

Arbeitsberichte der Hochschule für Wirtschaft FHNW – Nr. 12

Profile und Technologien der Personalisierung

Arbeitsbericht Nr. 37 des
Competence Center E-Business Basel

Uwe Leimstoll, Henrik Stormer, Raoul Schneider,
Michael Quade, Michael Pülz

ISSN Nr. 1662-3266 (Print)
 Nr. 1662-3274 (Online)

ISBN Nr. 978-3-03724-105-9

Institut Institut für Wirtschaftsinformatik IWI

Datum Oktober 2008

© 2008 Hochschule für Wirtschaft FHNW und die Autoren. Jede
Reproduktion, auch von Teilen und unabhängig vom Medium, ist
nur mit Genehmigung der Hochschule für Wirtschaft FHNW und
der Autoren gestattet.

Mit freundlicher
Unterstützung von

KTI/CTI

Projektpartner



Profile und Technologien der Personalisierung von

Uwe Leimstoll *), Henrik Stormer **), Raoul Schneider *), Michael Quade *), Michael Pülz *)

*) Institut für Wirtschaftsinformatik IWI – www.fhnw.ch/iwi **) Edorex Informatik AG – www.edorex.ch

Vorwort

Dieser Arbeitsbericht entstand im Rahmen des Forschungsprojekts „Personalisierung von E-Commerce-Applikationen II“ (PersECA II), das am Institut für Wirtschaftsinformatik der Fachhochschule Nordwestschweiz durchgeführt wurde. In den PersECA-Projekten werden in Zusammenarbeit mit mehreren Wirtschaftspartnern Wege erarbeitet, wie E-Commerce-Applikationen zu verbesserten Kundenbeziehungen und zu einem erhöhten Nutzen für die Kunden und das Unternehmen beitragen können.

Personalisierung basiert auf dem Aufbau und der Nutzung von Profilen (z.B. Kunden- und Produktprofile). In den Profilen werden die Daten gespeichert, die für die Realisierung von Personalisierungsfunktionen nötig sind. Zur Erfassung, Speicherung und Auswertung der Profile werden wiederum besondere Technologien benötigt, die diese Prozesse unterstützen und automatisieren. Daraus ergaben sich die zwei Themenschwerpunkte des Projekts PersECA II:

- Profile und
- Technologien

Diese Themen bilden auch den Gegenstand des vorliegenden Arbeitsberichts. Im Bereich „Profile“ werden Input- und Output-Profile definiert. Im Bereich „Technologien“ steht die Nutzung von Technologien, wie Collaborative Filtering, zur Automatisierung von Personalisierungsfunktionen im Vordergrund.

Das Projekt PersECA II wurde in separaten Teilprojekten mit sechs Wirtschaftspartnern durchgeführt. Dazu zählen die Anwendungspartner Actebis (Schweiz) AG, Brütsch/Rüegger Tools, buch.ch, Kantonale Drucksachen- und Materialzentrale Zürich kdmz sowie die Lösungsanbieter bbv Software Services AG und Opacc Software AG. In mehreren Workshops wurden zu Beginn Ideen für neue Personalisierungsfunktionen gesammelt, anschliessend verdichtet und priorisiert. Für die zur Umsetzung ausgewählten Funktionen wurden die Detailkonzeption und die Lösungsansätze erarbeitet. Die meisten Partner setzten die neuen Funktionen noch während des Projekts zuerst prototypisch und dann in der Live-Version um.

Die in diesem Arbeitsbericht zusammengestellten Projektergebnisse entstanden zum grössten Teil in der direkten Zusammenarbeit mit den Wirtschaftspartnern. Die zunächst theoretisch erarbeiteten Lösungsansätze wurden mit Daten und in den Systemen der Partner getestet und anschliessend verfeinert. An dieser Stelle danken wir unseren Wirtschaftspartnern für die vertrauensvolle und enge Zusammenarbeit, die massgeblich zum Gelingen des Projekts beigetragen hat. Unserem Hochschulpartner, der Information Systems Research Group an der Universität Fribourg, insbesondere Prof. Dr. Andreas Meier, danken wir für die kompetente Unterstützung bei der Entwicklung innovativer Lösungsansätze.

Das Projekt PersECA II wurde gefördert von der Kommission für Technologie und Innovation (KTI/CTI) am Bundesamt für Berufsbildung und Technologie BBT. An dieser Stelle danken wir der KTI für die finanzielle Unterstützung, ohne die dieses Projekt nicht möglich gewesen wäre.

Basel und Bern, im Oktober 2008

Uwe Leimstoll, Henrik Stormer,
Raoul Schneider, Michael Quade, Michael Pülz

Inhaltsverzeichnis

Vorwort	i
Inhaltsverzeichnis	iii
Abbildungs- und Tabellenverzeichnis	v
1 Einleitung.....	1
2 Profile für die Personalisierung	2
2.1 Kundendaten und Kundenprofile	2
2.2 Ansatz zur Entwicklung eines Profil-Frameworks	5
2.2.1 Kundenprofile mit CP/2	5
2.2.2 CP/2 Systemprofil.....	6
2.2.3 CP/2 Identifikationsprofil.....	7
2.2.4 CP/2 Transaktionsprofil	8
2.2.5 CP/2 Interaktionsprofil	9
2.2.6 CP/2 Präferenzprofil	11
3 Technologien für die Personalisierung	12
3.1 Allgemeines: Offline- und Online-Berechnungen	12
3.2 Auswertungsmethoden und Algorithmen für Empfehlungssysteme.....	14
3.2.1 Content-based Filtering	14
3.2.2 Collaborative Filtering.....	15
3.2.2.1 User-based Collaborative Filtering	15
3.2.2.2 Item-based Collaborative Filtering.....	18
3.2.3 Hybride Systeme	21
3.2.4 Aktuelle Forschungsthemen	21
3.3 Berechnung der Güte von Empfehlungen.....	22
3.3.1 Hitrate-Verfahren.....	22
3.3.2 Response-Verfahren	22
3.4 Optimierung von Empfehlungen	22
3.4.1 Alpha-Wert im Item-based-Collaborative-Filtering-Algorithmus	23
3.4.2 Gute und schlechte Scouts	23
3.4.3 Berücksichtigung älterer Daten	23
3.4.4 Schwellenwerte zur Beurteilung der Güte der Empfehlungen.....	24
3.4.5 Regeln zum Aussondern von Ähnlichkeiten.....	24
3.4.6 Produktgruppen berücksichtigen	24
3.4.7 Weitere Optimierungskriterien.....	24
3.4.8 Referenzprozesse für die Optimierung persönlicher Empfehlungen	25

4	Demonstrationsanwendung myLandscapes	27
5	Der Profil-Browser	28
5.1	Sinn und Zweck des Profil-Browsers	28
5.2	Aufbau der Applikation	28
5.2.1	Verwendete Technologien	29
5.2.2	Datenbankmodell	29
5.2.3	Funktionen und ihre Klassen	30
5.2.3.1	Login, Logout und Benutzerverwaltung	30
5.2.3.2	Import und Verwaltung von Transaktionsdaten	30
5.2.3.3	Verwaltung und Berechnung von Empfehlungen	31
5.2.3.4	Auswertung und Darstellung der Transaktionsdaten und Empfehlungen	31
5.3	Ausblick auf die Erweiterung des Profil-Browsers	32
6	Zusammenfassung und Schlussfolgerungen	35
	Literaturverzeichnis	36

Abbildungs- und Tabellenverzeichnis

Abb. 3.1:	Customer Profile Life Cycle	13
Abb. 3.2:	Technische Sicht der Profilverarbeitung in Personalisierungssystemen	13
Abb. 3.3:	Prozess der Offline-Berechnung.....	25
Abb. 3.4:	Prozess der Online-Filterung und -Sortierung	26
Abb. 4.1:	Fotogalerie myLandscapes.....	27
Abb. 5.1:	Datenbankmodell.....	29
Tab. 2.1:	Explizite Profile	3
Tab. 2.2:	Implizite Profile	4
Tab. 2.3:	Kalkulierte Profile.....	5
Tab. 2.4:	Systemprofil in CP/2	6
Tab. 2.5:	Identifikationsprofil in CP/2	7
Tab. 2.6:	Detailprofil zu Adressen.....	7
Tab. 2.7:	Transaktionsprofil in CP/2.....	8
Tab. 2.8:	Detailprofil zum Gesamtpreis.....	8
Tab. 2.9:	Interaktionsprofil in CP/2.....	9
Tab. 2.10:	Detailprofil zu aufgerufenen Webseiten.....	10
Tab. 2.11:	Präferenzprofil in CP/2.....	11
Tab. 3.1:	Einfache Bewertungsmatrix für User-based Collaborative Filtering	16
Tab. 3.2:	Ähnlichkeitsmatrix für User-based Collaborative Filtering	17
Tab. 3.3:	Einfache Bewertungsmatrix für Item-based Collaborative Filtering	19
Tab. 3.4:	Ähnlichkeitsmatrix für Item-based Collaborative Filtering.....	19
Tab. 3.5:	Speicherung der Ähnlichkeitsmatrix für Item-based Collaborative Filtering	19
Tab. 3.6:	Beispiel einer Bewertungsmatrix zur Qualitätssicherung.....	22
Tab. 5.1:	Übersicht über die verwendeten Technologien	29
Tab. 5.2:	Aufbau einer CSV-Datei für den Import.....	31

1 Einleitung

Über die letzten 15 Jahre hat das Internet die Beziehung von Unternehmen zu ihren Kunden – und umgekehrt – tiefgreifend verändert. Durch den einfachen Informationszugang hat die Markttransparenz erheblich zugenommen, und der Kunde kann mit wenigen Klicks zwischen den Anbietern wechseln. Infolgedessen verwenden Unternehmen immer mehr Ressourcen darauf, die Kunden durch besseren Service, individualisierte Angebote oder Kundenbindungsprogramme an das Unternehmen zu binden. Eine zunehmend bedeutende Rolle spielt dabei die Personalisierung. Unter Personalisierung wird im E-Commerce die individuelle Anpassung von Inhalten und Funktionen von E-Commerce-Applikationen an die speziellen Bedürfnisse eines Nutzers verstanden. Die Anpassung erfolgt auf der Basis von Daten, die über den Nutzer gespeichert werden.

Das Know-how zur Realisierung von Personalisierungsfunktionen ist heute noch nicht allgemein verfügbar, vor allem wenn die Auswertung von Kundendaten komplexe und aufwendige Berechnungen erfordert. Schwierig für Shopbetreiber sind auch Auswahl und Gestaltung von Funktionen, die speziell für ihre Kunden geeignet sind. Das Ziel des Projekts bestand deshalb darin, Methoden und Lösungsansätze zu erarbeiten, die Unternehmen bei der Einführung personalisierter Funktionen unterstützen. Der vorliegende Arbeitsbericht dokumentiert ausgewählte Ergebnisse des Projekts, die in dem Sinne verallgemeinerbar sind, dass sie auf konkrete Anwendungsbereiche übertragen werden können.

Das folgende Kapitel 2 befasst sich mit dem Aufbau und der Pflege von Profilen. Kapitel 3 geht anschliessend auf die grundlegenden Technologien ein, mit denen die Profile ausgewertet und kundenspezifische Informationen gewonnen werden können. Zwei Anwendungen, die im Rahmen des Projekts PersECA II entstanden sind, werden in Kapitel 4 und 5 vorgestellt. Ein Schlusskapitel fasst die Untersuchungsergebnisse zusammen und zieht Schlussfolgerungen (Kapitel 6).

2 Profile für die Personalisierung

2.1 Kundendaten und Kundenprofile

Personalisierung basiert auf Kundendaten. Dies geht bereits aus der Definition der Personalisierung hervor. In Anlehnung an andere Definitionen [z.B. Riecken 2000; Wu et al. 2003] wird Personalisierung hier wie folgt definiert:

Personalisierung ist die individuelle Anpassung von Inhalten und Funktionen von E-Commerce-Applikationen an die speziellen Bedürfnisse eines Nutzers. Die Anpassung erfolgt auf Basis explizit oder implizit erhaltener und gespeicherter Daten über den Nutzer.

Da es im Bereich des E-Commerce primär darum geht, Personalisierung zu automatisieren, muss das Wissen über den Kunden in einer maschinell verarbeitbaren Form – also in Form von Daten – verfügbar gemacht werden. Alternativ oder ergänzend dazu besteht die Möglichkeit, aus Daten, die über Kunden bereits gespeichert sind, neues Wissen über den Kunden zu generieren.

Kundendaten fallen im Unternehmen zu ganz unterschiedlichen Gelegenheiten an. Transaktionsorientierte Systeme liefern Daten über die verkauften Produkte und darüber, welche Kunden diese Produkte gekauft haben. Weitere Daten sind im Kunden- und Produktstamm gespeichert oder werden in Customer-Relationship-Management-Systemen (CRM-Systemen) vorgehalten. Im E-Commerce stehen weitere Möglichkeiten offen, Kundendaten zu erfassen: In Logfiles wird unter anderem gespeichert, welche Produkte ein Kunde in einem Katalog betrachtet hat. Auch Produktbewertungen durch Kunden, die Abfrage von Interessen oder die Einrichtung von Community-Plattformen dienen unter anderem der Gewinnung von Kundendaten.

Werden Kundendaten in verschiedenen Informationssystemen gehalten, stellen sie in der Regel noch nicht den gewünschten Ausgangspunkt für eine automatisierte Verwendung dar. Dazu müssen die Daten vereinheitlicht und zusammengeführt werden. Um kundenbezogene Auswertungen machen zu können, müssen die Kundendaten ausserdem in Bezug zu einem konkreten Kunden gesetzt werden. Erst diese Zuordnung der Daten zu konkreten Kunden führt zum Kundenprofil, das wie folgt definiert wird [Risch 2008, S. 23]:

Ein Kundenprofil umfasst alle relevanten Kundendaten eines Unternehmens, die einer eindeutig bestimmbar Person zugeordnet werden können und die Aufschluss über eine bestehende oder potenzielle Geschäftsbeziehung geben.

