

2009

EINSATZ VON RECOMMENDER- SYSTEMEN ZUR PERSONALISIERTEN INFORMATIONSVERSORGUNG IM STANDBERICHTSWESEN VON DATA- WAREHOUSE-SYSTEMEN

Carsten Jürck

Otto-Friedrich-Universität Bamberg

Tim-Oliver Förtsch

Otto-Friedrich-Universität Bamberg

Bernd U. Jahn

Otto-Friedrich-Universität Bamberg

Achim Ulbrich-vom Ende

Otto-Friedrich-Universität Bamberg

Follow this and additional works at: <http://aisel.aisnet.org/wi2009>

Recommended Citation

Jürck, Carsten; Förtsch, Tim-Oliver; Jahn, Bernd U.; and Ende, Achim Ulbrich-vom, "EINSATZ VON RECOMMENDER-SYSTEMEN ZUR PERSONALISIERTEN INFORMATIONSVERSORGUNG IM STANDBERICHTSWESEN VON DATA-WAREHOUSE-SYSTEMEN" (2009). *Wirtschaftsinformatik Proceedings 2009*. 114.

<http://aisel.aisnet.org/wi2009/114>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik Proceedings 2009 by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

EINSATZ VON RECOMMENDER-SYSTEMEN ZUR PERSONALISIERTEN INFORMATIONSVERSORGUNG IM STANDARDBERICHTSWESEN VON DATA-WAREHOUSE-SYSTEMEN

Carsten Jürck, Tim-Oliver Förtsch, Bernd U. Jahn,
Achim Ulbrich-vom Ende¹

Kurzfassung

Entscheidungsträger stehen vor der Herausforderung, bei wachsendem Informationsangebot und unter zunehmendem Zeitdruck komplexe Entscheidungen treffen zu müssen. Vor diesem Hintergrund haben sich bei wiederkehrenden Informationsnachfragen Standardberichtssysteme als Hilfsmittel zur Informationsversorgung etabliert. Da mittlerweile auch Standardberichtssysteme stark an Umfang gewonnen haben und kontinuierlichen Änderungen unterworfen sind, wird es für den Entscheidungsträger immer schwieriger, die für ihn geeigneten Berichte aufzufinden. In diesem Beitrag wird untersucht, wie die aus dem eCommerce-Umfeld bekannten Konzepte von Recommender-Systemen im Standardberichtswesen anwendbar sind, um Entscheidungsträger bei der personalisierten Informationsversorgung zu unterstützen. Ausgehend von theoretischen Untersuchungen wird abschließend die Integration eines kollaborativen Recommender-Systems in ein bestehendes Data-Warehouse-System vorgestellt.

1. Motivation und Zielsetzung

Eine adäquate Informationsversorgung ist für die Qualität von Managemententscheidungen von wesentlicher Bedeutung [2]. Das stetig wachsende Informationsangebot in den Unternehmen führt jedoch zum Problem der Informationsüberflutung. Die gezielte Selektion von Informationen für Entscheidungsprozesse wird mehr und mehr zur Herausforderung. Führungsinformationssysteme stellen Anwendungssysteme dar, welche die Unternehmensführung mit entscheidungsrelevanten Informationen versorgen. Als Datenbasis von Führungsinformationssystemen haben sich in den letzten Jahren Data-Warehouse-Systeme (DWH-Systeme) etabliert. Sie bieten Entscheidungsträgern einen zentralen Zugriff auf konsolidierte und historisierte, an Entscheidungsobjekten ausgerichtete Daten aus unternehmensinternen und -externen Datenquellen [7]. Dazu werden die relevanten Daten aus den Datenquellen extrahiert, aufbereitet und in vereinheitlichter Form in einer separaten Datenbasis, dem so genannten Data-Warehouse, gespeichert. Die Datenbereitstellung in Form multidimensionaler Datenstrukturen sowie spezielle Analyseoperatoren gestatten im Rahmen des Online-Analytical-Processings (OLAP) eine interaktive Datenanalyse und sollen dem Nutzer einen intuitiven Zugang zu den Daten im Data-Warehouse ermöglichen.

¹ Otto-Friedrich-Universität Bamberg, D-96045 Bamberg, Feldkirchenstraße 21.

