt skizzieren wir informell kurz die inhaltliche Beschreibung der Anwendungsobjekte, ohne jedoch auf die theoretischen Grundlagen genauer einzugehen. Die inhaltliche Beschreibung von Sendungen stellt eine Charakterisierung der Sendungen mit Hilfe des sog. begrifflichen „Hintergrundwissens“ dar. Grundlage für die Beschreibung der Anwendungsobjekte ist ein logisches Repräsentationssystem. Durch sog. Konzept- und Relationenbeschreibungsterme können Objekte und deren Beziehungen in der realen Welt computerintern beschrieben werden. Inferenzverfahren, die auf diesen Beschreibungsstrukturen operieren, ermöglichen die Realisierung der Anwendungsdienste (s.u.).

Im Prototypen wurde ein sehr einfaches Grundmodell automatisch aus einer Datenbasis von ca. 10000 Sendungen für eine Programmwoche erstellt. Die Grundlage hierfür sind die in den bereitstehenden Testdaten enthaltenen rudimentären Kategorisierungen. Es hat sich herausgestellt, daß in der Anwendung mit diesem automatisch gewonnenen Grundmodell schon recht eindrucksvolle Resultate erzielt werden können. Bei einer Ergänzung dieser Modelle durch weitere Beschreibungen der Sendungen (durch manuell erstellte Spezifikationen oder durch Integration weiterer automatisch ausgewerteter Datenquellen) ließe sich in vielen Fällen die Leistung des TV-Assistenten noch weiter verbessern. Die manuelle Bereitstellung von Modellen ist jedoch relativ aufwendig und damit kostenintensiv. Das Projekt BAND hat gezeigt, daß in dem Anwendungskontext aber auch mit automatischen Verfahren gearbeitet werden kann. Die Verarbeitungsalgorithmen für die Beschreibungsterme der Sendebeiträge sind schon jetzt auf mögliche umfangreiche zukünftige Beschreibungen von Sendungen ausgerichtet.

3.4 Verallgemeinerungsoperatoren und ihr Einsatz im Rahmen der perzeptiven Informationsfilterung

Wir haben in den oben geschilderten Anwendungsbeispielen gesehen, daß der Benutzer Suchanfragen auf der Basis von Gruppen aus seiner Sammlung starten kann. Die Gruppen bestehen aus einzelnen Sendungen. Die jeweiligen Beschreibungsterme für diese Sendungen werden durch Inferenzalgorithmen zu neuen Beschreibungen verarbeitet. Die neuen Beschreibungsterme repräsentieren die Gemeinsamkeiten aus den Beispielsendungen in den Gruppen. Der im BAND-Prototypen verwendete Algorithmus zur Berechnung der Gemeinsamkeiten aus den Eingabebeschreibungen heißt LCS-Algorithmus (least-common subsumer). Mit diesen neuen, ggf. komplexen Beschreibungen können dann die zu präsentierenden Grunddaten gefiltert werden, so daß sich die Menge der „interessanten Sendungen“ ergibt. In den hier gezeigten Abbildungen werden die „interessanten Sendungen“ durch Techniken der Hervorhebung vor einem Grunddatenbestand, der Programmtabelle, angezeigt. Eine absolute Filterung mit einer Auflistung der interessanten Sendungen ist leicht zu realisieren. Denkbar ist auch eine Anwendung zur Suche eines interessanten Ausschnittes der Programmtabelle.

Man beachte, daß es sich bei den berechneten Gemeinsamkeiten je nach Benutzer um völlig neue Beschreibungsformen handelt, die durch einen Modellierer nicht vorher mit einem Namen versehen werden müssen. Dieses ist bei der Vielfalt der möglichen Begriffsbildungen im TV-Bereich auch kaum möglich. Falls sich zu den berechneten Gemeinsamkeiten noch keine Sendungen bestimmen lassen, da eventuell die Gemeinsamkeiten zu speziell sind, erfolgt ein Abgleich mit dem modellierten Hintergrundwissen. Es wird nach einer vordefinierten (benannten) Beschreibung von Sendungen gesucht, wobei die gesuchte, vorbereitete Beschreibung die speziellste Generalisierung der Eingangsterme bezüglich des Hintergrundwissens darstellt. Da der neue, aus dem Hintergrundwissen gewonnene Beschreibungsterm allgemeiner ist als die Eingangsterme, besteht eine größere Möglichkeit,

daß nun Sendungen, die zu der Beschreibung passen, in der aktuellen Programmtabelle gefunden werden können.

