

**Arbeitsberichte der
Hochschule für Wirtschaft FHNW – Nr. 9**

Entwicklung eines konfigurierbaren Standard-Empfehlungssystems für die Shop-Software von myfactory.BusinessWorld

Arbeitsbericht Nr. 34 des
Competence Center E-Business Basel

Michael Pülz

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

ISBN Nr. 978-3-03724-102-8

Institut Institut für Wirtschaftsinformatik IWI

Datum Oktober 2008

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

Mit freundlicher
Unterstützung von

KTI/CTI

Projektpartner

bbv
behind things.



Vorwort

Die Bedeutung der Personalisierung nimmt im E-Commerce immer mehr zu [Schubert/Leimstoll 2002]. Die ursprüngliche Idee bestand darin, den konventionellen "Tante-Emma"-Laden mit seiner persönlichen Kundenberatung im Internet nachzuahmen. Hieraus entstand eine Fülle von Möglichkeiten, Inhalte und Funktionen einer Website an die individuellen Bedürfnisse eines Benutzers anzupassen. Diese Anpassung dient nicht nur der Verbesserung der Mensch-Maschine-Interaktion, sondern bietet dem Nutzer zusätzliche Informationen und Prozesse, die für ihn einen Mehrwert generieren sollen. Gleichzeitig erlaubt die Personalisierung die Umsetzung von Verkaufsstrategien, wie zum Beispiel das Cross- und Up-Selling.

Seit mehreren Jahren führt das Institut für Wirtschaftsinformatik an der Hochschule für Wirtschaft der Fachhochschule Nordwestschweiz Projekte zur Personalisierung von E-Commerce-Applikationen (PersECA) durch. Ziel der Projekte ist es, in Zusammenarbeit mit Wirtschaftspartnern innovative Lösungen zu erarbeiten, die im elektronischen Verkaufskanal zu verbesserten Kundenbeziehungen beitragen und einen zusätzlichen Nutzen sowohl für die Kunden als auch für das anbietende Unternehmen schaffen können.

Das Projekt PersECA II wird 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.

In der vorliegenden Fallstudie mit der Firma bbv Software Services AG mit Hauptsitz in Luzern wurden neue Personalisierungsfunktionen entwickelt, die in der ERP-II-Software myfactory umgesetzt wurden. Die vorliegende Fallstudie dokumentiert das Vorgehen im Projekt und die erarbeiteten Personalisierungslösungen im Detail.

Grossen Dank richten wir an unsere Projektpartner. Unser Dank geht an die bbv Software Services AG, insbesondere an Urs Gehrig und Adrian Bachofen. Der Information Systems Research Group an der Universität Fribourg, insbesondere Prof. Dr. Andreas Meier und Dr. Henrik Stormer, danken wir für die kompetente Unterstützung bei der Entwicklung innovativer Lösungsansätze. Prof. Dr. Uwe Leimstoll, Competence Center E-Business Basel, der die Gesamtverantwortung für das Projekt PersECA II trägt, danken wir für die inhaltliche und methodische Unterstützung.