Bei der Nutzung von DWH-Systemen, insbesondere bei der Nutzung von OLAP-Servern, lassen sich in der Praxis u. a. folgende Probleme identifizieren. Einerseits benötigen Entscheidungsträger für die adäquate Formulierung ihres Informationsbedarfs umfangreiche Kenntnisse über das zugrundeliegende Datenschema und die Semantik der verfügbaren Daten im Data-Warehouse. Andererseits muss der Nutzer für eine gezielte Informationsrecherche mit der Bedienung der jeweiligen Analyseanwendung vertraut sein. Nur unter diesen Voraussetzungen ist es einem Entscheidungsträger möglich, den persönlichen Informationsbedarf zu decken. Diese Voraussetzungen sind nicht immer gegeben, da die Nutzung des Systems oft in unregelmäßigen Abständen und unter Zeitdruck erfolgt. Dies hat zur Folge, dass die selbstständige Informationsversorgung durch das DWH-System mit einem Einarbeitungsaufwand verbunden ist, der in vielen Fällen den wahrgenommenen Nutzen übersteigt. Daher verwenden viele Unternehmen Standardberichtssysteme, um die wiederkehrende Informationsnachfragen für unterschiedliche Anwendungsgebiete zentral zur Verfügung zu stellen. Standardberichte fördern eine kontinuierliche und gemeinsame Ausrichtung aller Entscheidungsträger auf die wesentlichen Unternehmensziele. Ihr formaler Aufbau ist durchgängig gleich, die inhaltliche Ausgestaltung jedoch auf die spezifischen Informationsbedarfe der Nutzer zugeschnitten [9]. Da Informationsanforderungen der Entscheidungsträger jedoch in der Regel a priori nicht vollständig erfasst werden können, muss das Standardberichtssystem im Laufe des Systembetriebs an veränderte Unternehmensziele und Entscheidungssituationen angepasst werden [11]. Hieraus ergeben sich stetiges Wachstum und Wandel des Standardberichtssystems, was wiederum zu einer Informationsüberflutung der Nutzer führen kann. Im Rahmen des Projektes CEUS², welches den Hintergrund des vorliegenden Beitrags bildet, konnten zudem zwei weitere Problembereiche eines Standardberichtssystems identifiziert werden. Zum einen müssen die Nutzer für den effizienten Einsatz des Standardberichtwesens einen vollständigen Überblick über die vorhandenen Berichte haben. Diese Forderung ist in der Regel unrealistisch, da die konkrete Ausgestaltung eines Standardberichtssystems schnell mehrere hundert Berichte umfassen kann und diese gleichzeitig einer kontinuierlichen Anpassung unterliegen. Zum anderen liegt eine weitere Problematik im Auffinden potentiell interessanter Berichtsinhalte nach der Aktualisierung der Datenbasis im Data-Warehouse. Dabei ist es einem Nutzer nicht zuzumuten, alle für ihn potentiell relevanten Berichte auszuführen und deren Inhalte zu analysieren.

Ziel des vorliegenden Beitrags ist es daher, die personalisierte Informationsversorgung in OLAP-basierten Standardberichtssystemen durch den Einsatz von Recommender-Systemen zu verbessern. Hierzu werden im folgenden Abschnitt etablierte Ansätze zur automatisierten und personalisierten Informationsversorgung im DWH-Umfeld vorgestellt. Abschnitt 3 erläutert die grundlegenden Konzepte und Realisierungsansätze von Recommender-Systemen. Darauf aufbauend wird untersucht, welche der Realisierungsansätze auf das Standardberichtswesen eines Data-Warehouse-Systems übertragbar sind. Abschnitt 4 stellt eine prototypische Realisierung eines Recommender-Systems zur personalisierten Empfehlung von Standardberichten vor. Auf der Grundlage von ähnlichen Nutzerprofilen werden dem Entscheidungsträger potentiell interessante Berichte automatisiert vorgeschlagen, die sich jedoch thematisch von den bisher genutzten Berichten unterscheiden und somit einen Neuigkeitswert besitzen. Ein Fazit und ein Ausblick bilden in Abschnitt 5 den Abschluss dieses Beitrags.

2. Personalisierte Informationsversorgung in Data-Warehouse-Systemen

Der Informationsbedarf von Entscheidungsträgern ist sehr unterschiedlich ausgeprägt, d. h. jeder Entscheider benötigt personalisierte Informationen, um seine Entscheidungen gezielt treffen zu können. Eine *personalisierte Informationsversorgung* passt die Bereitstellung der Informationen an

² Computerbasiertes Entscheidungsunterstützungssystem für die Hochschulen in Bayern (<http://www.ceushb.de>).

die jeweiligen Bedürfnisse des Entscheiders an. Diese Informationsbereitstellung kann durch ein DWH-System aktiv oder passiv erfolgen, was häufig mit Push- bzw. Pull-Mechanismen gleichgesetzt wird. Klassische DWH-Systeme mit einer passiven Datenbereitstellung per Pull-Mechanismus sind nicht dafür ausgelegt, um den Entscheidungsträger personalisierte Informationen *automatisiert* zur Verfügung zu stellen [4]. Hierzu sind Erweiterungen um „aktive“ Systemkomponenten erforderlich. Informationen werden durch das DWH-System zur Verfügung gestellt, nachdem der Nutzer einen geeigneten Push-Mechanismus definiert hat. Aufgrund eines eintretenden Ereignisses werden hierbei vordefinierte Operationen ausgelöst, ohne dass dafür eine zeitgleiche Nutzerinteraktion erforderlich ist. Die aktive Informationsbereitstellung per Push-Mechanismus kann im DWH-Umfeld u. a. durch das Event-Condition-Action-Modell oder den Benachrichtigungsdienst realisiert werden, welche im Folgenden zur Abgrenzung des in diesem Beitrag vorgestellten Ansatzes kurz skizziert werden.