Die Tab. 2.1 bis Tab. 2.3 zeigen verschiedene Kundenprofile in einer Systematik, die in Abhängigkeit von der Herkunft und der Art der Daten gebildet wurde [Quade et al. 2008, S. 44 ff. in Anlehnung an Schubert/Leimstoll 2002, S. 24). Darin wird zwischen

- impliziten,
- expliziten und
- kalkulierten Profilen

unterschieden. Implizite Profile basieren auf Daten, die von Transaktionssystemen oder im Logfile aufgezeichnet werden, ohne dass sich der Kunde dessen bewusst sein muss. Die Da-

ten der expliziten Profile hingegen gibt der Kunde ausdrücklich bekannt. Die kalkulierten Profile wiederum enthalten Daten, die von analytischen Systemen berechnet worden sind, und auf deren Basis zum Beispiel Empfehlungen angezeigt werden können.

Tab. 2.1: Explizite Profile

Profiltyp	Nr.	Profilsegmente	Eigenschaft/Attribute
Identifikationsprofil	A.1	Benutzeridentifikation	Benutzername, E-Mail-Adresse, Signatur
	A.2	Kundenidentifikation	Bei Privatpersonen: Vorname, Name, Kundennummer Bei Unternehmen: Vorname, Name, Ansprechpartner, Firmenname, Organisationseinheit, Kundennummer, Handelsregister, Organisationsform, MWST Nummer, weitere E-Mail-Adressen und Signaturen Kunden-URL
	A.3	Funktion	Vor allem bei Unternehmen: Rollen und Berechtigungen (Einkäufer, Bedarfssteller, Administrator usw.)
	A.4	Adressen (geografisch)	Bestell-, Rechnungs- und Lieferadressen, Telefonnummern
	A.5	Demografisch	Selbstkategorisierung seitens des Kunden in vordefinierte Kategorien Personen: Alter, Geschlecht, Einkommen, Herkunft, Beruf, Tätigkeit, Anzahl Kinder, Wohnsituation, Religionszugehörigkeit usw. Unternehmen: Branche, Grösse / personenbezogen: Sprache, Hierarchiestufe, Verantwortlichkeit, Geschlecht usw.
	A.6	Psychografisch	Personen: Lebensstil, Einstellung (Politik), Persönlichkeit Unternehmen: Marktauftritt, Unternehmenskultur
Konditionenprofil	B.1	Vertrag	Rabattstufe, individuelle Preise, Mindestmengen (Bonus/Malus)
	B.2	Zahlungsart	Zahlungsarten (Rechnung, Kreditkartennummern, Kontoangaben), Limiten, Lastschriften
	B.3	Versandart	Lieferarten, Verpackung, Kennzeichnung (Gravuren, Aufdruck)
Präferenzprofil	C.1	Themen	Selbstausswahl von angebotenen Themen: Hobbys, Interessen, Vorlieben, Geschmack, Genres
	C.2	Produkte	Auswahl von angebotenen Produktgruppen oder Produkttypen (z.B. Bücher, DVD, CD's), Ausprägungen von Produkten (z.B. Premium, Low Cost, Business, Consumer)
	C.3	Katalog	Anzeige von Teilsortiment, Einträge in Wunschlisten, Bestelllisten und Einkaufslisten
	C.4	Interaktionseinstellungen	(Browser-)Einstellungen, voreingestellte Suche/Suchbegriffe, Favoriten Produktgruppen, Startseite, Navigation und Menu, Auswahl von Portlets
	C.5	Informationskanäle	Abonnierte Newsletter, Kataloge, Kundenzeitschrift, Werbeflyer
Gemeinschaftsprofil	D.1	Beziehungen	Angabe von besonderen Beziehungen zu anderen Kunden (z.B. Community, Soulsister, Affinitätsgruppen)
	D.2	Bewertungen	Bewertungen von Produkten, Webseiten, Reviews oder anderen Kunden anhand vorgefertigter Skalen (z.B. Musikgeschmack: 1 für „sehr gut“ bis 5 für „sehr schlecht“)

Profiltyp	Nr.	Profilsegmente	Eigenschaft/Attribute
	D.3	Meinungen	Freitexteingaben von Meinungen und Erfahrungen, typische Community Beiträge, Multimedia-Beiträge wie Bilder, Fotos, Videos, MP3-Dateien.
	D.4	Inhalte	Einträge in Foren zu bestimmten Themen und Kategorien, Inhalte der Kommunikation (Chat, E-Mail, Telefongespräche)

Tab. 2.2: Implizite Profile

Profiltyp	Nr.	Profilsegmente	Eigenschaft /Attribut
Transaktionsprofil (TA-Profil)	E.1	Offertanfragen	Speicherung von Offertanfragen mit Zuordnung zu den Vorgabekategorien (Metadaten)
	E.2	Kauf	Speicherung der durchgeführten Transaktionen (Käufe, Zahlungen, Inanspruchnahme von Dienstleistungen etc.), gekaufte Produkte / Dienstleistungen werden den Vorgabekategorien (Metadaten) zugeordnet.
	E.3	Rücksendungen	Defekte Ware, Rücksendungen wegen Falschbestellungen, Nichtgefallen
Interaktionsprofil	F.1	Navigation	Clickstream (angeschaute Webpages verlinkt mit vordefinierten Kategorien (Metadaten) der angezeigten Produkte oder Informationen), die ein vermeintliches Interesse widerspiegeln können (Politik, Computer, Weltgeschehen, Börse, etc.) Web Analytics
	F.2	Suchanfragen	Aufgezeichnete Suchanfragen und die dazu ausgewählten Suchergebnisse
Verhaltensprofil	G.1	Ereignisse	Registrierte und aufgezeichnete direkte Kundenkontakte (z.B. Hotline) und indirekt (Kontaktformular, E-Mail, Chats), Kundenreklamationen und Beschwerden, i. d. R. Text
	G.2	Zahlungsverhalten	Gutschriften, Zahlungszeiten (Skonto), Mahnungen, Sperrungen, Betreibungen
	G.3	Kommunikation	Genutzte Kommunikationswege (Telefon, E-Mail, Fax), Tonfall in der Kommunikation (freundlich, sachlich, unfreundlich), Zeiten der Kommunikation (Uhrzeit, Wochentag, Jahreszeit)
	G.4	Reaktion	Das Reaktionsverhalten der Kunden z.B. auf Newsletter oder Marketingkampagnen.
Externe Profile	H.1	Marktforschung	Ergebnisse der Marktforschung (bspw. Kundenverhalten, Kundentypen, Kaufverhalten in unterschiedlichen Regionen)
	H.2	Bonität	Bonitätsprüfung (bspw. bei Deltavista, Intrum Justizia)
	H.3	GIS-Daten	Demografische Daten aufgrund des Wohnorts/Wohnlage (bspw. aufgrund Mietpreise in Strasse, Bonität in PLZ-Kreis usw.)
	H.4	Kontext	Daten aus anderen Quellen (bspw. Wetterbericht, Lokale News, Events, Google, GIS, usw.)

Tab. 2.3: Kalkulierte Profile

Profiltyp	Nr.	Profilsegmente	Eigenschaft
Verwandtschaftsprofil (Collaborative Filtering)	I.1	Nachbarschafts-clique (User-based)	Eine Anzahl von anderen Kunden, zu denen der Kunde z.B. anhand von Bewertungen, Transaktionen, Interaktionen oder seines Verhaltens einen hohen berechneten Verwandtschaftsgrad aufweist.
	I.2	Produktinteressen (Item-based)	Empfehlungen für Produkte, die anhand der reinen Produktkäufe anderer Kunden, den Kunden interessieren können.
	I.3	Segmentierung	Einteilung der Kunden in Marketingsegmente, die z.B. anhand von Transaktionsvolumen, Art der gekauften Produkten, Häufigkeit der Shopbesuche berechnet wurden.
Inhaltsprofil (Content-based Filtering)	J.1	Produkte	Empfehlungen für Produkte, die zu den vom Kunden gekauften Produkten eine grosse Ähnlichkeit aufweisen (z.B. aus dem Genre Sc-Fi Filme)
	J.2	Themen	Empfehlung z.B. für Informationen aus bestimmten Themenbereichen, aus denen sich der Kunde schon öfters Informationen abgerufen hat (z.B. Sport, Politik)

Die im vorhergehenden Abschnitt beschriebenen Profile helfen dabei, Kundendaten im Unternehmen zu identifizieren und zu systematischen Kundenprofilen zusammenzuführen. Von dieser Basis ausgehend kann analysiert werden, wie die Daten verwendet werden können, um dem Kunden einen Mehrwert zu bieten oder um Verkaufsstrategien umzusetzen. Umgekehrt kann geprüft werden, welche Daten im Unternehmen zur Verfügung stehen und welche künftig neu erhoben werden müssten, um konkrete Personalisierungsfunktionen realisieren zu können.

2.2 Ansatz zur Entwicklung eines Profil-Frameworks

Für die Systematisierung der Kundenprofile und vor allem für deren Speicherung und Austausch in und zwischen den an der Personalisierung beteiligten Informationssystemen, ist ein technischer Standard sehr hilfreich. Ein erster Schritt in Richtung eines solchen Standards wurde im Projekt PersECA II mit der Entwicklung eines Ansatzes für die Bildung eines umfassenden Profilframeworks getan. Die folgenden Abschnitte beschreiben die Grundidee des Ansatzes sowie Beispiele für in XML beschriebene Profile.

2.2.1 Kundenprofile mit CP/2

CP/2 ist eine im PersECA-II-Projekt entwickelte Möglichkeit, Kundenprofile in Form von XML-Dokumenten standardkonform zu speichern und auszutauschen. CP/2 kommt zur weiteren Vereinfachung ohne XML-Attribute aus. In XML lassen sich Informationen entweder als Kindelement (Child Element) eines Tags oder als Attribut ablegen. Es gibt einige Fälle, bei denen nur das Kindelement möglich ist (etwa bei speziell formatierten Informationen), in den meisten Fällen kann der Designer des XML Schemas selbst entscheiden [Harold 2003]. In der Vergangenheit wurde bereits häufig über die Vor- und Nachteile von XML-Attributen diskutiert [Ogbuji 2004; Harold 2003]. Da CP/2 in dieser Arbeit insbesondere für die Darstellung von Implementierungsdetails eingesetzt wird, wurde auf XML-Attribute verzichtet. Ein Wert aus

dem CP/2-Profil lässt sich in diesem Falle über den Pfad (z.B. *profile.system.username*) darstellen.

Das CP/2-Profil wird von dem Feld *cp2* umschlossen. Jedes Kundenprofil wird mit *profile* eingeleitet:

```
<cp2>
  <profile>
    <system>...</system>
    <identification>...</identification>
    <transaction>...</transaction>
    <interaction>...</interaction>
    <preference>...</preference>
  </profile>
  <profile>
    ...
  </profile>
</cp2>
```

CP/2 erlaubt die Verwaltung von Profilen, die nicht alle Unterprofile definiert haben. Bietet ein Onlineshop seinen Kunden beispielsweise keine Bewertungsfunktionalität an, kann das Präferenzprofil weggelassen werden. Im Folgenden werden die Unterprofile genauer vorgestellt.

2.2.2 CP/2 Systemprofil

Das Systemprofil speichert die wesentlichen Systeminformationen eines Kunden. In CP/2 wird davon ausgegangen, dass sich ein Kunde beim Onlineshop registriert hat oder eine automatische Registrierung über den Shop Administrator erfolgt ist. Die Informationen über das Kundenkonto, etwa der eindeutige Benutzername oder die bevorzugte Sprache, werden im Systemprofil unter dem Feld *system* abgelegt.

Tab. 2.4: Systemprofil in CP/2

Feld	Beschreibung
username	Der Benutzername des Kunden
creationDate	Erstellungsdatum des Kundenprofils
modificationDate	Datum der letzten Veränderung am Kundenprofil
Description	Vom Shop Administrator gepflegte Kurzbeschreibung des Kunden
preferredLanguage	Bevorzugte Sprache des Kunden

Ein Systemprofilbeispiel sieht folgendermassen aus:

```
<system>
  <username>stormerh</username>
  <creationDate>12.02.2006</creationDate>
  <modificationDate>08.12.2006</modificationDate>
  <description>Kundenkonto stormerh</description>
  <preferredLanguage>de</preferredLanguage>
</system>
```

2.2.3 CP/2 Identifikationsprofil

Im Identifikationsprofil werden Anschrift und Rollen des Kunden abgelegt:

Tab. 2.5: Identifikationsprofil in CP/2

Feld	Beschreibung
name	Der Name des Kunden. Bei Geschäftskunden der Name der Firma
customerNumber	Die Kundennummer
customerRole	Die Rolle des Kunden (Privatkunde, Geschäftskunde)
addresses	Eine Liste mit Kundenadressen

In *addresses* werden alle vom Kunden eingetragenen Adressen mit folgenden Feldern abgespeichert:

Tab. 2.6: Detailprofil zu Adressen

Feld	Beschreibung
addressLine1	Erste Zeile der Adresse
addressLine2	Zweite Zeile der Adresse
postalCode	Postleitzahl
City	Der Name der Stadt
country	Der Name des Landes
defaultInvoiceAddress	Ein Flag das angibt, dass diese Adresse die Standardrechnungsadresse ist
defaultShippingAddress	Ein Flag das angibt, dass diese Adresse die Standardversandadresse ist

Das Identifikationsprofil erwartet, dass es bei mehreren Adressen genau eine Standardrechnungsadresse (*defaultInvoiceAddress*) und genau eine Standardversandadresse (*defaultShippingAddress*) gibt. Normalerweise werden die Adressen der letzten Bestellung übernommen.