Im BAND-Projekt wurden Verallgemeinerungsoperatoren für Beschreibungssprachen verschiedener Ausdrucksstärke untersucht. Schon die einfachen Operatoren bieten in der Anwendung gute Leistungen. Allerdings kann es in einigen Fällen dazu kommen, daß die verwendeten Beispiele wenig Gemeinsamkeiten haben. Anders ausgedrückt, der ermittelte Beschreibungsterm für die Gemeinsamkeiten ist vergleichbar mit dem Begriff „Ding“. Das hat zur Folge, daß in einer Programmtabelle quasi alles hervorgehoben wird, da alle Sendungen als „Ding“ angesehen werden können. In neueren Arbeiten werden mit Blick auf die automatische Akquisition weitere Beschreibungstechniken untersucht, die die Inferenzleistungen des BAND-Systems noch verbessern (z.B. eine probabilistische Modellierung der Überlappung von Begriffen) mit dem Ziel, eine Informationsüberflutung durch zuviele Hervorhebungen zu verhindern.

3.5 Generizität des Ansatzes

Die Grundmechanismen von BAND lassen sich in vielfältiger Weise kombinieren und einsetzen. In den Beispielen aus den vorigen Abschnitten wurden Sendungen zu Gruppen zusammengefaßt, die wiederum in Sammlungen verwaltet werden. Auf Basis dieser Gruppen werden für den Benutzer die relevanten Sendungen in der jeweiligen Programmtabelle hervorgehoben. Nun müssen Gruppen von Filmen nicht nur durch den Benutzer selbst zusammengestellt werden. Ein Redakteur kann z.B. eine Gruppe als „Tip des Tages“ für verschiedene Nutzergruppen zusammenstellen. Der „Tip des Tages“ wird dann auf einer Vorschaltseite den Nutzern angeboten (insbesondere, wenn er zu den vorhandenen Gruppen aus der Sammlung des jeweiligen Benutzers paßt). Ein „Tip des Tages“ kann dann durch einfaches Anklicken als Gruppe in die eigene Sammlung übernommen werden. Falls der Redakteur noch Bilder mit Zusatzinformationen usw. bereitstellt, so stehen diese Informationen sofort für den Benutzer zur Verfügung. Diese Erweiterung verdeutlicht die Generizität des im Projekt BAND verfolgten Ansatzes. Eine weitere Produktidee wäre z.B. ein automatisches System zur benutzeradaptiven Berechnung des „Tips des Tages“. An der Oberfläche können wiederum spezielle Anzeichen für den „Tip des Tages“ verwendet werden (z.B. ein Bild einer Chipstüte eines geeigneten Herstellers mit entsprechender Presenter-Werbung usw.).

In weiteren Arbeiten zum Projekt BAND wird zur Zeit untersucht, in welcher Weise sich Gruppen von Sendungen automatisch bestimmen lassen. Aus der Interaktion des Benutzers mit dem System (Anzeige von Hintergrundinformationen zu Filmen, Schauspielern usw.) lassen sich Rückschlüsse auf Benutzerpräferenzen ziehen. Eine automatische Clusterbildung durch Abstraktionsoperatoren wie z.B. der erwähnte LCS-Algorithmus bietet vielfältige Möglichkeiten.