Das *Event-Condition-Action-Modell* (ECA-Modell) stellt ein Konzept dar, um Auswertungen oder Analysen (teil-)automatisiert durchführen zu können. Es findet beispielsweise in aktiven Datenbanken Verwendung. Dieses Modell beinhaltet Regeln, die aus dem Tripel (Event, Condition, Action) bestehen und auf der Idee basieren, dass bei Eintritt eines Ereignisses (Event) eine vordefinierte Bedingung (Condition) überprüft und falls diese Bedingung erfüllt ist, eine entsprechende Aktion (Action) ausgeführt wird. Als Erweiterung zum klassischen ECA-Modell werden im Active-Data-Warehouse-Konzept³ komplexere Bedingungen verwendet, da sie sich auf multidimensionale Analysen im Data-Warehouse beziehen [19]. Ein Active-Data-Warehouse-System (ADWH-System) verwendet hierfür Analyseregeln, mit denen Aktionen ereignisgesteuert ausgelöst werden können. Der zentrale Punkt eines ADWH-Systems ist die Ereignissteuerung in der Analyseregeln [4], die Entscheidungsprobleme automatisiert erkennt und somit einen Anstoß durch den Entscheidungsträger nicht mehr erforderlich macht [19]. Somit lassen sich, entsprechend den definierten Regeln, beim Eintritt eines bestimmten Ereignisses korrespondierende Daten im Data-Warehouse analysieren. Nach einem Abgleich mit den definierten Bedingungen wird automatisch eine Entscheidung über eine Aktion getroffen. Das ADWH-System übernimmt somit die permanente Analyse der Daten und entlastet hierdurch den Entscheidungsträger von diesen Aufgaben.

Das Konzept des *Benachrichtigungsdienstes* stellt einem Entscheidungsträger ebenfalls automatisiert personalisierte Informationen per Push-Mechanismus bereit. Entscheidungsträger erstellen hierbei ein Abonnement für eine bestimmte Art von Ereignis. Statt dem Ereignis kann auch ein Zeitpunkt bzw. Zeitplan oder ein Schwellwert definiert werden. Sobald ein auslösendes Ereignis auftritt, wird automatisiert eine Benachrichtigung an den Abonnenten versendet, ohne dass dafür eine erneute Nutzerinteraktion erforderlich ist. Bei den generierten Benachrichtigungen handelt es sich damit um personalisierte Informationen, welche die Präferenzen des Abonnenten wiedergeben.

3. Einsatz von Recommender-Systemen in Data-Warehouse-Systemen

Die aktive Informationsbereitstellung per Push-Mechanismus durch die beiden im vorherigen Abschnitt dargestellten Ansätze unterstützt zwar die automatisierte und personalisierte Informationsversorgung eines Entscheidungsträgers, jedoch muss der Entscheider seinen Informationsbedarf im Vorfeld explizit formuliert haben. Recommender-Systeme beheben diese Einschränkung, indem sie auf unterschiedliche Weise die Anforderungen eines Nutzers erfassen, ihm Vorschläge bzw. Empfehlungen unterbreiten und diese seinen veränderten Präferenzen anpassen. Im Folgenden werden zunächst Recommender-Systeme im Allgemeinen charakterisiert,

³ Eine detaillierte Beschreibung des Active-Data-Warehouse-Konzepts findet sich beispielsweise bei [19].

um anschließend mögliche Ansätze zur Nutzung von Recommender-Systemen zur personalisierten Informationsversorgung im Standardberichtswesen eines DWH-Systems zu präsentieren⁴.

3.1. Grundlagen von Recommender-Systemen

Recommender-Systeme⁵ haben sich in den letzten Jahren als wichtige Bestandteile u. a. im eCommerce etabliert, indem sie potenziellen Käufern Kaufempfehlungen geben (vgl. beispielsweise das Produktempfehlungssystem von Amazon)⁶. Sie schlagen Nutzern basierend auf deren konkreten Bedürfnissen und Interessen Produkte oder Dienstleistungen vor. Nach [3] lassen sich im Allgemeinen die drei Systemkomponenten *Hintergrundinformationen* („background data“), *Nutzerinformationen* („input data“) sowie der *Informationsverarbeitungsalgorithmus* („algorithm“) unterscheiden. Im Rahmen der Aufgabendurchführung „Empfehlung generieren“ werden Nutzerinformationen mithilfe eines Informationsverarbeitungsalgorithmus auf Basis bestehender Hintergrundinformationen in Empfehlungen, beispielsweise für Produkte, umgewandelt. Diese Grundstruktur ist in *Abbildung 1* dargestellt.

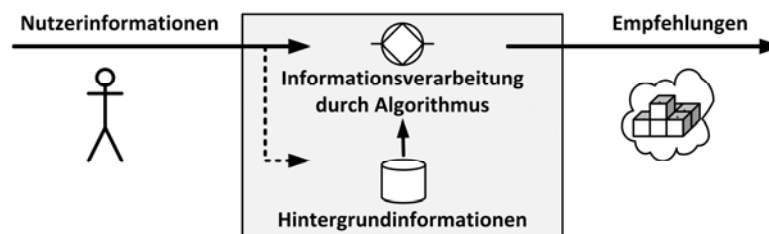


Abbildung 1: Allgemeine Struktur einer Empfehlungsaufgabe

Empfehlungen haben einen hohen Nutzwert, wenn sie sich am Informationsstand des individuellen Entscheidungsträgers orientieren. Dies impliziert, dass Empfehlungen geeignet sind, den Informationsstand des Individuums zu erweitern und das Individuum angeregt wird, die Informationen auch nachzufragen. Unpersonalisierte Empfehlungen, die bestehende Informationen substituieren (z. B. Top N verkaufter Produkte) sind dafür ebenso wenig geeignet wie Empfehlungen für Objekte, die das Individuum ohnehin nachfragen würde (vgl. [6], [1]). Zusammenfassend kann ein Recommender-System als Anwendungssystem definiert werden, welches automatisiert Nutzerinformationen auf Basis bestehender Hintergrundinformationen in personalisierte Empfehlungen überführt.