Ein einfaches Beispiel für ein Identifikationsprofil könnte folgendermassen aussehen:

```
<identification>
  <name>Henrik Stormer</name>
  <customerNumber>12321</customerNumber>
  <customerRole>Private</customerRole>
  <addresses>
    <address>
      <addressLine1>Bd der Perolles 18a</addressLine1>
      <addressLine2></addressLine2>
      <postalCode>1700</postalCode>
      <city>Fribourg</city>
      <country>Switzerland</country>
      <defaultInvoiceAddress>true</defaultInvoiceAddress>
```

```

    <defaultShippingAddress>true</defaultShippingAddress>
  </address>
</addresses>
</identification>

```

2.2.4 CP/2 Transaktionsprofil

Das Transaktionsprofil speichert die Bestellungen eines Kunden. Sollte die Bestellung mit Teillieferungen erfolgt sein, wird dies nicht berücksichtigt. Es werden nur Bestellungen erfasst, die tatsächlich ausgeführt wurden; zurückgezogene, gelöschte und retournierte Bestellungen fließen nicht in das Transaktionsprofil ein. Da die eingegangenen Bestellungen von den meisten Onlineshops nach diesen Regeln abgelegt werden, kann das Transaktionsprofil sehr einfach aus der Shop-Datenbank ausgelesen werden. Weitere Informationen über die Lieferung und eventuelle Rückläufe werden zumeist im ERP-System abgelegt. Die folgenden Felder müssen im Transaktionsprofil angelegt werden:

Tab. 2.7: Transaktionsprofil in CP/2

Feld	Beschreibung
orderDate	Das Datum der Bestellung
totalPrice	Der Gesamtpreis der Bestellung
shippingAddressId	Die ID der Versandadresse aus dem Identifikationsprofil
shippingCosts	Die gesamten Versandkosten
shippingType	Der Name der ausgewählten Versandart
paymentCosts	Die zusätzlich anfallenden Gebühren aufgrund der ausgewählten Zahlungsart
paymentType	Der Name der ausgewählten Zahlungsart
orderPositions	Eine Liste der einzelnen Bestellpositionen

Die *shippingAddressId* bezieht sich auf die Position der Adresse innerhalb des Identifikationsprofils. Die erste gelistete Adresse hat Id eins, die zweite zwei usw. Bei digitalen Produkten ist die *shippingAddressId* null. Das Transaktionsprofil bedingt, dass die angegebene Adresse im Identifikationsprofil vorhanden sein muss, falls *shippingAddressId* einen Wert ungleich null hat.

Die einzelnen Bestellpositionen werden in der Liste *orderPositions* gespeichert. Der Gesamtpreis aller Bestellpositionen ergibt sich aus $totalPrice - paymentCosts - shippingCosts$. Für jede Bestellposition müssen folgende Felder angelegt werden:

Tab. 2.8: Detailprofil zum Gesamtpreis

Feld	Beschreibung
itemNumber	Die eindeutige Produktnummer
quantity	Die bestellten Einheiten des Produkts
price	Der Gesamtpreis der Orderposition

Jedes Produkt muss eine eindeutige Nummer (*itemNumber*) besitzen. Der Preis für eine Einheit eines Produkts wird nicht im Transaktionsprofil abgelegt, kann aber über die Formel $price / quantity$ berechnet werden.

Ein Transaktionsprofil mit einer Bestellung von zwei Produkten hat folgendes Aussehen:

```
<transaction>
  <order>
    <orderDate>21.5.2006</orderDate>
    <totalPrice>22.30</totalPrice>
    <shippingAddressId>1</shippingAddressId>
    <shippingCosts>2.30</shippingCosts>
    <shippingType>Expressversand Schweizer Post</shippingType>
    <paymentCosts>0</paymentCosts>
    <paymentType>Kreditkarte</paymentType>
    <orderPosition>
      <itemNumber>3432</itemNumber>
      <quantity>1</quantity>
      <price>12</price>
    </orderPosition>
    <orderPosition>
      <itemNumber>2172</itemNumber>
      <quantity>2</quantity>
      <price>8</price>
    </orderPosition>
  </order>
</transaction>
```

2.2.5 CP/2 Interaktionsprofil

Das Interaktionsprofil speichert den Clickstream des Kunden auf dem Onlineshop. Hierbei wird insbesondere abgelegt, welche Produkte er sich angesehen und welche Suchbegriffe er eingegeben hat. Für jede vom Kunden geöffnete Session werden folgende Felder angelegt:

Tab. 2.9: Interaktionsprofil in CP/2

Feld	Beschreibung
creationDate	Das Datum an dem die Session angelegt wurde
userAgent	Der <i>userAgent</i> , der für diese Session übergeben wurde
ipAddress	Die IP-Adresse des Kunden
referrerURL	URL, von der aus der Kunde den Shop aufgerufen hat
websiteVisits	Eine Liste der Webseiten, die aufgerufen wurden

Für jeden Webseitenbesuch werden folgende Felder gespeichert:

Tab. 2.10: Detailprofil zu aufgerufenen Webseiten

Feld	Beschreibung
linkURL	Die URL der aufgerufenen Webseite
inputParameters	Die Eingabeparameter der Seite

Für die Auswertung des Clickstreams ist die aufgerufene URL nur teilweise von Interesse. Lediglich die aufgerufene Seite, die im Folgenden, mit *action* bezeichnet werden soll, sowie die möglicherweise übergebenen Parameter sind interessant. Die übrigen Teile der URL sind üblicherweise konstant und dem Shopadministrator bekannt. Der eventuell vorhandene Query String, also die an den Server übergebenen Parameter werden in *inputParameters* abgelegt. Da es möglich ist, dass mehr als ein Eingabeparameter übergeben wird, werden die *inputParameter* ebenfalls als Liste gespeichert, mit dem Name sowie Wert des Eingabeparameters. Eine URL der Art

`http://www.esarine.com/edvdshop/ViewProduct?productId=2`

wird in den oben definierten Feldern folgendermassen abgelegt:

```
<website>
<action>ViewProduct</action>
<inputParameters>
  <inputParameter>
    <name>productId</name>
    <value>2</value>
  </inputParameter>
</inputParameters>
</website>
```

Es wird also nur die Seite gespeichert, die aufgerufen wurde. Ein Beispiel, in dem ein Kunde nur eine einzige Session eröffnet und während dieser zwei Seiten betrachtet, sieht folgendermassen aus:

```
<interaction>
  <session>
    <creationDate>29.10.2006</orderDate>
    <userAgent>Mozilla 6.0 comp</userAgent>
    <ipAddress>21.212.22.7</ipAddress>
    <referrerURL>http://www.google.ch</referrerURL>
    <websiteVisits>
      <website>
        <action>Welcome</action>
      </website>
      <website>
        <action>ViewProduct</action>
        <inputParameters>
          <inputParameter>
            <name>productId</name>
            <value>134</value>
```

```

    </inputParameter>
  </inputParameters>
</website>
</websiteVisits>
</session>
</interaction>

```

2.2.6 CP/2 Präferenzprofil

Im Bewertungsprofil werden die Produktbewertungen, die der Kunde abgegeben hat, gespeichert. Das Bewertungsprofil umfasst folgende Felder:

Tab. 2.11: Präferenzprofil in CP/2

Feld	Beschreibung
itemNumber	Die eindeutige Nummer des bewerteten Produkts
creationDate	Das Datum, an dem die Bewertung abgegeben wurde
ratingValue	Die Produktbewertung, normalerweise ein Wert zwischen 1 und 5 (1 bis 5 Sterne)
description	Eine Kurzkritik zum Produkt

Ein Präferenzprofil mit einer Produktbewertung hat folgendes Aussehen:

```

<preference>
  <rating>
    <itemNumber>3432</itemNumber>
    <creationDate>23.11.2006</creationDate>
    <ratingValue>4</ratingValue>
    <description>
      Ein guter Film, der trotz des Alters nichts von seiner
      Faszination verloren hat.
    </description>
  </rating>
</preference>

```

3 Technologien für die Personalisierung

3.1 Allgemeines: Offline- und Online-Berechnungen

Der gesamte Berechnungsprozess, der für die Erstellung von Outputprofilen benötigt wird, kann bei grossen Datenmengen sehr aufwendig werden. Insbesondere bei der Berechnung kalkulierter Profile, wie sie für Empfehlungssysteme benötigt werden, kann der Rechenvorgang je nach Leistungsfähigkeit der zur Verfügung stehenden Hardware-Ressourcen zwischen einigen Stunden und mehreren Tagen betragen.

Im Onlineshop ist es in der Regel notwendig, eine schnelle Bereitstellung von Ergebnissen zu gewährleisten. Dies bedeutet, dass möglichst viele Rechenvorgänge schon abgeschlossen sein müssen, bevor der Kunde im Shop aktiv wird. Die Berechnungsvorgänge, die während des Zugriffs des Kunden auf den Shop noch durchgeführt werden müssen, werden entsprechend minimiert.

Der gesamte Berechnungsprozess kann damit unterteilt werden in

- Offline- und
- Online-Berechnungen

Offline-Berechnungen finden unabhängig von den Zugriffen der Kunden auf den Shop statt. Sie werden oft nachts oder am Wochenende auf nicht genutzten Back-Up-Systemen durchgeführt. Das Ziel einer Lösungskonzeption sollte sein, den Berechnungsprozess so zu gestalten, dass die zeitaufwendigen Berechnungen offline durchgeführt werden können. Die aus den Offline-Prozessen entstehenden Daten werden als „Outputprofile“ bezeichnet.

Online-Berechnungen finden statt, sobald der Kunde mit dem Shop interagiert und eine entsprechende Funktion aufruft. Die Online-Berechnungen greifen auf die vorverarbeiteten Outputprofile zu und holen dort die Daten, die für die Generierung und Anzeige des Outputs benötigt werden. Das Ziel besteht darin, dass der Zugriff auf die vorverarbeiteten Daten möglichst effizient erfolgen kann und damit eine schnelle Ergebnisanzeige ermöglicht wird.

Der Bearbeitungszyklus von Kundendaten läuft in mehreren Schritten ab, die zusammenfassend als „Customer Profile Life Cycle“ bezeichnet werden können (Abb. 3.1). In einem ersten Schritt werden die Profile definiert, bevor im zweiten Schritt die Daten erfasst werden. Nach einer Zusammenführung und Bereinigung der Daten sind die Input-Profile entstanden. In einem weiteren Schritt, der Offlineverarbeitung (Processing), werden die zeitintensiven Berechnungen durchgeführt. Die Ergebnisse werden in den so genannten Output-Profilen gespeichert, auf die die E-Commerce-Applikation anschliessend zugreift. Im Rahmen dieses Echtzeit-Zugriffsprozesses wird nur noch eine Filterung der Daten vorgenommen. Der letzte Schritt der Customer Profile Life Cycle besteht letztlich in der Ausgabe der personalisierten Information an den Benutzer. Da der Benutzer in der Regel auf diese Information reagiert, entstehen aus seinem Verhalten neue Daten, die wiederum in die Input-Profile aufgenommen werden können. Mitunter geben sie auch Hinweise auf ein nötiges Redesign der Profile.

In Abb. 3.1 lassen sich auch die Schwerpunkte dieses Arbeitsberichts erkennen. Sie liegen wie bereits einleitend erwähnt im Bereich des Data Input und des Data Processing. Die Aspekte des Information Output in der Kommunikation mit dem Kunden, insbesondere konkrete Personalisierungsfunktionen, werden in einem separaten Arbeitsbericht beschrieben [Leimstoll et al. 2008].

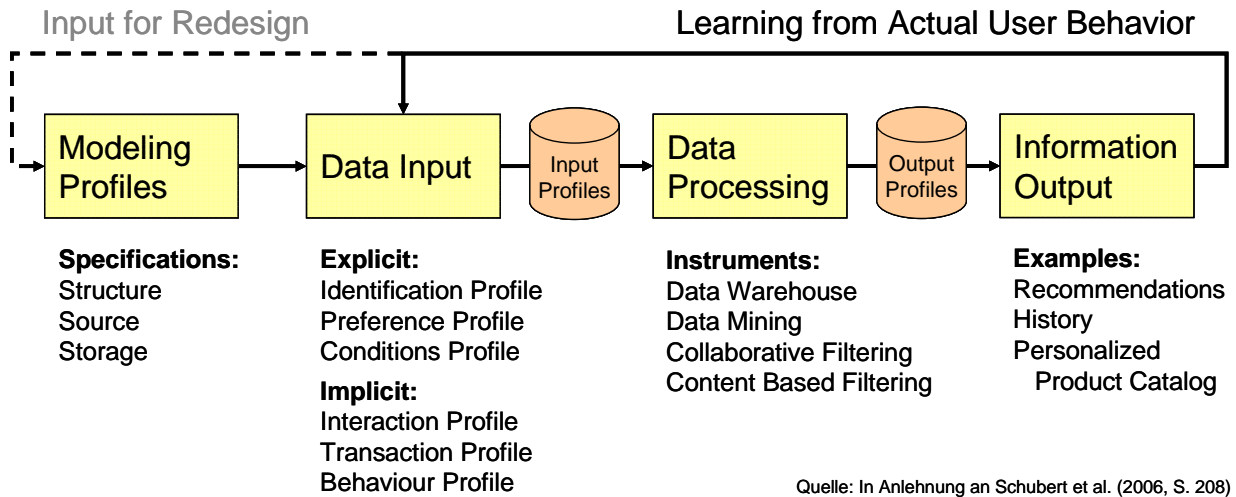


Abb. 3.1: Customer Profile Life Cycle

Die folgende Abb. 3.2 zeigt die beteiligten Systeme aus einer mehr technischen Sicht [Risch/Schubert 2005]:

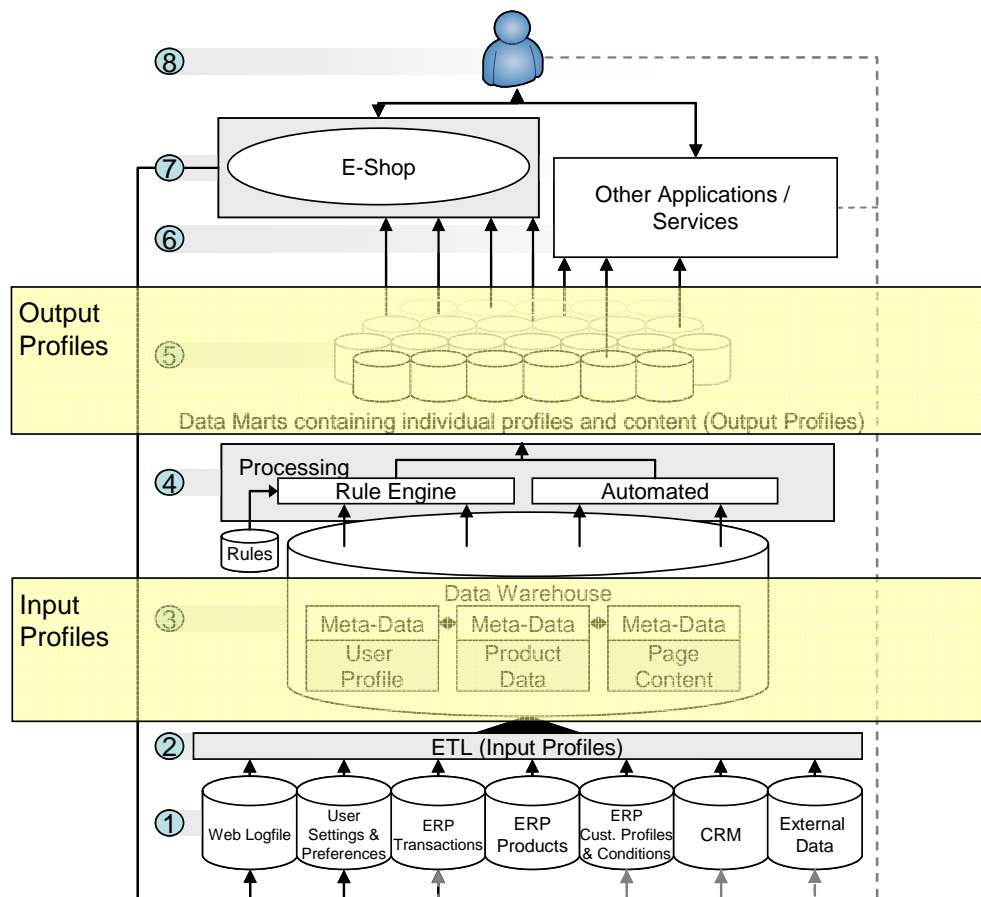


Abb. 3.2: Technische Sicht der Profilverarbeitung in Personalisierungssystemen

3.2 Auswertungsmethoden und Algorithmen für Empfehlungssysteme

Die folgenden Ausführungen konzentrieren sich auf die Datenverarbeitungsprozesse in Empfehlungssystemen. Dies liegt daran, dass Empfehlungssysteme im Vergleich zu anderen Personalisierungsfunktionen zum einen besonders aufwendige Berechnungen erfordern und zum anderen im Projekt PersECA II einen eindeutigen Schwerpunkt bildeten.

Grundsätzlich lassen für die Berechnung von Empfehlungen verschiedene Verfahren unterscheiden. Zu den Verfahren, die eine Aufteilung in Offline- und Online-Berechnungen erlauben, gehören im Wesentlichen die Folgenden:

- Content-based Filtering,
- Collaborative Filtering, das sich weiter unterteilen lässt in:
 - User-based Collaborative Filtering und
 - Item-based Collaborative Filtering.

Die Verfahren werden in den folgenden Abschnitten beschrieben.

3.2.1 Content-based Filtering

Um Empfehlungen mit Hilfe des Content-based Filtering berechnen zu können, müssen die angebotenen Produkte zunächst kategorisiert werden. Dazu sind geeignete Kategorien zu definieren (z.B. Autor, Thema, Sachgebiet, Erscheinungsjahr) und jedes einzelne Produkt muss möglichst in jeder Kategorie mit einem passenden Stichwort – auch: Parameter, Schlüsselwort – beschrieben, also klassifiziert werden (beim Sachgebiet z.B.: Mathematik, Deutsch, Biologie, Geographie). Diese Klassifizierung wird heute überwiegend manuell durchgeführt, sofern nicht geeignete Technologien zur Verfügung stehen, die solche Stichwörter aus unstrukturierten Produktbeschreibungen extrahieren. Mathematische Verfahren wie die Inverse Document Frequency (IDF¹) können dafür zum Einsatz kommen.

Um Empfehlungen zu generieren, werden die Parameter der Produktbeschreibungen anschliessend miteinander verglichen. Auf diese Weise werden ähnliche Produkte gefunden (beispielsweise Bücher mit denselben Autoren oder aus demselben Sachgebiet). Die Anzahl gleicher Schlüsselwörter ist der Indikator für die Ähnlichkeit zweier Produkte.

Vorgehen:

1. Die Ähnlichkeiten eines Produkts mit jedem anderen Produkt werden über die Schlüsselwörter statisch berechnet und in einer Ähnlichkeitsmatrix gespeichert.
2. Interessiert sich ein Kunde i für ein Produkt x , wird in der Ähnlichkeitsmatrix nach den ähnlichsten Produkten y gesucht, die der Kunde noch nicht gekauft hat (dazu muss der Kunde jedoch identifiziert worden sein).
3. Die so aufgefundenen Produkte y können dann empfohlen werden.

¹ Die "Inverse Document Frequency" ist ein Mass für die Bedeutung eines Begriffs. Diese Methode kommt beim "Information Retrieval" und "Text Mining" zum Einsatz.

Beim Content-based Filtering treten folgende Probleme auf:

- Die manuelle Verwaltung der Schlüsselwörter ist aufwändig. Wie oben erwähnt, kann in der Regel nicht davon ausgegangen werden, dass Shopbetreiber bereit sind, diesen Aufwand zu leisten.
- Um automatisches Information Retrieval anwenden zu können, müssen neben geeigneten Technologien auch entsprechende Daten (z.B. in Form von Produktbeschreibungen) vorhanden sein. Bei dem im Projekt verwendeten Beispielshop sind zu wenige Daten vorhanden, um mit diesem Verfahren zu arbeiten.

3.2.2 Collaborative Filtering

Die Grundidee des Collaborative Filtering besteht darin, zur Berechnung von Ähnlichkeiten auf das Verhalten der Benutzer zurückzugreifen. Im Gegensatz zum Content-based Filtering werden also nicht die Ähnlichkeiten der Produkteigenschaften verglichen, sondern es wird auf Grund des Benutzerverhaltens auf sinnvolle Empfehlungen geschlossen. Das Benutzerverhalten kann sich dabei auf

- getätigte Käufe,
- explizite Bewertungen,
- einfaches Anklicken von Produktinformationen (Clickstream-Auswertung)
- und Ähnliches

beziehen. Voraussetzung ist die Verfügbarkeit relativ vieler Daten. Der Begriff "Collaborative Filtering" leitet sich von einer Arbeit von Resnick et al. [1994] ab, die bereits Anfang der 90er Jahre ein Collaborative-Filtering-System namens GroupLens für die Bewertung von Beiträgen in Foren entwickelten.

Wie oben beschrieben unterscheidet man beim Collaborative Filtering weiter zwischen:

- User-based Collaborative Filtering und
- Item-based Collaborative Filtering.

Beide Verfahren werden im Folgenden erläutert.

3.2.2.1 User-based Collaborative Filtering

In einer Bewertungsmatrix werden Kunden und Produkte einander gegenübergestellt. In den Feldern der Matrix wird die "Bewertung" des Produkts p durch den Kunden a gespeichert. Bewertungen beziehen sich dabei auf explizite Bewertungen, getätigte Käufe usw. Die Werte bewegen sich zwischen -1 (Produkt p gefällt dem Kunden a überhaupt nicht) und $+1$ (Produkt p gefällt dem Kunden a sehr). \emptyset (NULL = Leer) bedeutet, dass das Produkt noch nicht gekauft wurde (bzw. dass noch keine Bewertungen vorliegen). Die Bewertungsmatrix kann man sich (vereinfacht) folgendermassen vorstellen (Tab. 3.1).

Tab. 3.1: Einfache Bewertungsmatrix für User-based Collaborative Filtering

	Kunde a	Kunde b	Kunde c	...
Produkt 1	1	∅	1	...
Produkt 2	∅	∅	1	...
Produkt 3	∅	1	∅	...
Produkt 4	1	0.5	∅	...
Produkt 5	-1	1	-0.3	...
...

Für jeden Kunden ergibt sich nun ein spezifischer Bewertungsvektor x . Der Kunde a hat z.B. den folgenden Bewertungsvektor:

$$x_{\text{Kunde a}} = (1, \emptyset, \emptyset, 1, -1)$$

Der Kunde b hat z.B. den folgenden Bewertungsvektor:

$$x_{\text{Kunde b}} = (\emptyset, \emptyset, 1, 0.5, 1)$$

Der Kunde c hat z.B. den folgenden Bewertungsvektor:

$$x_{\text{Kunde c}} = (1, 1, \emptyset, \emptyset, -0.3)$$

Ähnlichkeiten werden nun über einen Vergleich der Bewertungsvektoren der Kunden errechnet. Dabei wird jeder Kunde mit allen anderen Kunden verglichen. Eine Möglichkeit zur Berechnung der Ähnlichkeit der Vektoren ist die Cosinus-Distanzfunktion.

- x sei der Bewertungsvektor des Kunden a
- y sei der Bewertungsvektor des Kunden c
- i ist der Index für das bewertete Produkt

Die Cosinus-Distanzfunktion lautet nun:

$$d(x, y) = \cos \alpha(x, y) = \frac{x \cdot y}{|x||y|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

Durch Vergleich der beiden Kunden a und c ergibt sich folgende Rechnung:

Der *Zähler* berechnet sich zu:

$$\begin{aligned} (1 \cdot 1) + (\emptyset \cdot 1) + (\emptyset \cdot \emptyset) + (1 \cdot \emptyset) + (-1 \cdot -0.3) \\ = 1 + 0.3 = 1.3 \end{aligned}$$

Der *Nenner* errechnet sich wie folgt:

Die erste Summe des Nenners lautet:

$$1 + \emptyset + \emptyset + 1 + 1 = 3$$

Die Wurzel daraus ergibt den Wert 1.73

Die zweite Summe des Nenners ergibt:

$$1 + 1 + \emptyset + \emptyset + 0.09 = 2.09$$

Die Wurzel daraus ergibt: 1.45

Der gesamte Nenner hat also den Wert:

$$1.73 * 1.45 = 2.51$$

Nun errechnet sich die Cosinus-Distanz der beiden Kunden a und c wie folgt:

$$1.3 / 2.51 = 0.52$$

Zwischen diesen beiden Kunden besteht also eine recht hohe Ähnlichkeit.

Die Ähnlichkeit zwischen den Kunden a und b beträgt:

$$-0.5 / (1.73 * 1.5) = -0.19$$

Die Ähnlichkeit zwischen den Kunden b und c beträgt:

$$-0.3 / (1.5 * 1.45) = -0.14$$

Diese Ähnlichkeiten werden nun in einer weiteren Matrix gespeichert. In unserem Beispiel sieht diese Matrix wie folgt aus.

Tab. 3.2: Ähnlichkeitsmatrix für User-based Collaborative Filtering

	Kunde a	Kunde b	Kunde c	...
Kunde a		-0.19	0.52	
Kunde b			-0.14	
Kunde c				
...				

Zwischen den Kunden können sogenannte Nachbarschaftscliquen berechnet werden, also Gruppen von Kunden, die ähnlich bewerten oder kaufen. Befindet sich in unserem Beispiel Kunde a nun im E-Shop (Informationsphase), so können ihm aufgrund seiner Ähnlichkeit zu Kunde c Produkte empfohlen werden, die:

- Kunde c gut bewertet hat und
- Kunde a noch nicht gekauft hat.

In unserem Beispiel würde Kunde a also das Produkt 2 empfohlen werden.

Beim User-based Collaborative Filtering existieren die folgenden Probleme:

- Die Bewertungsmatrix ist normalerweise sehr spärlich gefüllt, deshalb können für viele Produkte keine Aussagen gemacht werden. Für die Beispieldaten unseres Projekts (siehe unten) war die Bewertungsmatrix hauptsächlich mit \emptyset gefüllt. Dies liegt daran, dass die meisten Kunden nur sehr wenige Käufe getätigt haben und auch keine aktive Bewertung der Produkte vornehmen. Dadurch errechnet sich die Cosinus-Distanz zwischen den Kunden meist zu \emptyset und es können keine Empfehlungen generiert werden.
- Der Rechenaufwand für die Berechnungen hängt sowohl von der Anzahl der Kunden als auch von der Anzahl der Produkte ab. Bei grossen Matrizen kann dies zu Skalierungsproblemen (überproportionale Steigerung der Rechenzeit) führen.

- Damit ein Produkt überhaupt empfohlen werden kann, muss zu diesem Produkt mindestens eine positive Bewertung vorliegen ("First-Rater Problem").

3.2.2.2 Item-based Collaborative Filtering

Die Methode des Item-based Collaborative Filtering verfolgt im Wesentlichen den Ansatz: "Benutzer, die das Produkt x gekauft haben, haben auch das Produkt y gekauft". Das Verfahren ist dreistufig:

1. Ausgehend von der Kunden/Produkte-Bewertungsmatrix (wie oben beim User-based Collaborative Filtering beschrieben) werden Ähnlichkeiten zwischen den Produkten berechnet (und nicht Ähnlichkeiten zwischen Kunden). Es entsteht eine Produkte/Produkte-Ähnlichkeitsmatrix. Zur Berechnung der Ähnlichkeiten hat sich der Algorithmus von Deshpande und Karypis [2004] als sinnvoll erwiesen [Leimstoll/Stormer 2007]. Die Berechnungen basieren bei diesem Verfahren auf den von den Kunden getätigten Käufen (Transaktionsprofil). Da die Berechnung der Produkte/Produkte-Ähnlichkeitsmatrix sehr zweitaufwändig ist, kann sie nicht zur Laufzeit durchgeführt werden, wenn ein Kunde im Shop agiert. Stattdessen müssen die Ähnlichkeiten schon vorab berechnet werden – in einem Offline-Prozess (3.1).