Ideen für neue Produkte im Rahmen des TV-Assistenten lassen sich viele aufzeigen. Benutzer könnten ihre Sammlungen austauschen und über die Inhalte diskutieren. Vergleichbare Vorschläge wurden in einigen Informationssystemen in etwas anderer Form schon in

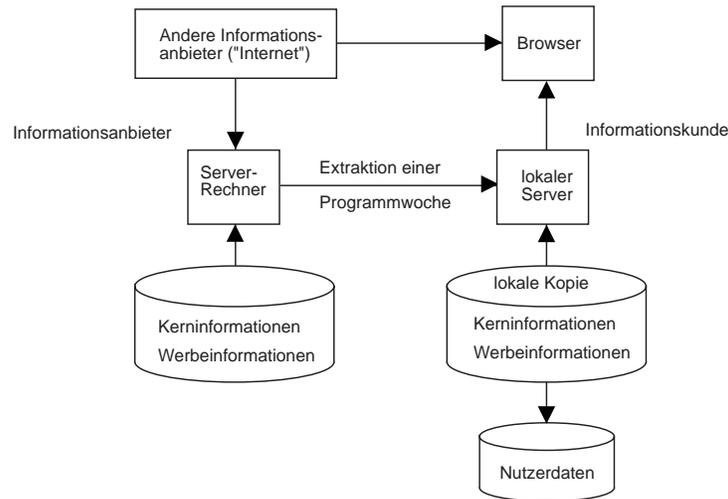


Abb. 8. Skizze für ein Konzept zum Schutz des Benutzers vor Mißbrauch in den benutzerspezifischen Sammlungen enthaltenen Informationen.

die Tat umgesetzt. Die Auswertung von Benutzerinformationen bietet viele Möglichkeiten, allerdings sind auch die möglichen Gefahren nicht außer acht zu lassen. Dieses leitet über zum Thema Datensicherheit.

3.6 Konzept zur Gewährleistung der Datensicherheit

Die Speicherung von benutzerspezifischen Informationen kann eventuell auch zu einem Mißbrauch führen. In diesem Abschnitt zeigen wir ein Szenario auf, bei dem ein Mißbrauch vermieden werden kann. Das Bild 8 zeigt eine mögliche Konfiguration der im BAND-Projekt entwickelten Softwarebausteine, mit der die Sicherheit der benutzerspezifischen Daten gewährleistet werden kann. Zur Benutzung des TV-Assistenten installiert der Benutzer ein geprüftes lokales Serverprogramm, welches in zeitlichen Abschnitten einen (kodierten) Ausschnitt der Programm- und Werbeinformationen über das Internet auf den lokalen Rechner kopiert. Der lokale Server speichert die benutzerspezifischen Informationen auf dem lokalen Rechner (Intranet). Auf den lokalen Server bzw. auf dem lokalen Netzwerk kann der Benutzer dann wie gewohnt mit seinem Internet-Browser zugreifen. Durch eine entsprechende Kodierung der Programminformationen kann der Anbieter sicherstellen, daß der Kunde die Informationen nur im Zusammenhang mit dem TV-Assistenten verwendet.

Ob und wie dieses Konzept in einem Produkt durchgesetzt wird, kann im Kontext dieses Berichts jedoch nicht vertiefend diskutiert werden. Aus Sicht des Informationsanbieters lassen sich bei dem in Abbildung 8 vorgestellten Konzept sicherlich nicht mehr die gewünschten Marktanalysen durchführen. Zu beachten ist auch, daß die schon existierenden Programminformationsdienste in den entsprechenden Logging-Protokollen schon jetzt umfangreiche Daten über die Benutzer speichern. Eine Zuordnung der Internet-Adressen zu



Abb. 9. Detaildarstellung für den Sendebeitrag „Traumschiff“.

Namen und Wohnorten scheint nicht ausgeschlossen. Da dieses Problem jedoch nicht spezifisch für BAND ist, möchten wir im Rahmen dieses Reports auf eine genauere Diskussion verzichten.