Hinsichtlich der Realisierung von Recommender-Systemen lassen sich im Allgemeinen zwei Hauptaktivitäten unterscheiden: Die *Generierung von Nutzerprofilen* und die *Verwertung der Profile* [12]. Nutzerprofile dienen der Abgrenzung des subjektiven Informationsbedarfs eines Nutzers. Die Erstellung der Profile erfolgt anhand bestehender Nutzerinformationen. Dabei handelt es sich typischerweise um Präferenzen, Wünsche, Kenntnisse, Fähigkeiten oder Interessen eines Individuums. Bei der Verwertung von Nutzerprofilen steht der Abgleich des Informationsangebots und des subjektiven Informationsbedarfs im Vordergrund. Die beiden Hauptaktivitäten müssen aufeinander abgestimmt werden, da sich Abhängigkeiten zwischen der Art der Profilgenerierung und der Profilverwertung ergeben. [3] unterscheidet daher anhand der benötigten Inputdaten (zur Profilgenerierung) und der Art der Informationsverarbeitung (zur Profilverwertung) fünf Ansätze von Recommender-Systemen.

⁴ Ein Ansatz zur Unterstützung der Navigation in OLAP-Systemen durch Recommender-Systeme findet sich in [5].

⁵ Es existieren in der Literatur heterogene Begriffsbestimmungen und Klassifikationen (vgl. u. a. [15], [18]).

⁶ Eine Übersicht bisheriger Anwendungsdomänen für Recommender-Systeme findet sich bei [12].

- Der als *kollaboratives Filtern* bezeichnete Ansatz gilt als die am häufigsten verwendete Realisierungsalternative und nutzt Ähnlichkeiten zwischen Nutzern für die Generierung von Empfehlungen. Im Rahmen der Profilgenerierung werden Relevanzbewertungen erfasst, die Nutzer für Objekte abgeben. Relevanzbewertungen einer Nutzergemeinschaft fungieren als Hintergrundinformationen. Relevanzbewertungen des einzelnen Anwenders dienen als Nutzerinformationen und bilden dessen Nutzerprofil. Im Rahmen der Profilverwertung werden Ähnlichkeiten zwischen den Bewertungsmustern des Anwenders und der Nutzergemeinschaft gesucht. Werden Individuen der Nutzergemeinschaft identifiziert, die Objekte sehr ähnlich bewertet haben wie der Anwender, so werden deren Relevanzbewertungen für vom Anwender noch nicht bewertete Objekte auf diesen übertragen. Sind darunter Objekte mit hoher Relevanzbewertung werden diese dem Anwender empfohlen.
- Beim Ansatz des *inhaltsbasierten Filterns* werden Eigenschaften von Objekten bewertet und verglichen. Als Nutzerinformationen zur Profilgenerierung werden Relevanzbewertungen des Anwenders hinsichtlich der Eigenschaften von bekannten oder verwendeten Objekten erfasst. Als Hintergrundinformationen stehen die anhand ihrer tatsächlichen Eigenschaften bewerteten Objekte des Informationsangebots zur Verfügung. Im Rahmen des Informationsverarbeitungsprozesses werden diejenigen Objekte ermittelt und empfohlen, deren Eigenschaften den vom Anwender als relevant eingestuften Eigenschaften am ähnlichsten sind.
- Der Ansatz des *demographischen Filterns* basiert auf demographischen Ähnlichkeiten zwischen Nutzern, z. B. dessen Alter, Familienstand, Bildungsgrad. Während des Informationsverarbeitungsprozesses werden Individuen der Nutzergemeinschaft ermittelt, die hinsichtlich der demographischen Merkmale eine hohe Ähnlichkeit mit dem Anwender aufweisen. Die Präferenzen der so ermittelten Individuen werden auf den Anwender übertragen und für Empfehlungen verwendet.
- Das *nutzenbasierte Filtern* empfiehlt anhand der individuellen Nutzenfunktion des Anwenders das Objekt mit dem für ihn höchsten Nutzen. Da als Nutzerinformation eine individuelle Nutzenfunktion vorausgesetzt wird, die im vorliegenden Kontext nicht bekannt ist und auch nicht (ohne intensive Beanspruchung der Anwender) ermittelt werden kann, wird dieser Ansatz nicht näher charakterisiert.
- Der Ansatz des *wissensbasierten Filterns* sieht vor, dass auf Basis des individuellen Bedarfs des Nutzers und seiner Präferenzen auf zu empfehlende Objekte geschlossen werden kann. Alle bisher verfügbaren Systeme verlangen, dass der Anwender seinen Bedarf bezüglich einzelner Objekteigenschaften manuell erfasst. Dies widerspricht dem Formalziel einer hohen Automatisierung. Daher wird auch dieser Ansatz nicht näher beschrieben.