Das eigentliche Item-based Collaborative Filtering ist mit diesem Schritt beendet. Auf Basis der Produkte/Produkte-Ähnlichkeitsmatrix können für jedes Produkt die jeweils ähnlichsten Produkte bestimmt werden. Um von diesen noch nicht personalisierten Informationen zu personalisierten Empfehlungen zu gelangen, kann das Verfahren wie im Projekt um die folgenden beiden Schritte ergänzt werden:

2. Zur Laufzeit wird dann für den jeweiligen Kunden eine individuelle Empfehlungsliste ("Top-N-Liste") berechnet. Hierzu werden entweder die von ihm getätigten Käufe oder – falls noch keine Käufe vorliegen – die von ihm angeklickten Produkte herangezogen. Über die unter 1. beschriebene Produkte/Produkte-Matrix werden dann die ähnlichsten Produkte bestimmt, indem je Produkt die Summen der Ähnlichkeiten berechnet werden. Hierdurch entsteht eine persönliche Top-N-Liste.
3. In einem dritten Schritt muss die persönliche Empfehlungsliste noch gefiltert werden, um z.B. zu vermeiden, dass bereits gekaufte Produkte nochmals empfohlen werden, falls dies für dieses Produkt keinen Sinn macht (z.B. eine bestimmte DVD).

Der unter 1. genannte Algorithmus von Deshpande und Karypis [2004] berechnet Ähnlichkeiten zwischen Produkten aufgrund der Kunden/Produkte-Bewertungsmatrix. Die Berechnung der Produkte/Produkte-Matrix erfolgt nach folgender Formel²:

$$\text{sim}(i, j) = \frac{\text{Freq}(ij)}{\text{Freq}(i) \times (\text{Freq}(j))^\alpha}$$

Dabei bedeuten:

- $\text{sim}(i, j)$: Ähnlichkeit zwischen den beiden Produkten i und j

² Die tatsächlich verwendete Formel ist noch etwas komplizierter und berücksichtigt die unterschiedliche Häufigkeit von Käufen einzelner Kunden. Kunden, die viele Produkte gekauft haben, erhalten ein geringeres Gewicht als solche, die weniger häufig Produkte gekauft haben.

- Freq (ij): Häufigkeit, mit der die Produkte i und j gemeinsam gekauft wurden
- Freq (i): Häufigkeit, mit der Produkt i gekauft wurde
- Alpha: Gewichtungsparemeter – Dieser Wert bewegt sich zwischen 0 und 1.

Gehen wir davon aus, dass die folgende Matrix angibt, welche Kunden welche Produkte gekauft haben (Kauf = 1, \emptyset = Leer, kein Kauf).

Tab. 3.3: Einfache Bewertungsmatrix für Item-based Collaborative Filtering

	Kunde a	Kunde b	Kunde c	...
Produkt 1	1	\emptyset	1	
Produkt 2	\emptyset	\emptyset	1	
Produkt 3	\emptyset	1	\emptyset	
Produkt 4	1	1	\emptyset	
Produkt 5	1	1	1	
...				

Die Ähnlichkeiten zwischen den Produkten 1 und 2 errechnen sich dann wie folgt:

$$\text{sim}(\text{Produkt 1}, \text{Produkt 2}) = 1 / (2 * 1^{0.5}) = 1 / 2 = 0.5$$

Die folgende Tabelle zeigt die Ähnlichkeitsmatrix der Produkte (Tab. 3.4):

Tab. 3.4: Ähnlichkeitsmatrix für Item-based Collaborative Filtering

	Produkt 1	Produkt 2	Produkt 3	Produkt 4	Produkt 5	...
Produkt 1		0.5	\emptyset	0.35	0.58	
Produkt 2			\emptyset	\emptyset	0.58	
Produkt 3				0.7	0.58	
Produkt 4					0.58	
Produkt 5						
...						

Diese Matrix kann wie folgt gespeichert werden (Tab. 3.5):

Tab. 3.5: Speicherung der Ähnlichkeitsmatrix für Item-based Collaborative Filtering

P1#	P2#	Similarity
Produkt 1	Produkt 2	0.5
Produkt 1	Produkt 3	\emptyset
Produkt 1	Produkt 4	0.35
Produkt 1	Produkt 5	0.58
Produkt 2	Produkt 1	0.5
Produkt 2	Produkt 3	\emptyset
Produkt 2	Produkt 4	\emptyset
Produkt 2	Produkt 5	0.58

P1#	P2#	Similarity
Produkt 3	Produkt 1	∅
Produkt 3	Produkt 2	∅
Produkt 3	Produkt 4	0.7
Produkt 3	Produkt 5	0.58
Produkt 4	Produkt 1	0.35
Produkt 4	Produkt 2	∅
Produkt 4	Produkt 3	0.7
Produkt 4	Produkt 5	0.58
Produkt 5	Produkt 1	0.58
Produkt 5	Produkt 2	0.58
Produkt 5	Produkt 3	0.58
Produkt 5	Produkt 4	0.58
...

P1# und P2# symbolisieren die Produktnummern zweier unterschiedlicher Produkte, deren Ähnlichkeit an Hand des oben beschriebenen Algorithmus berechnet und in der Spalte *Similarity* gespeichert wird. Wie oben beschrieben wird diese Tabelle nicht zur Laufzeit, sondern offline berechnet. Anschliessend laufen die Punkte 2 und 3 wie oben beschrieben ab, um einem spezifischen Kunden Produktempfehlungen geben zu können.

Die individuelle Empfehlung für einen Kunden erfolgt dann dynamisch im Moment des Besuchs des E-Shops. Mit folgendem SQL³-Befehl lassen sich auch bei einer grossen Bewertungsmatrix dynamisch individuelle Empfehlungen bestimmen:

```

SELECT      P2#
FROM        P1# IN (Px1, Px2, ..., Pxn)
GROUP BY   P2#
ORDER BY   SUM (Similarity) DESC;
    
```

Die Liste (P_{x1}, ..., P_{xn}) wird dynamisch erzeugt und beinhaltet die schon gekauften Produkte bzw. die gerade angeschauten (angeklickten) Produkte des jeweiligen Kunden, für den eine Empfehlung berechnet werden soll. Es wird eine Liste von Produkten P2# ("Top-N-Liste") erzeugt, die absteigend nach Ähnlichkeit mit den bereits gekauften (bzw. angeklickten, falls der Kunde noch keine Käufe getätigt hat) sortiert ist. Diese Liste dient dann als Grundlage für die weitere Verarbeitung. So müssen möglicherweise gewisse Produkte aus der Liste herausgefiltert werden, die nicht angeboten werden sollen (Details hierzu siehe unter 3.4).

³ SQL (Structured Query Language) ist eine standardisierte Datenbanksprache zur Definition, Abfrage und Manipulation von Daten in relationalen Datenbanken.

Vorteil des Item-based Collaborative Filtering

Im Gegensatz zum User-based Collaborative Filtering nimmt die Komplexität nur mit der Anzahl der Produkte zu (nicht auch noch mit der Anzahl der Kunden). Hierdurch ist die Berechnung weniger laufezeitintensiv und besser skalierbar.

Nachteil des Item-based Collaborative Filtering

Ein bekanntes Problem beim Item-based Collaborative Filtering ist das sogenannte "Cold-Start Problem". Einem neuen Kunden, der noch keine Produkte gekauft hat, können keine Produkte auf Basis seiner bisherigen Käufe empfohlen werden. Zur Lösung dieses Problems werden Clickstreams aufgezeichnet und verwendet. Die Berechnungen basieren in diesem Fall also auf den Produkten, die ein Kunde anklickt oder die er während seines Einkaufs in seinen virtuellen Warenkorb legt.

3.2.3 Hybride Systeme

Hybride Systeme verbinden die beiden grundlegenden Methoden Content-based Filtering und Collaborative Filtering miteinander. Die jeweiligen Nachteile sollen dabei eliminiert werden. Im Wesentlichen geht man dabei wie folgt vor:

- Wann immer möglich, wird das Collaborative Filtering genutzt (User- oder Item-based).
- Sollte dies nicht möglich sein (z.B. auf Grund des Cold-Start-Problems oder weil die Bewertungsmatrix zu spärlich gefüllt ist), wird das Content-based Filtering verwendet.

Darüber hinaus kann man versuchen, die Informationen über Produkteigenschaften (Content-based; z.B. übereinstimmende Schlagwörter oder Ähnliches) für die Verbesserung der Bewertungsmatrix zu nutzen, die dann wiederum im Collaborative Filtering eingesetzt wird.

3.2.4 Aktuelle Forschungsthemen

Empfehlungssysteme werden in der Literatur umfassend behandelt. Zu den aktuellen Forschungsfeldern im Bereich der Empfehlungssysteme zählen unter anderem die nachfolgend aufgelisteten Themen. Sie werden hier im Sinne eines Ausblicks genannt und um die Relevanz der im Projekt PersECA vorgenommenen Untersuchungen zu unterstreichen:

- Einbezug von Kontextinformationen bei der Berechnung der Empfehlungen (z.B. des aktuellen Standorts oder der Tageszeit), um deren Qualität zu erhöhen [z.B. Kwon/Keun Shin 2008; Krüger et al. 2007]
- Messung der Güte einer Bewertungsmatrix: Wie viele Bewertungen benötigt eine Bewertungsmatrix, um gute Empfehlungen zu liefern [Lee et al. 2007]?
- Optimierung der Performance (Laufzeit- und Speicherverhalten des Algorithmus) insbesondere unter Miteinbezug aktueller Daten [z.B. Berkovsky et al. 2008]
- Datenschutzfragen: Algorithmen, bei denen die Profile des Kunden beim Kunden selbst hinterlegt werden und das Empfehlungssystem die Daten anonym abfragt [Castagnos/Boyer 2006; Kobsa 2007]

3.3 Berechnung der Güte von Empfehlungen

Prinzipiell sollte die Güte von Empfehlungen auf zwei unterschiedlichen Wegen beurteilt werden. Zum einen sind die Empfehlungen mit dem gesunden Menschenverstand kritisch zu reflektieren. Diese Aufgabe sollte von Personen durchgeführt werden, die mit den Kundenbedürfnissen vertraut sind und abschätzen können, wann ein Kunde sich für welche Produkte interessieren könnte. In der Regel sind Mitarbeitende aus dem Vertrieb dazu geeignet.

Zum anderen sollte die Güte von Empfehlungen auch objektiv gemessen werden. Hierzu existieren im Wesentlichen zwei Verfahren:

3.3.1 Hitrate-Verfahren

Das Beispiel bezieht sich auf das Item-based Collaborative Filtering. In einem Testset wird die Information über den Kauf eines bestimmten Produktes durch einen bestimmten Kunden bewusst weggelassen. In der folgenden Bewertungsmatrix würde z.B. der Kauf des Produktes 4 durch den Kunden a gelöscht, also der Zelleninhalt auf \emptyset gesetzt (grau hinterlegte Zelle).

Tab. 3.6: Beispiel einer Bewertungsmatrix zur Qualitätssicherung

	Kunde a	Kunde b	Kunde c	...
Produkt 1	1	\emptyset	1	
Produkt 2	\emptyset	\emptyset	1	
Produkt 3	\emptyset	1	\emptyset	
Produkt 4	\emptyset	1	\emptyset	
Produkt 5	1	1	1	
...				

Hierdurch ändert sich der Bewertungsvektor des Kunden a. Nun wird die Ähnlichkeitsmatrix, wie unter 3.2.2.2 beschrieben berechnet. Anschliessend wird die Empfehlungsliste für den Kunden a berechnet. Da der Kunde das Produkt 4 tatsächlich gekauft hatte, sollte dieses Produkt relativ weit oben in der Empfehlungsliste auftauchen.

Das Ergebnis des Tests ist umso besser, je mehr der zuvor aus dem Testset entfernten Produkte in der Empfehlungsliste erscheinen. Der Anteil der zuvor entfernten Produkte an der Empfehlungsliste wird als Trefferquote oder auch „Hitrate“ bezeichnet.

3.3.2 Response-Verfahren

Im laufenden Betrieb wird überprüft, wie die Kunden auf die berechneten Empfehlungen reagieren. Dieses Verfahren ist aufwändiger und erst durchführbar, wenn die Empfehlungen bereits im Shop angezeigt werden. Mit Hilfe von Clickstream-Aufzeichnungen muss dazu die Reaktion der Kunden auf die Empfehlungen festgehalten werden: Wie häufig werden die Empfehlungen angeklickt? Führen die Empfehlungen zu zusätzlichem Umsatz?

3.4 Optimierung von Empfehlungen

Die von den Algorithmen berechneten persönlichen Empfehlungen sind in der Regel noch nicht sinnvoll verwendbar. Sie müssen zunächst im Rahmen des Onlineprozesses optimiert

werden. Dabei werden zum Beispiel Saisonartikel, nicht auf Lager befindliche Artikel oder Artikel, die der Kunde schon gekauft hat, herausgefiltert. Umgekehrt lassen sich Artikel auch gezielt anbieten, z.B. Artikel, die dem Verkäufer eine höhere Marge bringen oder die aus anderen Gründen abverkauft werden sollen. Die Möglichkeiten hierbei, aber auch die Anforderungen an eine sinnvolle Optimierung, sind vielfältig [Leimstoll/Stormer 2007]. In den folgenden Abschnitten werden die wesentlichen Aspekte der Optimierung von Empfehlungen beschrieben.