3.7 Zeichenkettenbasierte Assoziation von Sendebeiträgen

Zum Vergleich der in BAND neu entwickelten Techniken soll in diesem Abschnitt eine zeichenkettenbasierte Assoziation von Sendebeiträgen aufgezeigt werden. Im aktuellen BAND-Prototypen wurde eine sehr komfortable Version dieser etablierten Suchtechnik implementiert. Neben einer Eingabe von Suchzeichenketten werden automatisch Querverweise erzeugt. Wenn also in einer Detaildarstellung eines Sendebeitrages eine Zeichenkette verwendet wird, die auch in der Detaildarstellung eines anderen Sendebeitrages vorkommt, so ist die Zeichenkette anklickbar. Die Erzeugung der Hypertext-Funktionalität erfolgt automatisch. Abbildung 9 zeigt eine Detaildarstellung der Sendung „Traumschiff“. Durch Anklicken von „Kommentare“ wird eine Zusammenfassung des Inhalts präsentiert (siehe Abbildung 10). Die automatisch erzeugten Hypertextverweise sind durch Unterstreichungen markiert. Nehmen wir an, ein Benutzer klicke auf „Berlin“. In Abbildung 11 wird das Resultat gezeigt: Eine Liste von Sendebeiträgen wird zur Auswahl in einer Liste angezeigt. Wir sehen in Abbildung 12, in welcher Weise das Wort „Berlin“ in dem Film „Die Unschuld vom Lande“ bedeutsam ist.

Wir sehen also, daß der TV-Assistent auch eine zeichenkettenbasierte Assoziation von Sendungen unterstützt. Das Beispiel macht aber auch die Vorteile der inhaltsbasierten As-



Abb. 10. Kommentare zum „Traumschiff“.



Abb. 11. Ergebnis der Zeichenkettensuche nach „Berlin“.

soziation deutlich. In dem Beispiel wäre vermutlich die Auflistung von weiteren Sendungen über Kreuzfahrten für den Benutzer angemessen gewesen.

3.8 Wiederverwendbare generische Softwarebausteine, Unterstützte Plattformen

Neben dem Code zur Generierung von Programmtabellen mit zeitlicher Ausrichtung von Sendebeiträgen verschiedener Sender (siehe die entsprechenden Abbildungen) umfaßt die in BAND erstellte Software noch weitere wiederverwendbare, generische Bausteine. Zur einfachen Manipulation von Anwendungsobjekten in HTML-Dokumenten dient das Modul HTTP-CLIM. Die Berechnung von Gemeinsamkeiten von Konzepttermen wird durch die Programmpakete LCS und PCLASSIC möglich. Diese Programmpakete lassen sich auch in anderen Anwendungskontexten als den TV-Assistenten wiederverwenden.

Die im Projekt BAND erstellte Software läuft auf allen gängigen Plattformen: Windows, MacOS und Unix. Als Internet-Browser werden „Internet Explorer“ und „Netscape“ unterstützt.



Abb. 12. Detaildarstellung eines der Suchergebnisse zur Zeichenkette „Berlin“.

4 Zusammenfassung

Die in diesem Bericht aufgeführten Beispiele belegen, daß mit der im Projekt BAND entwickelten Technologie ein TV-Assistent entwickelt werden konnte, der in seinen Möglichkeiten beträchtlich über das hinausgeht, was bisherige Systeme an Dienstleistungen anbieten können. Die Ausgabe von Programmdateien erfolgt benutzeradaptiv in dem Sinne, daß unter Bezugnahme auf die Sammlung des Benutzers für ihn relevante Sendungen in entsprechender Weise besonders ausgezeichnet werden (z.B. durch Hervorhebungen oder durch spezielle Auflistungen). Im BAND-Prototypen erfolgte die Hervorhebung in einer speziell konstruierten Programmtabelle. Der in dieser Tabelle dargestellte Grunddatenbestand wird in Bezug auf die Datenindizierung mit klassischen Techniken der Informationsrecherche bestimmt. Durch diese nahtlose Integration von bewährten Standarddatenstrukturen und Algorithmen kann eine hohe Performanz garantiert werden.