Basel, im Oktober 2008

Michael Pülz

Inhaltsverzeichnis

Vorwort	ii
Inhaltsverzeichnis	iii
Abbildungs- und Tabellenverzeichnis	vi
Abkürzungsverzeichnis	vii
1 Einleitung	1
2 Das Unternehmen: bbv Software Services AG	2
2.1 Hintergrund	2
2.2 Branche, Produkte, Zielgruppen und Märkte	2
2.2.1 Softwareentwicklung	3
2.2.2 Softwarequalität	4
2.2.3 Softwaremanagement	4
2.3 Partner	4
2.3.1 Allgemeine Partner	4
2.3.2 Development Partner von myfactory	5
2.4 Unternehmensvision	5
2.5 Motivation für die Teilnahme am Personalisierungsprojekt PersECA II	5
3 IST-Situation: ERP-II-Software vor dem Personalisierungsprojekt	7
3.1 Leistungsumfang der Software	7
3.2 Elektronischer Produktkatalog	7
3.3 Technologie	10
3.3.1 Software-Lösung/Programmierung	10
3.3.2 IT-Architektur	10
4 Soll-Konzept: ERP-II-Software mit Personalisierungsfunktionalität	12
4.1 Auswahl der Personalisierungsfunktionen	12
4.1.1 Diskussion der Personalisierungslandkarte	12
4.1.2 Schwerpunkt: Personalisierung während der Informationsphase	12
4.1.3 Notwendige Erweiterungen der Kundenprofildaten	13
4.2 Erwarteter Nutzen	13
4.3 Erwarteter Aufwand und Risiken	13
5 Detailkonzept und Umsetzung	14
5.1 Allgemeine Rahmenbedingungen	14
5.2 Auswertungsmethoden und Algorithmen für Empfehlungssysteme	14
5.2.1 Content-based Filtering	15

5.2.2	Collaborative Filtering.....	15
5.2.2.1	User-based Collaborative Filtering	16
5.2.2.2	Item-based Collaborative Filtering.....	18
5.2.3	Hybride Systeme	22
5.2.4	Aktuelle Forschungsthemen	22
5.2.5	Qualitätssicherung.....	22
5.3	Ablauf der Berechnung von Empfehlungen	23
5.3.1	Input-Profile	23
5.3.1.1	Aufzeichnung des Clickstreams	23
5.3.1.2	Füllen der Rating-Matrix	24
5.3.2	Berechnung der Ähnlichkeitsmatrix.....	24
5.3.3	Output-Profile	24
5.4	E-Shop-Klassifizierung und -Parametrisierung für ein Standard- Empfehlungssystem.....	25
5.4.1	Fragestellung.....	25
5.4.2	Klassifizierung von E-Shops.....	25
5.4.3	Zusammenführung von Shop-Kategorien und Empfehlungs- Parametern.....	27
5.4.4	Parametrisierung	29
5.5	Programmierung und Umsetzung	31
5.5.1	Umsetzung des Item-based Collaborative Filtering.....	31
5.5.2	Implementierung von Cookies.....	31
5.5.3	Implementierung der Clickstream-Aufzeichnung.....	32
5.5.4	Empfehlungen in der myfactory.BusinessWorld-Software	32
5.5.5	Implementierung der Parametereinstellungen bei der Konfiguration eines E-Shops	34
5.5.6	Weitere Implementierungen	41
5.5.7	Entwicklungsumgebung	41
5.5.8	Entwicklungssoftware	42
5.6	Feinabstimmung und potenzielle Verbesserungen.....	44
5.6.1	Alpha-Wert im Item-based-Collaborative-Filtering-Algorithmus	44
5.6.2	Gute und schlechte Scouts	44
5.6.3	Berücksichtigung älterer Daten	45
5.7	Vermarktung	45
6	Projektmanagement	48
6.1	Ablauf des Gesamtprojekts.....	48
6.2	Vorgehensweise im Teilprojekt mit bbv.....	48
6.3	Lessons Learned	50

6.3.1	Schwierigkeiten bei der Erarbeitung und Implementierung	50
6.3.2	Begünstigende Faktoren	51
7	Zusammenfassung und Schlussfolgerungen	52
	Literaturverzeichnis	53