3.4.1 Alpha-Wert im Item-based-Collaborative-Filtering-Algorithmus

Im Rahmen der PersECA-II-Teilprojekte wurden verschiedene Alpha-Werte für den Item-based-Collaborative Filtering-Algorithmus nach Deshpande und Karypis [2004] getestet. Die Ergebnisse weichen kaum voneinander ab. Mit gut abgestimmten Empfehlungssystemen lassen sich durchaus Trefferquoten in Höhe von 60 % und mehr realisieren [Leimstoll/Alioski 2008; Quade et al. 2008], d.h. 60 % der tatsächlich gekauften Produkte werden empfohlen, wenn man den Kauf zu Testzwecken vorab aus den Daten eliminiert (siehe 3.3.1).

Im Allgemeinen scheinen Alpha-Werte zwischen 0.3 und 0.6 gute Ergebnisse zu bringen. Welcher Alpha-Wert passend ist, muss im Einzelfall geprüft werden. Eine allgemeingültige Empfehlung über die exakte Höhe des Alpha-Wertes kann nicht abgegeben werden. Bislang ist noch unklar, wovon die optimale Höhe des Alpha-Wertes bestimmt wird.

3.4.2 Gute und schlechte Scouts

Eine Methode, um personalisierte Empfehlungen zu verbessern, besteht im Erkennen von so genannten „Scouts“ [Mohan et al. 2007]. Häufig werden für die Berechnung von Empfehlungen alle Produkte herangezogen, die ein Kunde in einem Shop in einem bestimmten Zeitraum gekauft hat. Es kann aber sein, dass nicht alle diese Produkte zu guten Empfehlungen führen. Deshalb wird zwischen guten und schlechten Scouts unterschieden. Gute Scouts haben eine positive Aussagekraft für Empfehlungen, während schlechte Scouts kaum eine oder sogar eine negative Aussagekraft haben.

Schlechte Scouts können "berechnet" und dann bei der Berechnung der Empfehlungen weggelassen werden. Die Basis für die Berechnung der Empfehlungen bilden dann nur die guten Scouts. Der Scout-Ansatz setzt damit eine relativ grosse Datenmenge voraus, weil das Weglassen der schlechten Scouts eine Reduktion der Daten bewirkt, die für die Berechnung von Empfehlungen letztlich genutzt werden können.

Zur Identifizierung der schlechten Scouts muss aus der vorhandenen Bewertungsmatrix – wie schon bei der Hitrate-Berechnung (siehe 3.3.1) – ein Testset extrahiert werden. Auf den übrigen Daten (dem Trainingsset) werden anschliessend die Produktähnlichkeiten berechnet.

3.4.3 Berücksichtigung älterer Daten

Zur Berechnung von Produktähnlichkeiten werden in der Regel möglichst viele Daten benötigt. Häufig besteht das Problem aber nicht in einer zu geringen Anzahl an Datensätzen, sondern in einer zu geringen Besetzung der Bewertungsmatrix. Daher empfiehlt es sich, für die Berechnung der Ähnlichkeitsmatrix alle greifbaren Daten heranzuziehen, auch wenn Käufe schon länger zurückliegen. Dies steigert die Chance, eine Matrix mit höherer Besetzungsdichte zu erhalten. Zur Ermittlung der personalisierten Empfehlung, also für die Berechnung der

persönlichen Top-N-Liste, sollten dann aber nur die Käufe herangezogen werden, die nicht zu weit zurückliegen. Eine Empfehlung für die Länge des optimalen Zeitraums kann nicht gegeben werden, weil zu viele Einflussfaktoren das Ergebnis beeinflussen. Dazu zählen zum Beispiel die Datenlage insgesamt, die Veränderlichkeit der Kaufgewohnheiten im Zeitablauf, die Unterschiede in den Kaufgewohnheiten der einzelnen Kunden oder die Heterogenität des Angebots. Im Einzelfall ist durch Tests zu ermitteln (3.3), welche Datenbasis zu den besten Ergebnissen führt.

3.4.4 Schwellenwerte zur Beurteilung der Güte der Empfehlungen

Um unpassende Empfehlungen zu vermeiden, können Schwellenwerte angegeben werden. Schwellenwerte sagt aus, ab wann eine Ähnlichkeit keine Bedeutung mehr hat. Es werden dann nur Artikel mit einer ausreichend grossen Ähnlichkeit berücksichtigt.

Das Problem dabei ist die Berechnung des Schwellenwertes. Er kann im Prinzip nicht allgemeingültig bestimmt werden, denn eine Ähnlichkeit kann auch bei einem hohen Wert zu unpassenden Empfehlungen führen. Aber es gibt ein anderes Vorgehen, wie man potenziell nicht so gute Ähnlichkeiten ausfiltern kann (siehe dazu 3.4.5).

3.4.5 Regeln zum Aussondern von Ähnlichkeiten

Bisweilen können auch hohe Ähnlichkeitswerte zu ungeeigneten Empfehlungen führen. Dies ist der Fall, wenn Produkte nur sehr selten gekauft wurden, also wenn ein Produkt und das vermeintlich ähnliche Produkt dazu z.B. weniger als drei- bis viermal gekauft wurden; oder wenn zwei Produkte immer nur von den gleichen zwei bis drei Kunden gekauft wurden. In diesen Fällen sagen die Ähnlichkeitswerte wenig aus und sollten ausgefiltert werden. Ein solche Filterung wird idealerweise auf die Ergebnisse des Empfehlungssystems angewendet, bevor die Daten im ERP-System abgespeichert werden.

3.4.6 Produktgruppen berücksichtigen

Bei heterogenen Produktgruppen (z.B. Haushaltsgeräte, Schuhe, Wein in einem Shop) kann es Sinn machen, mit den Empfehlungen innerhalb einer Produktgruppe zu bleiben. Aufgrund des paarweisen Vergleichs bei der Ähnlichkeitsberechnung kann diese wie gewohnt über das gesamte Sortiment durchgeführt werden. Die Auswahl von Produktgruppen, aus denen eine Empfehlung letztlich stammen soll, kann durch Filterung erfolgen, wenn die Top-N-Listen erstellt werden oder auch noch später, sofern die Top-N-Listen zunächst über die gesamte Produktpalette erstellt wurden.

3.4.7 Weitere Optimierungskriterien

Um gute Empfehlungen zu erhalten, kann die Top-N-Liste nach weiteren Kriterien gefiltert werden [Leimstoll/Stromer 2007]. Auch hierbei ist im Einzelfall ist zu prüfen, welche Kriterien Sinn machen:

- Produkte, die der Kunde bereits gekauft hat, werden ausgefiltert. Dies verhindert, dass einem Kunden, der einen Fernseher gekauft hat, erneut ein Fernseher angeboten wird. In der Regel macht es nur in Branchen mit Wiederholungskäufen Sinn, auch Produkte zu empfehlen, die der Kunde schon kennt.
- Saisonprodukte, die nicht der aktuellen Saison entsprechen, werden ausgefiltert. Dies vermeidet, dass im Sommer Streusalz empfohlen wird.
- Es kann Sinn machen, Produkte hoher Unterschiedlichkeit in den Empfehlungen anzuzeigen. Die Top-N-Liste wird dann so umgeordnet, dass Produkte mit hoher Unterschiedlichkeit nach oben rücken [Ziegler et al. 2005].
- Mit Hilfe einer Umordnung der Top-N-Liste können ferner Verkaufsstrategien umgesetzt werden:
 - Marge: Der Verkäufer ist in der Regel daran interessiert, Produkte mit höherer Marge zu empfehlen
 - Qualität: Der Verkäufer ist in der Regel daran interessiert, hochwertige Produkte zu empfehlen, um die Kundenzufriedenheit zu sichern und Servicekosten zu vermeiden.
 - Verfügbarkeit: Der Verkäufer ist in der Regel mehr daran interessiert, Produkte zu verkaufen, die ab Lager verfügbar sind, als Produkte beim Lieferanten neu zu ordern.

3.4.8 Referenzprozesse für die Optimierung persönlicher Empfehlungen

Eine geeignete Reihenfolge der Arbeitsschritte, die im Rahmen der beschriebenen Berechnungs- und Optimierungsprozesse zu durchlaufen sind, lassen sich in Form eines Referenzprozesses darstellen. Er wird in zwei Teilprozesse unterteilt: Die Offline-Berechnung (Abb. 3.3) und die Online-Filterung und -Sortierung (Abb. 3.4).

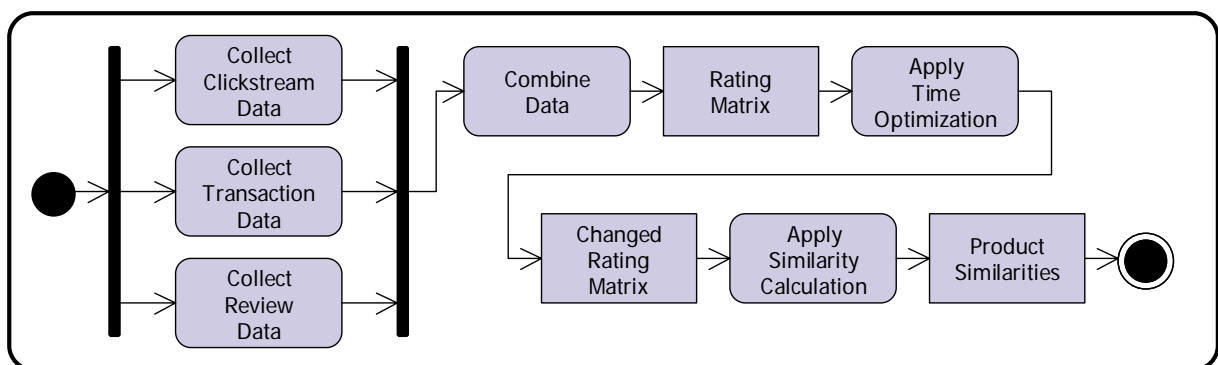


Abb. 3.3: Prozess der Offline-Berechnung

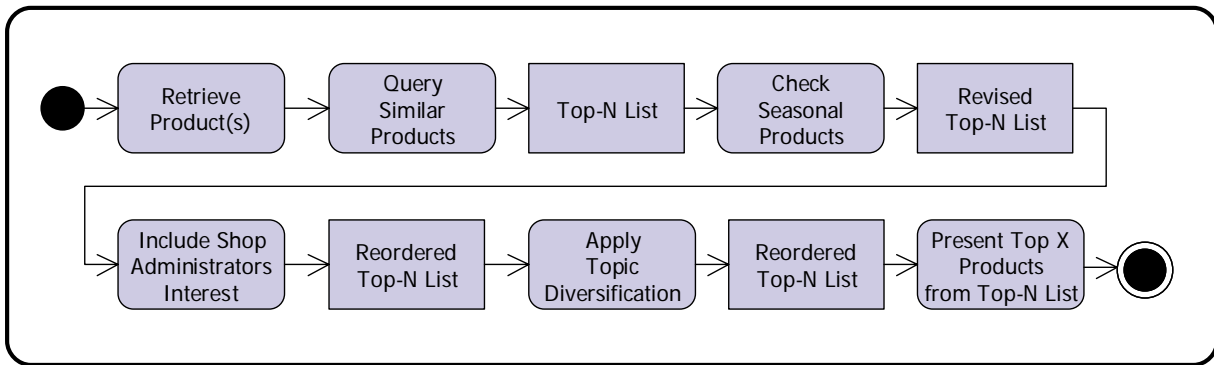


Abb. 3.4: Prozess der Online-Filterung und -Sortierung

4 Demonstrationsanwendung myLandscapes

Um die Unterschiede zwischen Content-based Filtering und Collaborative Filtering zu veranschaulichen, wurde im Rahmen des Projekts PersECA II eine Demonstrationsplattform geschaffen. Die webbasierte Anwendung „myLandscapes“ (Abb. 4.1) zeigt und erläutert am Beispiel einer Fotogalerie, wie Ähnlichkeiten zustande kommen und wie sie sich je nach Berechnungsverfahren unterscheiden.



The screenshot shows the homepage of myLandscapes.ch. At the top, there are navigation links for 'DE | EN | Home | Kontakt' and the logo for 'n|w Fachhochschule Nordwestschweiz Hochschule für Wirtschaft'. The main header features the site name 'myLandscapes.ch' and a subtitle 'personalisierungs demonstrator | beta 0.8'. Below the header is a large image of tall grasses against a sunset sky. To the right of the image is a login form with fields for 'Username' and 'Password', a 'GO' button, and a link for 'jetzt registrieren'. A red error message below the form reads 'Benutzername oder Passwort nicht korrekt'. Below the image, there is a welcome message: 'Willkommen bei myLandscapes.ch, dem Personalisierungs-Demonstrator des Competence Center E-Business Basel der Fachhochschule Nordwestschweiz'. To the right of this is a section titled 'Hallo! Das erste Mal hier?' which explains that users must register and rate ten photos to use the service.

Abb. 4.1: Fotogalerie myLandscapes

myLandscapes ist unter der folgenden URL zugänglich:

www.mylandscapes.ch

Benutzer müssen ein Konto eröffnen und mindestens zehn Fotos bewerten. Anschliessend werden über Nacht Ähnlichkeiten zwischen Bildern und zwischen Usern berechnet. Am nächsten Tag kann der Benutzer seine persönlichen Empfehlungen einsehen und nachvollziehen, weshalb ihm diese Empfehlungen gemacht werden.

5 Der Profil-Browser

5.1 Sinn und Zweck des Profil-Browsers

Der Profil-Browser ist ein webbasiertes Tool, das Informationen aus Transaktionsdaten grafisch oder tabellarisch aufbereitet. Dies ermöglicht eine Übersicht beispielsweise über das Kaufverhalten von Kunden oder die Anzahl von Produkten, die pro Einkauf eingekauft werden. Aus den Transaktionsdaten können ausserdem Produkt-Empfehlungen berechnet und dargestellt werden.