3.2. Personalisierte Empfehlung von Standardberichten

Nach der Einführung der unterschiedlichen Recommender-Systemtypen wird im Folgenden deren Einsatz für eine automatisierte und personalisierte Empfehlung von Standardberichten in DWH-Systemen untersucht. Hierzu werden zunächst verfügbare OLAP-Metadaten identifiziert, die potentiell als Input für Recommender-Systeme zur Verfügung stehen. Diese lassen sich entsprechend ihrer Bezugsobjekte folgendermaßen kategorisieren:

- *Nutzerinformationen*⁷:
Hierzu gehören alle Stammdaten des Nutzers eines OLAP-Systems z. B. Name, organisatorische Zugehörigkeit, Alter, Aufgabengebiete. Zudem werden die Berechtigungen eines Nutzers, die ihm Rahmen des Berechtigungskonzeptes zugeordnet werden, hierunter erfasst. Das

⁷ Dieser Begriff ist nicht mit der Bedeutung von „input data“ (vgl. Abschnitt 3.1) gleichzusetzen.

Berechtigungskonzept ordnet jedem Nutzer eine Menge von Zugriffsrechten auf Objekte des OLAP-Systems zu.

- *Berichtsinformationen:*

Diese beschreiben die Struktur eines Berichtes in Form der enthaltenen Dimensionen und Kennzahlen, sowie die Selektionsbedingungen eines Berichtes in Form eines Filters. Da die Ordnerstruktur in einem Standardberichtssystem häufig einer inhaltlichen Systematik unterliegt, kann der logische Speicherort eines Berichtes ebenfalls dieser Kategorie zugeordnet werden.

- *Nutzungsinformationen:*

Sie reflektieren die Nutzung der einzelnen Standardberichte durch die jeweiligen Nutzer. Die Nutzungsinformationen umfassen dabei alle Daten, die im Rahmen der Nutzung des OLAP-Servers protokolliert werden können. In vielen Fällen wird damit erfasst, welcher Nutzer welchen Bericht zu welcher Zeit ausgeführt hat. Des Weiteren kann die Interaktion eines Nutzers mit einem Bericht in diese Kategorie eingeordnet werden. Die Interaktionen umfassen z. B. die Drill-Funktionen, den Export, die Verweildauer oder die Pivotierung eines Berichtes.

Anhand der soeben identifizierten Informationsarten wird nachfolgend aufgezeigt, welche der eingeführten Realisierungsalternativen von Recommender-Systemen für die personalisierte Empfehlung von Standardberichten geeignet sind. Die Ansätze des *nutzenbasierten* und des *wissensbasierten Filterns* setzen Nutzerinformationen voraus, die in OLAP-Systemen im Allgemeinen nur schwer ermittelt werden können. Obgleich dies die Anwendung dieser Ansätze nicht ausschließt, werden sie jedoch im Folgenden aus diesem Grund nicht weiter behandelt.

Beim Ansatz des *kollaborativen Filterns* müssen Relevanzbewertungen des Nutzers für Standardberichte herangezogen werden, um Nutzerprofile erstellen zu können. Relevanzbewertungen lassen sich auf Basis der in OLAP-Systemen enthaltenen Nutzungsinformationen erzeugen. Im einfachsten Fall wird die Nutzungshäufigkeit als Relevanzbewertung angesehen („Wie oft hat Anwender X den Standardbericht Y aufgerufen?“). Diese kann um die Nutzungsintensität erweitert werden („Welche Interaktionen führte der Anwender X mit dem Standardbericht Y aus?“). Unabhängig von der Implementierung des Recommender-Systems ist festzuhalten, dass ein OLAP-System über geeignete Metadaten verfügt, um Profile für kollaboratives Filtern zu generieren. Analog verhält es sich mit den notwendigen Hintergrundinformationen: Relevanzbewertungen der Nutzergemeinschaft können auf dieselbe Art und Weise ermittelt werden wie die Bewertungen des einzelnen Anwenders. Hinsichtlich des Informationsverarbeitungsprozesses sind sehr unterschiedliche Ausgestaltungen denkbar (vgl. z. B. [16]). Geeignete Algorithmen finden sich beispielsweise bei [6] und [1].

Soll der Ansatz des *inhaltsbasierten Filterns* verfolgt werden, müssen Relevanzbewertungen des Anwenders hinsichtlich der Eigenschaften von Standardberichten vorhanden sein, um eine Profilgenerierung durchzuführen. Als Eigenschaften von Standardberichten können die in den Berichtsinformationen enthaltenen Strukturmerkmale angesehen werden. Die Bewertung der Eigenschaften erfolgt durch die Nutzung des jeweiligen Berichtes, welche in den Nutzungsinformationen protokolliert sind. Ruft ein Anwender A beispielsweise mehrere Berichte auf, welche die Dimension *Hochschulstandort* enthalten und setzt er hierbei jeweils den Filter „*Hochschulstandort* = *Wien*“, so ist davon auszugehen, dass für Anwender A Standardberichte eine hohe Relevanz aufweisen, die Informationen über Wiener Hochschulen beinhalten. Die Menge aller zur Verfügung stehenden Standardberichte kann ebenfalls anhand der Berichtsinformationen bezüglich ihrer tatsächlichen Eigenschaften bewertet werden. Somit sind die notwendigen Hintergrundinformationen für inhaltsbasiertes Filtern ebenfalls verfügbar. Die Ausgestaltung des Informationsverarbeitungsprozesses und dabei einzusetzende Algorithmen sind beispielsweise bei [1] beschrieben.