Abbildungs- und Tabellenverzeichnis

Abb. 3.1:	Beispiel eines Artikelkatalogs in myfactory.....	8
Abb. 3.2:	Beispiel einer Gruppenhierarchie im Artikelkatalog.....	9
Abb. 3.3:	Beispiel der Erfassung eines Verkaufs.....	10
Abb. 5.1:	Im Internet gekaufte Warengruppen [PayPal 2007].....	27
Abb. 5.2:	Mindmap zur Zuordnung von E-Shop-Kategorien und Parametern.....	28
Abb. 5.3:	Persönliche Empfehlungen.....	33
Abb. 5.4:	Zusatzmodul Empfehlungssystem.....	34
Abb. 5.5:	Kategorisierung des E-Shops.....	35
Abb. 5.6:	Anpassung der Parameter für die Filterung der Empfehlungen.....	36
Abb. 5.7:	Einstellung weiterer Parameter für das Empfehlungssystem.....	37
Abb. 5.8:	Bestätigung der Konfiguration des Empfehlungssystems.....	38
Abb. 5.9:	Abschluss der Konfiguration.....	39
Abb. 5.10:	Feinststeuerung auf Artekelebene.....	40
Abb. 5.11:	Zusammenstellung der Datenbankaufträge (Jobs).....	41
Abb. 5.12:	Personalisierungsmodul in der myfactory Modulliste.....	46
Abb. 5.13:	Produktflyer zum Empfehlungssystem zu myfactory.....	47
Abb. 6.1:	Das Gesamtprojekt PersECA II.....	48
Tab. 5.1:	Einfache Bewertungsmatrix für User-based Collaborative Filtering.....	16
Tab. 5.2:	Ähnlichkeitsmatrix für User-based Collaborative Filtering.....	18
Tab. 5.3:	Einfache Bewertungsmatrix für Item-based Collaborative Filtering.....	19
Tab. 5.4:	Ähnlichkeitsmatrix für Item-based Collaborative Filtering.....	20
Tab. 5.5:	Speicherung der Ähnlichkeitsmatrix für Item-based Collaborative Filtering.....	20
Tab. 5.6:	Beispiel einer Bewertungsmatrix zur Qualitätssicherung.....	23
Tab. 5.7:	Vorkonfiguration der Parameter.....	30
Tab. 5.8:	System Events.....	42
Tab. 5.9:	Stored Procedures.....	43
Tab. 6.1:	Teilschritte im PersECA-II-Projekt.....	49
Tab. 6.2:	Detailablauf des Teilprojekts mit bbv.....	49

Abkürzungsverzeichnis

bbv	Bachofen, Baumann, Vollmann
B2C	Business to Consumer
B2B	Business to Business
BCCS	British Classic Car Spares
KMU	kleinere und mittlere Unternehmen

1 Einleitung

Personalisierung basiert auf dem Aufbau und der Nutzung von Profilen (insbesondere 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.

Die bisherige Arbeit mit Wirtschaftspartnern im Rahmen der Projekte zur Personalisierung von E-Commerce-Applikationen (PersECA) hat gezeigt, dass sowohl der Aufbau als auch die Nutzung von Kundenprofilen in der Praxis meist nicht systematisch erfolgen. Im jetzt abgeschlossenen Projekt PersECA II haben wir uns deshalb auf die folgenden zwei Themenschwerpunkte konzentriert:

- Profile und
- Technologien

Im Bereich „Profile“ sollen Input- und Output-Profile definiert werden. Die Entwicklung geeigneter Methoden soll den gesamten Lebenszyklus der Profilbewirtschaftung (Daten- und Profilpflege) erleichtern. Im Bereich „Technologien“ steht die Nutzung von Technologien, wie Collaborative Filtering, zur Automatisierung von Personalisierungsfunktionen im Vordergrund.

Diese Fallstudie beschreibt das PersECA-II-Teilprojekt mit der bbv Software Services AG, Luzern und Zug. Das Teilprojekt wurde im Frühjahr 2006 gestartet und Anfang 2008 abgeschlossen. Ziel war es, die bestehende Business Software myfactory.BusinessWorld um Personalisierungsfunktionen zu erweitern. In diesem Rahmen wurde ein standardisiertes Empfehlungssystem entwickelt, das im myfactory-Shop-Modul zur Verfügung gestellt werden kann. Da die Funktionsweise eines Empfehlungssystems an die Art der gehandelten Produkte angepasst werden muss, wurden verschiedene Parameter entwickelt, mit denen sich das System kundenspezifisch konfigurieren lässt.