Die Motivation für die Entwicklung des Profil-Browsers war der Wunsch, Profildaten sichtbar zu machen. Normalerweise sind Profildaten in komplexen Systemen, wie etwa ERP-Systemen – gespeichert und daher nur einem ausgewählten Nutzerkreis zugänglich. Dies gilt auch für die in den PersECA-Projekten aufgebauten Kundenprofile. Ohne eine Sicht auf diese Profile wäre auch nicht erkennbar, zu welchen Ergebnissen beispielsweise die Empfehlungsalgorithmen führen, die im Rahmen des Projekts entwickelt und eingesetzt wurden.

Mit dem Profil-Browser sollte folglich eine Applikation geschaffen werden, die mit Hilfe der entwickelten Algorithmen aus den Transaktionsdaten (Inputprofile) Ergebnisse berechnet und darstellt (Outputprofile). Diese Applikation steht als Beispielanwendung auch den Projektpartnern zur Verfügung. Sie kann quasi als eine kleine Business-Intelligence-Anwendung betrachtet werden, die Transaktionsdaten verdichtet und nach verschiedenen Dimensionen darstellt.

Anwender können eigene Transaktionsdaten in den Profil-Browser einlesen und daraus Auswertungen erstellen oder Empfehlungen berechnen. Diese Daten können anschliessend zum Beispiel im Telefonmarketing genutzt werden. Ruft ein Kunde an, kann im Profil-Browser nachgesehen werden, was er in einem bestimmten Zeitraum gekauft hat, in welchen Mengen er bestellt, wie gross sein Umsatz ist etc. Mit den berechneten Empfehlungen können ihm auch Vorschläge gemacht werden, welche Produkte für ihn interessant sein könnten.

Der Profil-Browser ist unter der folgenden URL zugänglich:

www.fhnw.ch/wirtschaft/iwi/profil-browser

5.2 Aufbau der Applikation

Der Profil-Browser ist modular aufgebaut und lässt sich dadurch gut erweitern. Die meisten Funktionen sind im Source-Code dokumentiert und als JavaDoc verfügbar. Dieses Kapitel gibt einen Überblick über den Aufbau des Profil-Browsers und der verwendeten Technologien. Ausserdem werden die einzelnen (Auswertungs-)Funktionen beschrieben und aufgelistet, welche Java-Klassen zu den Funktionen gehören. Dies ist für Entwickler hilfreich, die wissen wollen, welche Klassen in welchen Funktionen verwendet werden.

Dieses Kapitel wendet sich hauptsächlich an Entwickler und Personen, die sich mit den verschiedenen Java-Technologien auskennen. Einzelne Begriffe werden deshalb vorausgesetzt und nicht näher erläutert.

5.2.1 Verwendete Technologien

Der Profil-Browser baut auf dem Webapplikations-Framework Struts 2 auf. Struts ist ein Open Source Framework für Java-Webanwendungen. Es verarbeitet HTTP-Anfragen in einem standardisierten Prozess und unterstützt so die Entwicklung von Webanwendungen. Das Ziel von Struts ist die Trennung der Präsentations-, Daten- und Anwendungslogik.

Tab. 5.1: Übersicht über die verwendeten Technologien

Applikationsserver	Apache Tomcat 5.5.23
Datenbank	MySQL 5.1.11
Datenbankanbindung	JDBC
Programmiersprachen	JavaSE 1.5, JSP, Ajax, JavaScript, HTML
Webapplikations-Framework	Apache Struts 2
Java Chart Library	JFreeChart 1.0.8

Das zentrale Steuerelement bei Struts2 ist das struts.xml. In diesem wird definiert, welche Action-Klasse nach einem Aufruf aus der JSP-Datei aufgerufen wird.

5.2.2 Datenbankmodell

Die Struktur der Daten lässt sich in drei Tabellen abbilden. Abb. 5.1 zeigt das daraus entstehende Datenbankmodell.

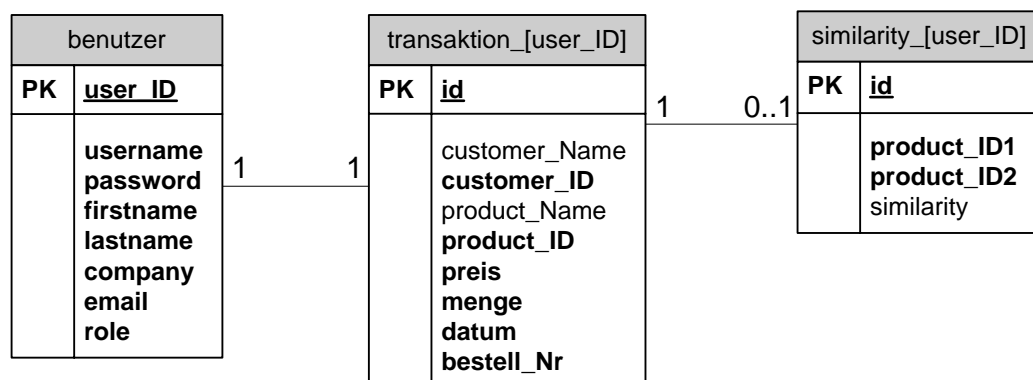


Abb. 5.1: Datenbankmodell

Tabelle benutzer: In dieser Tabelle sind die Anwender gespeichert, die den Profil-Browser benutzen dürfen. Der Profil-Browser kontrolliert beim Login, ob die Person in dieser Tabelle erfasst ist und ob das Passwort korrekt ist.

Tabelle transaktion_[Nummer]: Diese Tabelle wird automatisch erstellt, sobald ein Benutzer Transaktionsdaten in den Profil-Browser lädt. Die Daten werden in dieser Tabelle gespeichert und bilden damit die Basis für die verschiedenen Auswertungen. Die [Nummer] ist die User_ID des Benutzers. Pro Benutzer kann eine Tabelle transaktion_[Nummer] erstellt werden.

Tabelle similarity_[Nummer]: Diese Tabelle wird automatisch erstellt, sobald ein Benutzer aus seinen Transaktionsdaten Empfehlungen berechnen lässt. Die berechneten Ähnlichkeiten werden in dieser Tabelle abgelegt. Die [Nummer] ist die User_ID des Benutzers. Pro Benutzer kann eine Tabelle transaktion_[Nummer] erstellt werden.

5.2.3 Funktionen und ihre Klassen

Die folgenden Unterkapitel beschreiben die einzelnen Funktionen des Profil-Browsers. Die Klassen, die für die einzelnen Funktionen erforderlich sind, werden den Bereichen grob zugeordnet.

5.2.3.1 Login, Logout und Benutzerverwaltung

Neue Benutzer müssen in der aktuellen Version des Profil-Browsers direkt in der MySQL-Tabelle „Benutzer“ angelegt werden. Für das Login werden der Benutzername sowie das Passwort benötigt. Der Benutzername muss eindeutig sein. Die Anwender können alle ihre Daten mit Ausnahme des Benutzernamens selbst ändern.

Wenn das Login erfolgreich ist, wird eine Session geöffnet. Diese wird entweder beim Logout oder nach 20 Minuten Inaktivität gelöscht. Anschliessend muss der Benutzer sich wieder neu anmelden. Die Session ermöglicht die eindeutige Identifizierung des Benutzers während der Dauer der Sitzung.

Benötigte Klassen: LoginAction, LogoutAction, User und EditUserAction.

5.2.3.2 Import und Verwaltung von Transaktionsdaten

Es ist möglich, Dateien sowohl als CSV (Comma Separated Value) oder als XML-Datei zu importieren.

Import von Daten im XML CP/2-Format

Für den Import von Daten aus einer XML-Datei muss diese im CP/2-Format vorliegen. Dieses Format wurde im Rahmen des PersECA-II-Projektes definiert (2.2). Dieses Format beschreibt einen XML-Standard für die Beschreibung von Transaktionsprofilen.

Import von Daten im CSV-Format

Für den Import von Daten im CSV-Format müssen die Daten mit einem Semikolon (;) von einander getrennt sein (Excel-CSV-Standard). Der Aufbau der CSV-Datei ist folgendermassen:

Tab. 5.2: Aufbau einer CSV-Datei für den Import

Kundenname	Kunden-ID	Produktname	Produkt-ID	Produktpreis	Menge	Kaufdatum	Bestellnummer
H. Mustermann	159874	Stradivari	65402	15862.00	1	11.02.2008	5623
H. Mustermann	159874	Kopierpapier 80g Naturweiss	15001	2.50	4	11.02.2008	5623
Emilie Kaufgern	159821	Büroschere gross, weiss	12015	13.75	1	28.03.2008	6589

Die Angabe von „Kundenname“ und „Produktname“ ist freiwillig, wird jedoch zwecks einer besseren Verständlichkeit der Ergebnisse empfohlen. Sind keine Namen vorhanden, werden sie mit den jeweiligen IDs ersetzt.

Falls nur das Empfehlungssystem verwendet werden soll, ist ausserdem der Import von Produktpreis, Menge, Kaufdatum sowie Bestellnummer freiwillig. Die entsprechenden Felder werden dann mit Zufallsdaten aufgefüllt.

Anmerkungen zum Import

Die CSV-Dateien werden über den SQL-Befehl 'LoadDataInLine' eingelesen. Mit diesem können grosse Datensätze in kürzester Zeit eingelesen werden. Die XML-Datei wird hingegen geparkt. Dadurch dauert der Import von XML-Dateien bei grossen Dateien um ein vielfaches länger als der Import von CSV-Dateien.

Benötigte Klassen: UploadAction, InputProfile, InputProfile2CSV, InputProfileReaderXML, ImportCSV2DB, DeleteTransactionsAction

5.2.3.3 Verwaltung und Berechnung von Empfehlungen

Die Empfehlungen werden nach dem Deshpande/Karypis-Algorithmus berechnet (3.2.2.2). Die Ähnlichkeiten werden in einer eigenen Tabelle abgelegt. Werden sie neu berechnet, werden die zuvor berechneten Ähnlichkeiten automatisch gelöscht. Dabei wird die Tabelle gelöscht und anschliessend neu angelegt. Der Benutzer kann die Tabelle mit den Ähnlichkeiten aus dem Profil-Browser heraus löschen, wenn keine Berechnungen mehr abgerufen werden sollen.

Benötigte Klassen: ComputeRecommendationsAction, ComputeRecommendationThread, DeleteRecommendationsAction.

5.2.3.4 Auswertung und Darstellung der Transaktionsdaten und Empfehlungen

Transaktionsstatistik

- Anzahl Transaktionen pro Zeitraum: Grafik mit allen Transaktionen (über alle Kunden und alle Produkte) in einen bestimmten Zeitraum

- Umsatz pro Zeitraum: Grafische Darstellung des Gesamt-Umsatzes (Anzahl verkaufter Produkte multipliziert mit dem jeweiligen Preis) aller Kunden in einen bestimmten Zeitraum
- Warenkorb nach Zeitraum: Durchschnittlicher Wert eines Warenkorbes aller Kunden in einem bestimmten Zeitraum. Der Warenkorbwert wird folgendermassen berechnet: Umsatz (siehe oben) geteilt durch die Anzahl Transaktionen.

Transaktionen pro Produkt nach Zeitraum

- Menge pro Zeitraum: Anzahl der Verkäufe eines einzelnen Produkts innerhalb eines bestimmten Zeitraums
- Menge pro Transaktion: Anzahl der Produkte, die pro Transaktion verkauft werden

Transaktionen pro Kunde nach Zeitraum

- Anzahl Transaktionen pro Zeitraum: Anzahl der Transaktionen, die ein einzelner Kunde in einem bestimmten Zeitraum getätigt hat.
- Umsatz pro Zeitraum: Umsatz eines einzelnen Kunden in einem bestimmten Zeitraum.
- Warenkorb nach Zeitraum: Der durchschnittliche Wert des Warenkorbs eines einzelnen Kunden in einem bestimmten Zeitraum.

Produkte pro Kunde im Detail

- Auflistung der Produkte, der Menge, des Preises und des Kaufdatums der Einkäufe eines bestimmten Kunden

Kunde pro Produkt im Detail

- Auflistung, welche Kunden an welchem Datum, zu welcher Menge und zu welchem Preis ein bestimmtes Produkt gekauft haben.

Empfehlungssystem

- Auflistung der Produkte, die einem ausgewählten Produkt ähnlich sind.

Benötigte Klassen: DBQueries, DBUtilities, StatistikData, alle stat_*_*Action-Klassen

5.3 Ausblick auf die Erweiterung des Profil-Browsers

Dank des modularen Aufbaus lässt sich der Profil-Browser gut erweitern und um zusätzliche Auswertungen oder Funktionen ergänzen. Nachfolgend ein paar Ideen, mit welchen Funktionen der Profil-Browser in einer nächsten Version erweitert werden könnte.

Weitere Algorithmen zur Berechnung von Empfehlungen

In der aktuellen Version des Profil-Browsers werden die Empfehlungen nach dem Algorithmus von Deshpande berechnet. Es gibt eine Vielzahl von weiteren Algorithmen, die implementiert

werden können und die je nach Einsatzart oder -gebiet unter Umständen bessere Empfehlungen generieren.

Um weitere Algorithmen zu implementieren, muss ein neuer Thread analog dem „ComputeReccomendationsThread“ angelegt werden, der den neuen Algorithmus beinhaltet. Über den Menüpunkt Empfehlungen muss der Benutzer die Möglichkeit haben, einen oder mehrere der Algorithmen auszuwählen. Diese Auswahl kann anschliessend in der „ComputeRecommendationsAction“ abgefragt werden, und der oder die verschiedenen Threads gestartet werden. Da die Algorithmen in verschiedenen Threads berechnet werden, können mehrere verschiedene Berechnungen gleichzeitig durchgeführt werden. Die Resultate werden anschliessend in einer eigenen Tabelle nach dem Muster 'similarity_xyz_v' abgelegt, wobei xyz der Name des Algorithmus ist und v die ID des Kunden.