Wird der Ansatz des *demographischen Filterns* verfolgt, müssen demographische Daten über den Anwender zur Profilgenerierung zur Verfügung stehen. In OLAP-Systemen ist dies durch Nutzerinformationen der Fall, die demographische Stammdaten des Anwenders beinhalten. Ebenso stehen die demographischen Merkmale der Nutzergemeinschaft als Hintergrundinformationen zur Verfügung. Geht beispielsweise aus den Stammdaten des Anwenders A hervor, dass dieser in der Personalabteilung der Hochschule B arbeitet und sein Aufgabengebiet die Professorenbesoldung ist, können aus der Nutzergemeinschaft weitere Individuen ermittelt werden, die dasselbe Aufgabengebiet an anderen Hochschulen haben und deren Nutzungsinformationen herangezogen werden, um Empfehlungen für Anwender A zu generieren. Eine Implementierung des Informationsverarbeitungsprozesses kann mit unterschiedlichen Algorithmen erfolgen. Hierzu wird auf [10] und [13] verwiesen.

Zusammenfassend ist festzuhalten, dass sich die Ansätze des kollaborativen Filterns, des inhaltsbasierten Filterns und des demographischen Filterns für den Einsatz in einem OLAP-basierten Standardberichtssystem anbieten. Im nächsten Abschnitt wird vorgestellt, wie ein auf dem kollaborativen Filtern beruhendes Recommender-System in eine bestehende DWH-Architektur eingebunden werden kann.

4. Prototypische Realisierung eines Recommender-Systems zur personalisierten Empfehlung von Standardberichten

Ausgehend von den theoretischen Überlegungen in den vorangegangenen Abschnitten wird nachfolgend die Integration eines kollaborativen Recommender-Systems in ein bestehendes DWH-System vorgestellt. Als Anwendungsbeispiel und empirische Grundlage dient das seit 2005 betriebene DWH-System CEUS. Das System umfasst derzeit ca. 1.500 Standardberichte und wird von über 500 Personen genutzt. Während des Betriebs wurde deutlich, dass sich die meisten Entscheidungsträger keinen vollständigen Überblick über das Standardberichtswesen verschaffen können. Die prototypische Realisierung schlägt daher auf Basis von ähnlichen Nutzerprofilen Entscheidungsträgern potentiell interessante Berichte automatisiert vor. Die generierten Berichtsempfehlungen auf Basis des kollaborativen Filterns bieten personalisierte Informationen, die sich den veränderten Präferenzen der Entscheidungsträger anpassen. Nachfolgend wird zunächst die Architektur des Gesamtsystems und daran anschließend der Prozess der Empfehlungsgenerierung vorgestellt.

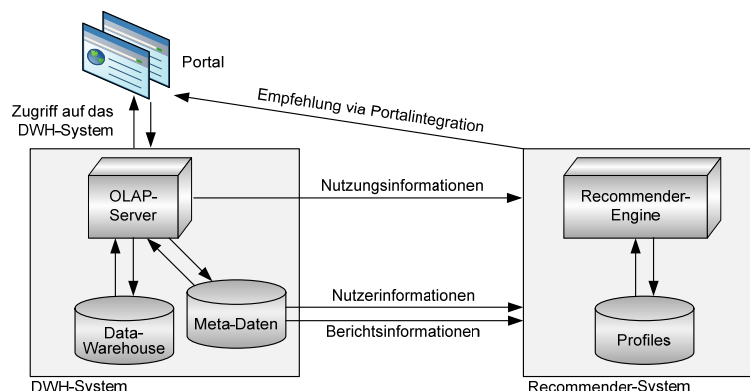


Abbildung 2: Architektur eines um ein Recommender-System erweitertes DWH-System

Die Architektur des Gesamtsystems (vgl. *Abbildung. 2*) umfasst ein DWH-System, welches die multi-dimensional aufbereiteten Daten als Grundlage für das Standardberichtssystem des OLAP-Servers zur Verfügung stellt. Der OLAP-Server übernimmt die Aufgabe, Berichtsaufrufe der Nutzer zu

verarbeiten und die entsprechenden Daten vom Data Warehouse abzurufen. Der eingesetzte OLAP-Server verfügt über ein Metadaten-Repository mit Berichtsinformationen, Nutzerinformationen und Nutzungsinformationen. Diese Informationen dienen als Inputdaten für das Recommender-System. Das Recommender-System wiederum umfasst einen Speicher für Berichts- und Nutzerprofile sowie eine Recommender-Engine. Ausgehend von den Daten des OLAP-Servers generiert die Recommender-Engine die spezifischen Profile für die einzelnen Nutzer. Auf Grundlage der gespeicherten Nutzerprofile berechnet die Recommender-Engine zur Laufzeit Empfehlungen für einzelne Entscheidungsträger. Das konzipierte Recommender-System sieht vor, die Profile der einzelnen Nutzer implizit, d. h. ohne Interaktion mit dem Nutzer, aus den Nutzungsdaten des Standardberichtssystems zu erstellen. Im vorliegenden Anwendungsbeispiel ist das Recommender-System eine eigene Systemkomponente, die unabhängig vom DWH-System fungiert.