Kapitel 2 beschreibt zunächst die Firma bbv Software Services AG, Luzern, mit der das Projekt durchgeführt wurde und die das Empfehlungssystem implementierte. Kapitel 3 beschreibt den Leistungsumfang der bestehenden Applikation vor dem Personalisierungsprojekt. Kapitel 4 beschreibt die Soll-Konzeption mit Personalisierungsfunktionen. In Kapitel 5 werden das Detailkonzept und die Umsetzung in der konkreten ERP-Software dargestellt. Hierzu gehört auch die Voreinstellung von Personalisierungsparametern. Kapitel 6 beschäftigt sich abschliessend mit dem Projektmanagement. Die Ergebnisse des Projektes werden in Kapitel 7 zusammengefasst und es werden Schlussfolgerungen gezogen.

2 Das Unternehmen: bbv Software Services AG

2.1 Hintergrund

bbv Software Services AG, im Folgenden kurz „bbv“ genannt, entwickelt seit 1995 erfolgreiche, qualitativ hochstehende Software für Industrie, Medizinaltechnik, Gesundheitswesen und Energiewirtschaft sowie für Dienstleistungsunternehmen wie Banken oder Versicherungen. An drei Standorten – Luzern, Zug und Bern – erarbeitet ein eingespieltes Team von Software-Ingenieuren, -Entwicklern, -Architekten, -Qualitätsexperten und -Projektleitern sowie Business-Analysten und Testmanagern innovative Lösungen. Dabei kommen modernsten Technologien und Entwicklungsmethoden zum Einsatz. Die individuellen Software-Lösungen zielen darauf ab, Wettbewerbsvorteile zu schaffen und den Anwendern zu einer starken Position im Markt zu verhelfen.

Seit ihrer Gründung wächst die Firma kontinuierlich. Im Geschäftsjahr 2007 wurde ein Umsatz von CHF 11.8 Mio. erwirtschaftet (2006: CHF 8.1 Mio.). bbv ist mit einem Aktienkapital von CHF 150'000 ausgestattet. Die Bilanzsumme belief sich per Ende 2007 auf CHF 7.8 Mio.

Per 31. Dezember 2007 beschäftigt bbv 70 Mitarbeitende. bbv nutzt zudem ein grosses Netzwerk von freischaffenden Spezialisten. Zurzeit werden drei Lehrlinge in Informatik, Fachrichtung Applikationsentwicklung, ausgebildet [bbv 2008].

Je nach individuellem Bedarf werden bbv-Spezialisten zu projektbezogenen Teams zusammengezogen. Die Spezialisten werden dabei auf Grund ihres Know-hows ausgewählt. Die Firma setzt auf Selbstverantwortung und laufende Weiterentwicklung der Mitarbeitenden und ist durch eine starke Teamkultur geprägt.

2.2 Branche, Produkte, Zielgruppen und Märkte

Die Softwareentwicklungsbranche verzeichnet in der Schweiz momentan geradezu einen Boom. Die Nachfrage nach individuellen Softwareprodukten und den zugehörigen Dienstleistungen ist trotz des umfassenden Angebots an Standardsoftware ungebrochen. Gleichzeitig zeichnet sich das Angebot an IT-Fachkräften eher durch Knappheit aus, so dass die Branche hinsichtlich der personellen Ressourcen stark unter Druck steht.

Im Markt hat bbv insbesondere mit zwei Gruppen von Mitbewerbern zu tun. Einerseits sind dies die Softwareentwicklungsabteilungen von (potenziellen) Kunden, die in externen Beratern und Dienstleistern oft eine Konkurrenz für die eigene Abteilung oder gar Person sehen. Andererseits gibt es die klassischen Mitbewerber, welche dieselben Leistungen wie bbv anbieten. Erschwerend kommt noch der Trend zum Offshoring in Billiglohnländer hinzu.

Die Dienstleistungen von bbv untergliedern sich im Wesentlichen in die Bereiche

- Softwareentwicklung,
- Softwarequalität und
- Softwaremanagement.

Diese drei Bereiche werden in den folgenden Kapiteln erläutert.