Anmerkung: In der aktuellen Version wird die Tabelle nach dem Muster 'similarity_v' gespeichert, also ohne Name des Algorithmus.

Variabler Alpha-Wert

Der Alpha-Wert, der zur Berechnung der Ähnlichkeiten nach Deshpande benötigt wird und die Qualität der Empfehlungen beeinflusst, ist zurzeit fest auf 0.5 gesetzt. Mehrere Tests haben gezeigt, dass dieser Wert in der Regel zu den besten Ergebnissen führt. In einer zukünftigen Version des Profil-Browsers könnte dieser Wert durch den Anwender oder Administrator frei gesetzt werden, bevor die Berechnungen gestartet werden.

Testsets

Ein Testset ist eine zufällig ausgewählte Untermenge der Transaktionsdaten. Die Daten des Testsets werden von den Ähnlichkeitsberechnungen ausgenommen. Nach dem Berechnen der Ähnlichkeiten mit den verbleibenden Transaktionsdaten wird überprüft, ob die Daten aus dem Testset empfohlen werden. So könnte die Güte der Empfehlungen überprüft werden.

Benutzerverwaltung

Das Hinzufügen von neuen Benutzern erfolgt zurzeit noch direkt in der MySQL-Datenbank. Es ist vorgesehen, dass in einer zukünftigen Version ein oder mehrere Administratoren neue Benutzer über die Weboberfläche administrieren können. In der Tabelle „Benutzer“ gibt es dazu schon das Feld „Rolle“, über das mit einem Integer-Wert verschiedene Rollen und Rechte gesteuert werden könnten. Denkbar sind Rollen wie zum Beispiel Administrator, Superuser und User. Der Administrator könnte neue Benutzer anlegen und löschen, Passwörter zurücksetzen oder Rollen vergeben. Der User könnte Daten ansehen und der Superuser Transaktionsdaten einlesen, Empfehlungen berechnen und löschen etc.

Momentan kann pro Datenbasis nur ein Benutzer/Login auf die Daten zugreifen. Dies könnte in einer folgenden Version so geändert werden, dass mehrere Anwender mit ihrem eigenen Login auf die Daten zugreifen können.

Dazu könnten in der Datenbank eine Tabelle Gruppe angelegt werden, in der die Anwender einer oder mehreren Gruppen hinzugefügt würden. Anstelle der UserID würde dann die GruppenID verwendet, um Tabellen vom Typ transaktion_[GruppenID] und similarity_[GruppenID] anzulegen.

Weitere Auswertungen hinzufügen

Durch den modularen Aufbau des Profil-Browsers ist es einfach, neue Auswertungsmodule hinzuzufügen. Wenn zusätzliche Daten zur Verfügung stünden, könnten zusätzliche Auswertungsmodule geschrieben werden, die auf diese Daten zugreifen, auswerten und darstellen.

6 Zusammenfassung und Schlussfolgerungen

Einen Schwerpunkt im Projekt PersECA II bildete die Entwicklung von Empfehlungssystemen zur Förderung des Cross- und Up-Selling. Als Lösungsansatz erwies sich das Collaborative Filtering in einer personalisierten Variante als besonders geeignet: Es erlaubt eine Aufteilung der Berechnungen in Offline- und Onlineprozesse, es erzielt schon mit einfachen Transaktionsdaten gute Ergebnisse, seine Anforderungen an die Datenpflege sind gering und nach der Initialisierung läuft es weitgehend automatisch ab.

Der Initialaufwand besteht zunächst in der Auswahl und Implementierung eines geeigneten Berechnungsalgorithmus. Anschliessend muss in jedem konkreten Fall systematisch getestet werden, mit welchen Daten die besten Ergebnisse erzielt werden und auf welcher Basis (z.B. Transaktionshistorie, Warenkorbinhalt, Clickstream) die Empfehlungen personalisiert werden sollen. Als besonders wichtig stellte sich heraus, dass die vom System berechneten Ergebnisse in einem letzten Schritt mit Hilfe von Filterregeln noch zu optimieren sind. Dabei werden zum Beispiel Saisonartikel, nicht auf Lager befindliche Artikel oder Artikel, die der Kunde schon gekauft hat, herausgefiltert. Umgekehrt lassen sich Artikel auch gezielt anbieten, z.B. Artikel, die dem Verkäufer eine höhere Marge bringen oder die aus anderen Gründen abverkauft werden sollen.

Insgesamt konnte festgestellt werden, dass die Implementierung und Optimierung eines Empfehlungssystems einen aufwendigen Prozess darstellt. Die beschriebenen Profile, Technologien und Methoden können für den Aufbau eines solchen Systems herangezogen werden. Der formulierte Referenzprozess erleichtert die Planung und Durchführung der nötigen Schritte.

Literaturverzeichnis

- Berkovsky, Shlomo; Eytani, Yaniv; Manevitz, Larry (2008): Retrieval of collaborative filtering nearest-neighbors in a content-addressable space, in: Manolopoulos, Yannis; Filipe, Joaquim; Constantopoulos, Panos; Cordeiro, José (eds.): Enterprise Information Systems, 8th International Conference, ICEIS 2006, Paphos, Cyprus, May 23-27, 2006, Revised Selected Papers, Berlin, Heidelberg: Springer, 2008.
- Castagnos, Sylvain; Boyer, Anne (2006): From implicit to explicit data: A way to enhance privacy, in: Workshop on Privacy-Enhanced Personalization PEP2006, April 2006, Montreal, Canada.
- Deshpande, Mukund; Karypis, George (2004): Item-based top-n recommendation algorithms, in: ACM Transactions on Information Systems, 22(1), 143-177, 2004.
- Harold, Elliotte Rusty (2003): Effective XML: 50 Specific Ways to Improve Your XML, Addison-Wesley Professional 2003.
- Kobsa, Alfred (2007): Privacy-enhanced personalization, in: Communications of the ACM 50(8), August 2007, 24-33.
- Krüger, Antonio; Baus, Jörg; Heckmann, Dominik; Kruppa, Michael (2007): Adaptive Mobile Guides. in: Brusilovsky, Peter; Kobsa, Alfred; Nejdl, Wolfgang (eds.): The Adaptive Web, Berlin, Heidelberg: Springer, 2007.
- Kwon, Ohbyung; Keun Shin, Myung (2008): LACO: A location-aware cooperative query system for securely personalized services, in: Expert Systems with Applications 34(4), May 2008, 2966-2975.
- Lee, Hong Joo; Kim, Jong Woo; Park, Sung Joo (2007): Understanding collaborative filtering parameters for personalized recommendations in e-commerce, in: Electronic Commerce Research 7(3-4), December 2007, 293-314.
- Leimstoll, Uwe; Alioski, Adrian (2008): Personalisierung im B2B-Werkzeughandel: Entwicklung neuer Funktionen für den E-Shop von Brüttsch/Rüegger Tools, Basel: Fachhochschule Nordwestschweiz, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik, Arbeitsberichte der Hochschule für Wirtschaft FHNW, Arbeitsbericht Nr. 32 des Competence Center E-Business Basel, 2008.
- Leimstoll, Uwe; Alioski, Adrian; Risch, Daniel (2008): Personalisierungsfunktionen im E-Commerce: Eine Systematisierung mit Beispielen, Basel: Fachhochschule Nordwestschweiz, Institut für Wirtschaftsinformatik IWI, Arbeitsberichte der Hochschule für Wirtschaft FHNW, Arbeitsbericht Nr. 36 des Competence Center E-Business Basel, 2008.
- Leimstoll, Uwe; Stormer, Henrik (2007): Collaborative recommender systems for online shops, in: Proceedings of the 13th Americas Conference on Information Systems, August 9-12, 2007, Keystone, Colorado.
- Mohan, Bharath Kumar; Keller, Benjamin J.; Ramakrishnan, Naren (2007): Scouts, promoters, and connectors: The roles of ratings in nearest-neighbor collaborative filtering, in: ACM Transactions on the Web TWEB 1(2), August 2007.
- Ogbuji, Uche (2004): Principles of XML design: When to use elements versus attributes, IBM Research [<http://www.ibm.com/developerworks/xml/library/x-eleatt.html>, abgerufen am 24. September 2007].

- Quade, Michael; Stormer, Henrik; Schneider, Raoul; Merz, Jürg (2008): Entwicklung und Umsetzung eines Systems für personalisierte Empfehlungen in einem B2B-E-Shop, Basel: Fachhochschule Nordwestschweiz, Institut für Wirtschaftsinformatik IWI, Arbeitsberichte der Hochschule für Wirtschaft FHNW, Arbeitsbericht E-Business Nr. 31, 2008.
- Resnick, Paul; Iacovou, Neophytos; Suchak, Mitesh; Bergstrom, Peter; Riedl, John (1994): GroupLens: An open architecture for collaborative filtering of netnews, in: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, October 1994, 22 - 26.
- Riecken, Doug (2000): Personalized views of personalization, in: Communications of the ACM 43(8), 2000.
- Risch, Daniel (2008): Nutzung von Kundenprofilen im E-Commerce: Dargestellt am Beispiel des B2C E-Commerce in der Schweiz, Lohmar – Köln: Josef Eul Verlag, 2008.
- Risch, Daniel; Schubert, Petra (2005): Customer profiles, personalization and privacy, in: Proceedings of COLLECTeR Europe 2005, June 2005, 1-12.
- Schubert, Petra; Kummer, Mathias; Leimstoll, Uwe (2006): Legal requirements for the personalization of commercial Internet applications in Europe, in: Journal of Organizational Computing and Electronic Commerce (JOCEC), Vol. 16, Issue 3/4, 2006, S. 201-221.
- Schubert, Petra; Leimstoll, Uwe (2002): Handbuch zur Personalisierung von E-Commerce-Applikationen, Basel: Fachhochschule beider Basel (FHBB), Institut für angewandte Betriebsökonomie (IAB), Arbeitsbericht E-Business Nr. 7, 2002.
- Wu, Dezhi; Im, Il; Tremaine, Marilyn; Instone, Keith; Turoff, Murray (2003): A framework for classifying personalization scheme used on e-commerce websites, in: Proceedings of the 36th Hawaii International Conference on System Sciences, HICSS'03, Hawaii, 2003.
- Ziegler, Cai-Nicolas; McNee, Sean M.; Konstan, Joseph A.; Lausen, Georg (2005): Improving recommendation lists through topic diversification, in: Proceedings of the 14th international conference on World Wide Web (WWW), 2005.

In der Reihe bereits erschienen

In der Reihe „Arbeitsberichte des Competence Center E-Business Basel“ sind bisher unter anderem die folgenden Titel erschienen:

- Tanner, Christian (2003): E-Procurement-Studie: E-Supplier - Situationsaufnahme bei E-Procurement-Betreibern, Basel: Fachhochschule beider Basel (FHBB), Institut für angewandte Betriebsökonomie (IAB), Arbeitsbericht E-Business Nr. 18, 2003.
- Hügli, Raphael; Schubert, Petra (2007): Billing Studie 2006 - Debitorenmanagement im Schweizer Gesundheitswesen, Basel: Fachhochschule Nordwestschweiz, HSW Basel (FHNW), Institut für Wirtschaftsinformatik (IWI), Arbeitsbericht E-Business Nr. 27, 2007.
- Risch, Daniel (2007): Kundenprofile im E-Commerce, Basel: Fachhochschule Nordwestschweiz FHNW, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik IWI, Arbeitsbericht E-Business Nr. 29, 2007.
- Quade, Michael; Stormer, Henrik; Schneider, Raoul; Merz, Jürg (2008): Entwicklung und Umsetzung eines Systems für personalisierte Empfehlungen in einem B2B-E-Shop, Basel: Fachhochschule Nordwestschweiz, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik, Arbeitsberichte der Hochschule für Wirtschaft FHNW, Arbeitsbericht Nr. 31 des Competence Center E-Business Basel, 2008.
- Leimstoll, Uwe; Alioski, Adrian (2008): Personalisierung im B2B-Werkzeughandel: Entwicklung neuer Funktionen für den E-Shop von Brüttsch/Rüegger Tools, Basel: Fachhochschule Nordwestschweiz, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik, Arbeitsberichte der Hochschule für Wirtschaft FHNW, Arbeitsbericht Nr. 32 des Competence Center E-Business Basel, 2008.
- Quade, Michael; Alioski, Adrian (2008): Entwicklung eines Konzepts für einen personalisierten Newsletter mit Empfehlungen, Basel: Fachhochschule Nordwestschweiz, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik, Arbeitsberichte der Hochschule für Wirtschaft FHNW, Arbeitsbericht Nr. 33 des Competence Center E-Business Basel, 2008.
- Pülz, Michael (2008): Entwicklung eines konfigurierbaren Standard-Empfehlungssystems für die Shop-Software von myfactory.BusinessWorld, Basel: Fachhochschule Nordwestschweiz, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik, Arbeitsberichte der Hochschule für Wirtschaft FHNW, Arbeitsbericht Nr. 34 des Competence Center E-Business Basel, 2008.
- Alioski, Adrian; Leimstoll, Uwe (2008): Entwicklung innovativer Personalisierungsfunktionen für den Onlineshop von buch.ch, Basel: Fachhochschule Nordwestschweiz, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik, Arbeitsberichte der Hochschule für Wirtschaft FHNW, Arbeitsbericht Nr. 35 des Competence Center E-Business Basel, 2008.
- Alioski, Adrian; Leimstoll, Uwe; Risch, Daniel (2008): Personalisierungsfunktionen im E-Commerce: Eine Systematisierung mit Beispielen, Basel: Fachhochschule Nordwestschweiz, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik, Arbeitsberichte der Hochschule für Wirtschaft FHNW, Arbeitsbericht Nr. 36 des Competence Center E-Business Basel, 2008.

Weitere Publikationen unter: www.hsw-basel.ch/iwi/publications.nsf