Ausgehend von der dargestellten Architektur wird der Einsatz des Recommender-Systems zur Empfehlung von Standardberichten aufgezeigt. Hierbei können zwei Phasen unterschieden werden:

- *Profilgenerierung:*
Ausgangspunkt für die Profilgenerierung ist die Annahme, dass die Gesamtheit der bisher vom Entscheidungsträger ausgeführten Berichte dessen Interessensgebiete widerspiegelt. Das jeweilige Nutzerprofil besteht somit aus einer Bewertung der ausgeführten Standardberichte hinsichtlich ihrer Aufrufhäufigkeit. Notwendige Voraussetzung hierfür ist der persönliche Login zum System⁸. Die Inputdaten können über eine geeignete Schnittstelle des OLAP-Servers aus dem Metadaten-Repository extrahiert werden. Im Anwendungsbeispiel steht ein umfangreiches Protokoll der Nutzungsinformationen der letzten zwei Jahre als Datengrundlage für die Profilgenerierung zur Verfügung. Um im Rahmen der Berichtsempfehlung für einen Nutzer Berichte von anderen Nutzern mit ähnlichen Interessensgebieten empfehlen zu können, ist es notwendig, zunächst auf Basis der einzelnen Profile entsprechende Nutzergruppen zu bilden. Die Berechnung der Nutzergruppen erfolgt über Distanz- oder Ähnlichkeitsmaße. Im Anwendungsbeispiel wird das *Clustering* der unterschiedlichen Profile über das Kosinusmaß⁹ vorgenommen. Der Output der Profilgenerierung sind Gruppen von Nutzern, die sich in der Nutzung des Standardberichtssystems ähnlich sind. Generell muss bei der Profilgenerierung mit dem Problem des so genannten „Kaltstarts“ umgegangen werden. Neue Entscheidungsträger, für die noch keine Bewertungen ermittelt wurden, lassen sich nicht sinnvoll in Nutzergruppen einordnen. Aus diesem Grund werden nur Nutzer in die Clusterbildung aufgenommen, die eine minimale Anzahl¹⁰ von Berichten ausgeführt haben¹¹.
- *Berichtsempfehlung:*
Auf Grundlage dieser Hintergrundinformationen kann die Recommender-Engine zur Laufzeit individuelle Berichtsempfehlungen für spezifische Nutzer generieren. Zu diesem Zweck wird das Profil des Nutzers mit seiner Nutzergruppe verglichen. Dabei werden die Berichte identifiziert, die der Nutzer bisher noch nicht ausgeführt hat, aber gleichzeitig innerhalb seiner Nutzergruppe im Sinne der Häufigkeit der Berichtsausführung besonders „interessant“ sind. Im dargestellten Anwendungsfall wurde die entsprechende Funktionalität in SQL implementiert, um eine einfache Integration in die bestehende Infrastruktur zu ermöglichen. Die generierten Berichtsempfehlungen stehen den Nutzern über die Integration

⁸ Bei funktionsorientierten, von mehreren Personen verwendeten Logins ist dies nur auf Umwegen möglich.

⁹ Eine Darstellung anderer möglicher Distanz- und Ähnlichkeitsmaße findet sich z. B. bei [17].

¹⁰ Die Anzahl der notwendigen Bewertungen für sinnvolle Empfehlungen muss im Einzelfall entschieden werden.

¹¹ Dieses Problem kann mit Hilfe von hybriden Recommender-Ansätzen minimiert werden (vgl. z. B. [3]).

in das bestehende Web-Portal des Standardberichtssystems zur Verfügung. Hierbei werden den Entscheidungsträgern die vier Berichte mit der höchsten Relevanzbewertung angezeigt (vgl. *Abbildung 3*).



Abbildung 3: Web-Portal mit generierten persönlichen Berichtsempfehlungen

Das aufgezeigte Vorgehen ermöglicht es, die Entscheidungsträger bei der Nutzung des Standardberichtssystems zu unterstützen, indem Berichte vorgeschlagen werden, die in der Interessengruppe des jeweiligen Nutzers eine hohe Relevanzbewertung erhalten haben. Hierdurch kann die Inter-subjektivierung des subjektiven Informationsbedarfs von Entscheidungsträgern gefördert werden.

Das zweite angesprochene Problem, der Identifizierung von Berichtsinhalten mit Neuigkeitswert nach einer Aktualisierung des Data-Warehouse, kann durch die Erweiterung der aufgezeigten Lösung um einen verhaltensorientierten kollaborativen Ansatz gelöst werden (vgl. [8], [14]). Hierbei wird zusätzlich die Nutzerinteraktion als Teil der Nutzungsinformationen für das Recommender-System verwendet. Eine positive Relevanzbewertung kann aus einer intensiven Interaktion mit einem Bericht gefolgert werden. Zur Berücksichtigung neuer Daten nach der Aktualisierung des Data-Warehouse wird davon ausgegangen, dass sich die Intensität der Interaktion mit dem entsprechenden Bericht verstärkt, falls der Bericht Informationen mit Neuigkeitswert liefert. Andernfalls würde sich die Intensität der Interaktion in der Regel nicht verändern. Durch den Vergleich der Nutzungsintensität kann dem Entscheidungsträger der spezifische Inhalt (Neuigkeitswert) eines Standardberichts empfohlen werden, falls dieser bereits für den Nutzer als relevant eingestuft wurde. Aus Platzgründen wird in diesem Beitrag auf eine detaillierte Darstellung dieser Erweiterung verzichtet.