2.2.1 Softwareentwicklung

bbv entwickelt qualitativ hochstehende Software für unterschiedliche Anwendungsfelder wie z.B. Medizintechnik, Gesundheitswesen und Energie. Zum Einsatz kommen insbesondere die Programmiersprachen C++ und Java sowie die Entwicklungsumgebung Microsoft .NET. An Datenbanken werden MS SQL Server, MS Access, Oracle und MySQL eingesetzt.

Der Bereich Softwareentwicklung ist weiter unterteilt in die Teilbereiche *System Services*, *Embedded Software* und *Mobile Services*. Die Kernkompetenz der *System Services* sind lokale, verteilte und internetbasierte Industrie- und Businessapplikationen in heterogenem Umfeld, basierend auf Microsoft-Technologien. Dies umfasst sowohl Desktop- und Client/Server-Applikationen als auch Wide-Area-Lösungen, die auf dem Internet basieren und mobile Geräte als Clients einbinden. Diese mobilen Applikationen müssen mit Standardapplikationen wie MS Exchange Server, LDAP und SAP kommunizieren können.

bbv verfügt über viel Erfahrung mit den zukunftsrelevanten Kerntechnologien. Die Softwareentwicklung wird mit dem objektorientierten Vorgehensmodell *Rational Unified Process* (RUP) abgewickelt. Dabei kommt in allen Projekten die *Unified Modelling Language* (UML) als Notationssprache zum Einsatz. Bei der Anforderungsanalyse wird mit *Use Cases* gearbeitet. Darüber hinaus hat die Firma konkrete Erfahrungen mit *Extreme Programming* gesammelt.

bbv unterstützt alle Microsoft-Plattformen von Windows 95 bis Windows Vista inklusive Windows CE. Als Microsoft Gold Certified Partner hat bbv umfangreiche Erfahrungen mit verschiedenen von Microsoft unterstützten Technologien wie .NET, COM, DCOM, COM+, ATL, XML, WTL, MAPI, TAPI, Active Directory usw.

Die System Services beschäftigen sich vorwiegend mit der Entwicklung von kundenspezifischen Softwarelösungen im Umfeld von Microsoft. Einen speziellen Schwerpunkt bildet dabei die Entwicklung mit Microsofts .NET-Technologie. Einige Projekte aus diesem Umfeld umfassen:

- Abrechnung von Medizinalleistungen nach TARMED,
- Modulares Unterrichtskonzept für Fachhochschulen,
- Produktionsstrasse für Brillengläser.

Zusätzlich bietet bbv Beratungs- und Entwicklungsdienstleistungen im Java-Umfeld an. Spezialitäten sind hierbei:

- Java EE (EE = Enterprise Edition) basierte Client/Server-Architekturen,
- Auf Standard Widget Toolkit (SWT) basierte und mit Eclipse entwickelte Rich Clients,
- Swing basierte Rich Clients,
- Java Server Pages (JSP), Struts (ein Open-Source-Framework für die Präsentations- und Steuerungsschicht von Java-Webanwendungen) und Java Server Faces (JSF) basierte Thin Clients.

Im Teilbereich *Embedded Software* wird kundenspezifische Software für Maschinen, Geräte und Steuerungen entwickelt. Hardware-Reviews, Simulation und Implementation von Entwicklungsprozessen gehören ebenso zum Angebot wie Design, Usability-Testing und Integration.

Die *Mobile Services* von bbv verbinden die Embedded Systems mit Business-Anwendungen. Auf PDAs und Smartphones werden Werkzeuge für Techniker und Aussendienstmitarbeitende realisiert, die sich nahtlos in die Unternehmenssoftware einfügen. Auch Echtzeitsystemlösungen gehören zum Repertoire von bbv.

2.2.2 Softwarequalität

Dieser Bereich umfasst die folgenden Dienstleistungen:

- Qualitätsmanagement,
- Testing Services,
- Testautomation und
- Unit Testing.