Für den Endanwender des TV-Assistenten bringen die im Projekt entwickelten wissensbasierte Techniken einen gesteigerten Nutzwert. Nach unserer Kenntnis gibt es zur Zeit kein weiteres (TV-)Auskunftssystem, das ähnliche Funktionalität bietet. Die in diesem Bericht aufgezeigten Beispiele können jedoch offensichtlich nur einen kleinen Eindruck von den Möglichkeiten der im Projekt BAND realisierten beispielorientierten Informationsrecherche geben. In weiteren Arbeiten kann für die Ausgestaltung eines vollständigen Anwendungssystems auch noch die graphische Darstellungsform detaillierter untersucht werden. Der erstellte Prototyp zeigt jedoch zweifelsfrei schon jetzt die erreichten Ergebnisse.

5 Ausblick: Perspektiven für Anwendungen und Einsatz

Zur Anwendung der im Projekt BAND entwickelten Technologien in den Bereichen Knowledge-Engineering, Inferenztechnik, Beschreibungslogiken für eine konkrete Aufgabenstellung für

einen Kunden sind einige weitere Voraussetzungen zu schaffen. Wir betrachten die notwendigen Schritte anhand der Leitdomäne „TV-Assistent“.

5.1 Einbindung in die vorhandene Infrastruktur beim Anwender

Bei allen Überlegungen, das BAND-System in einer realen Umgebung einzusetzen, müssen die entwickelten Technologien in einen infrastrukturellen Zusammenhang gesetzt werden. Ein TV-Assistent muß, um Anwender (Kunden) optimal bei der Planung und Nutzung von TV-Angeboten unterstützen zu können, mit aktuellen Programmdateien über Sendungen versorgt werden. Um diese Daten zur Verfügung zu haben, sind bindende Verfahren nötig, die über den einfachen Import von Textlisten hinausgehen. Im Anwendungsfall innerhalb einer Programmredaktion (z.B. in einem Programmzeitschriftenverlag) ist ein Online-Zugriff auf das Produktionssystem und dessen Datenquellen unabdingbar. Zur Veranschaulichung sei gesagt, daß bei ca. zehntausend Sendetiteln je Programmwoche (bezogen auf das deutsche Bundesgebiet) und bis zu fünfzig Beschreibungselementen je Sendung sowie bei den diversen Programmänderungen und -verschiebungen sich eine große, bei Anfragen zu berücksichtigende Datenflut ergibt. Diese Verschmelzung eines TV-Assistenten mit dem Datenpool eines Anbieters stellt die Grundvoraussetzung für einen Einsatz der BAND-Technologien in dieser Domäne dar. Vor der eigentlichen Inbetriebnahme steht also ein gewisser Aufwand bzgl. der Einbindung in die kundenspezifische EDV-Welt: Das BAND-System muß an die vorhandenen Datenbanken angeschlossen und in den Ablauf der redaktionellen Tätigkeit eingebunden werden; dazu gehört eben auch eine direkte Einflußnahme der Redakteure auf die Klassifikationen, Indizierungen und Auszeichnungen, die im BAND-System als Informationen zur Inferenzbildung herangezogen werden. In dem BAND-Projekt wurde durch die Entwicklung des TV-Assistenten gezeigt, daß die Anbindung an die in der kommerziellen EDV-Welt üblichen Technologien ohne größere Probleme möglich ist.

Die im BAND-System verwendeten Technologien können Ihre gewinnbringenden Qualitäten besonders bei einer längerfristigen und häufigen Benutzung durch Anwender ausspielen. Dazu ist das System nicht nur in einen reibungslosen Ablauf des digitalen Informationsflusses zu integrieren; auch und besonders wichtig ist die Bereitstellung von Bediener-schnittstellen für Administratoren und für redaktionell/inhaltliche Tätigkeiten, wie sie für das Bearbeiten der Wissensbasis sowie der Klassifikationsschemata ständig notwendig sind. Im vorliegenden Prototypen wurden diese Arbeiten durch Softwareexperten geleistet, die Werkzeuge auf Implementationsebene verwenden können. Für eine Überführung in einen Regelbetrieb ist die Entwicklung und Implementierung von entsprechenden Schnittstellen wünschenswert.