5. Fazit und Ausblick

Die Unterstützung von Entscheidungsträgern im Umgang mit einem wachsenden Informationsangebot gewinnt mehr und mehr an Bedeutung. In diesem Beitrag wurde der Einsatz von Recommender-Systemen als Unterstützung beim Umgang mit umfangreichen Standardberichtssystemen erörtert. Es wurde die prototypische Realisierung eines kollaborativen Recommender-Systems in die bestehende DWH-Systemarchitektur von CEUS aufgezeigt. Die Übertragbarkeit des Ansatzes ist dabei eng an die verfügbaren Metadaten des DWH-Systems geknüpft. Nur wenn diese einen Rückschluss auf die zugrundeliegenden Interessengebiete der Entscheidungsträger ermöglichen, ist ihr Einsatz möglich und sinnvoll. Um die Idee der verhaltensorientierten Ansätze des kollaborativen Filterns für ein Standardberichtssystem realisieren zu können, ist in weiteren Arbeiten zu untersuchen, welche Interaktionsformen eine Aussagekraft bezüglich der Relevanz von Berichtsinhalten ermöglichen und eine entsprechende Bewertung dieser in Bezug auf den subjektiven Neuigkeitswert für einen Entscheidungsträger zulassen. Des Weiteren ist für alle

Recommender-Systemtypen und Informationsarten zu evaluieren, welche Distanz- und Ähnlichkeitsmaße im Rahmen des Standardberichtswesens geeignet sind und zu akzeptierten Empfehlungen führen.

6. Literaturverzeichnis

- [1] ADOMAVICIUS, G., TUZHILIN, A., Toward the Next Generation of Recommender Systems: A Survey of the State-of- the-Art and Possible Extensions, in: IEEE Transactions on Knowledge and Data Engineering 17-6, 2005, S. 734-749.
- [2] BIETHAN, J., MUKSCH, H., RUF, W., Ganzheitliches Informationsmanagement - Band I, München 1990.
- [3] BURKE, R., Hybrid Recommender Systems: Survey and Experiments, in: User Modeling and User-Adapted Interaction 12-4, 2002, S. 331-370.
- [4] GELHOET, M., RIEGER, B., Mehrstufige Entscheidungsunterstützung durch Active Data Warehouses, in: Ferstl, O.K., Sinz, E.J., Eckert, S., Isselhorst, T. (Hrsg.), Wirtschaftsinformatik 2005, Heidelberg 2005, S. 1405-1419.
- [5] GIACOMETTI, A., MARCEL, P., NEGRE, E., A Framework for Recommending OLAP Queries, in: DOLAP '08: Proceeding of the ACM 11th international workshop on Data warehousing and OLAP, Napa Valley 2008, S. 73-80.
- [6] HERLOCKER, J.L., KONSTAN, J.A., TERVEEN, L.G., RIEDL, J.T., Evaluating Collaborative Filtering Recommender Systems, in: ACM Transactions on Information Systems 22-1, 2004, S. 5-53.
- [7] INMON, W.H., Building the Data Warehouse, Indianapolis 2005.
- [8] KIM, J., OARD, D.W., ROMANIK, K., User modelling for Information Filtering Based on Implicit Feedback, in: ISKO-France, France 2001.
- [9] KLIMEK, S., Entwicklung eines Führungsleitstands als Unterstützungssystem für das Management unter besonderer Berücksichtigung des FuE-Bereichs, Göttingen 1998.
- [10] KRULWICH, B., Lifestyle Finder: Intelligent User Profiling Using Large-Scale Demographic Data, in: AI Magazine 18-2, 1997, S. 37-45.
- [11] LEHNER, W., Datenbanktechnologien für Data-Warehouse-Systeme. Konzepte und Methoden, Heidelberg 2003.
- [12] MONTANER, M., LÓPEZ, B., DE LA ROSA, J. L., A Taxonomy of Recommender Agents on the Internet, in: Artificial Intelligence Review 19-4, 2003, S. 285-330.
- [13] NGUYEN, A.-T., DENOS, N., BERRUT, C., Improving New User Recommendations with Rule-based Induction on Cold User Data, in: Proceedings of the ACM Conference on Recommender Systems, New York 2007, S. 121-128.
- [14] OARD, D. W., KIM, J., Modeling information content using observable behaviour, in: Proceedings of the 64 Annual Meeting of the American Society for Information Science and Technology, USA 2001.
- [15] RESNICK, P., VARIAN, H.R., Recommender Systems, in: Communications of the ACM 40-3, 1997, S. 56-58.
- [16] SARWAR, B., KARYPIS, G., KONSTAN, J., RIEDL, J., Item-Based Collaborative Filtering Recommendation Algorithms, in: Proceedings of the 10th International Conference on World Wide Web, New York 2001, S. 285-295.
- [17] SCHMITT, I., Ähnlichkeitssuche in Multimedia-Datenbanken: Retrieval, Suchalgorithmen und Anfragebehandlung, München 2005, S. 215-238.
- [18] TERVEEN, L., HILL, W., Human-Computer Collaboration in Recommender Systems, in: J. Carroll (Hrsg.), Human Computer Interaction in the New Millenium, New York 2001, S. 487-509.
- [19] THALHAMMER, T., Active Data Warehouses: Complementing OLAP with Analysis Rules, Linz 2001.