Entscheidende Merkmale des BAND-Systems resultieren aus dem intelligenten Erfassen von Eingangsdaten, das über das einfache Speichern von Daten in Tabellen hinausgeht. Dazu muß, um eine Klassifikation von TV-Sendungen zu ermöglichen, eine Modellierung der Ausgangsdomäne bereitgestellt werden. Diese Modellierung ist aufgrund des dynamischen Anwendungsbereichs regelmäßig anzupassen. Dabei muß die Verfügbarkeit von Ex-

perten gewährleistet sein, die die Modellierungen in das System überführen. Eine gewisse Vertrautheit mit den Grundlagen der notwendigen formalen Modellierungstechniken kann von den Mitarbeitern bei Auftraggebern kaum erwartet werden. Darum erscheint eine begleitende Beratung durch die entsprechend ausgebildete Systementwickler auch langfristig unabdingbar.

5.2 Sekundärnutzen

Aufgrund der methodischen Vorgehensweise kann die entwickelte Technologie auch für andere, aber ähnlich gelagerte Anwendungsbereiche verwendet werden. Z.B. werden Fragen der Datenhaltung und der Datenmodellierung aus einem anderen Blickwinkel beleuchtet und können somit für den Auftraggeber interessante Perspektiven für eine weitergehende Nutzung der Datenbasis eröffnen. Im Kontext des BAND-Prototypen, der für die Verwendung im Internet über eine WWW-Schnittstelle entwickelt wurde, könnte sich eine Nutzbarkeit im Zusammenhang mit elektronischen Programmführern, z.B. im digitalen Fernsehen ergeben. Auch hierfür werden die Programminhalte auf vielfältige Weise präsentiert und für den Endbenutzer intuitiv verwendbar aufbereitet. Auch in diesem Falle ist die sorgfältige Klassifikation und Repräsentation von TV-Daten die Voraussetzung für eine erfolgreiche Anwendung. Die Flexibilität des BAND-Rahmenwerkes ermöglicht dann die Unterstützung von neuen Ausgabeformen.

Danksagungen

Die Beteiligten des Projektes BAND bedanken sich bei der Wirtschaftsbehörde der Stadt Hamburg für die Förderung des Projekts im Rahmen des Labors für Künstliche Intelligenz. Weiterhin gebührt den Studierenden Alissa Kaplunova, Irena Kortas und Claas Prien, die im Rahmen von BAND ihre Diplom- bzw. Studienarbeiten angefertigt haben oder noch anfertigen, besonderer Dank für die geleistete Arbeit. Für die Bereitstellung von Programmdateien möchten wir dem Heinrich Bauer Verlag, Hamburg ebenfalls unseren Dank aussprechen. Die Entwicklung des Prototypen wäre in der kurzen Zeit kaum möglich gewesen ohne die Verwendung des Systems CL-HTTP, entwickelt und bereitgestellt von John Mallery, Artificial Intelligence Laboratory, Massachusetts Institute of Technology. CL-HTTP umfaßt die Software für einen Webserver sowie auch die Textindizierungssoftware „Lambda-Vista“ für die zeichenkettenbasierte Informationsassoziation. Die inhaltliche Suche nach relevanten Sendungen basiert im aktuellen Prototypen auf dem Wissensrepräsentations- und Inferenzsystem CLASSIC, welches uns von den AT&T Bell Research Laboratories für Forschungszwecke zur Verfügung gestellt wurde.

Veröffentlichungen im Berichtszeitraum

- Möller, R., Haarslev, V., Neumann, B., Semantics-Based Information Retrieval, In: Proc. IT&KNOWS-98: International Conference on Information Technology and Knowledge Systems, 31. August-4. September, Vienna, Budapest, 1998, pp. 49-61.

- Mantay, T., Möller R., Content-Based Information Retrieval by Computation of Least Common Subsumers in a Probabilistic Description Logic , In: Proc. International Workshop on Intelligent Information Integration, ECAI'98, Aug. 23-28, Brighton, UK, erschienen als TZI-Report No. 10, Technologie-Zentrum Informatik Univ. Bremen, 1998, pp. 141-156.
- Mantay, T., Computing Least Common Subsumers in Expressive Description Logics. Erscheint in: Proc. of DL99, International Workshop on Description Logics, Linköping, 1999.