Zu den Hauptaufgaben des Qualitätsmanagements gehören sowohl der Aufbau eines Testcenters als auch die Definition der Entwicklungs- und Testprozesse. Es wird sichergestellt, dass die begleitenden Prozesse wie Release-, Versions- oder Changemanagement in die Organisation des Kunden passen.

Integrations- und Systemtests stellen sicher, dass die Softwareprodukte die beabsichtigte Funktionalität korrekt zur Verfügung stellen. Zu den Testing Services gehören Teststrategie, Testplan, manuelle oder automatisierte Testausführung.

2.2.3 Softwaremanagement

Das Softwaremanagement umfasst die folgenden Dienstleistungen:

- Coaching und Education: Management- und Wirtschaftlichkeitsberatung für IT Projekte, Architekturberatung, Vermittlung von Technologie- und Prozess-Know-how,
- Übernahme des Projekt Managements,
- Technologieberatung,
- Innovations- und Investitionsberatung.

2.3 Partner

2.3.1 Allgemeine Partner

bbv arbeitet mit einer Reihe von langfristigen Partnern zusammen. Dies ist notwendig, um die steigende Nachfrage der Kunden nach integrierten Leistungen und nach Leistungen aus einer Hand zu befriedigen.

Im Bereich der Client-/Server-Datenbanksysteme wird mit der Firma INTEGRATE Informatik GmbH aus Kriens gearbeitet. Net Solution GmbH aus Cham ist der Partner für Microsoft-Windows-Lösungen. Die MM Informatik AG aus Hünenberg ist Partner für Projektverwaltungs- und CRM- Systeme. Im Designfragen wird mit der Firma quadesign partner ag aus Zug zusammengearbeitet.

Die Abteilung Informatik der Hochschule Rapperswil gründete zusammen mit den Partnerunternehmen Microsoft, Siemens Building Technologies, bbv und Concurrent Engineering das erste .NET-Kompetenzzentrum an einer Hochschule.

Prozessberatung und Training in Usability und Requirements Engineering liefert die Firma Hauri Ergonomie & Coaching. Hierzu gehören das Coaching von Führungskräften, Teamentwicklung, Moderation von Gruppen und Begleitung sowie Steuerung von Organisationsentwicklungen.

Primelco Visual Data AG aus Baar ist ein Handels- und Produktionsunternehmen in den Markt Bereichen Visualisierungssysteme und Industriekomponenten. Die Firma verwirklicht aus Handelsprodukten und Einzelkomponenten gefertigte Prototypen.

2.3.2 Development Partner von myfactory

bbv ist Development Partner von myfactory [myfactory 2008]. Die Standard ERP-Lösung myfactory.BusinessWorld wird an die individuellen Anforderungen der Kunden angepasst und erweitert. Diese Erweiterungen fließen wieder in das myfactory-Produkt ein. Das in dieser Fallstudie beschriebene Projekt führte zur Erweiterung von myfactory.BusinessWorld. Ein entsprechendes Software-Modul steht nach Abschluss der Entwicklung den Partnern und Kunden von myfactory zur Verfügung.

2.4 Unternehmensvision

bbv sieht sich als kompetenter Anbieter für komplexe, softwarelastige Projekte, wobei der gesamte Projektzyklus abgedeckt wird. Diese Position will sie in den kommenden Jahren auf den gesamten deutschsprachigen Raum – auch im Ausland – ausweiten.

Ziel von bbv ist es nicht primär, ERP-Lizenzen zu verkaufen oder Support zu leisten. Das Hauptgeschäftsfeld besteht darin, Standardlösungen zu Individuallösungen weiterzuentwickeln.

2.5 Motivation für die Teilnahme am Personalisierungsprojekt PersECA II

bbv entwickelt massgeschneiderte Lösungen für Kunden aus Industrie und Dienstleistung. Die Personalisierung gewinnt dabei stetig an Bedeutung. Die Motivation für das Projekt bestand darin, Personalisierungsfunktionen zu entwickeln, die in den Standard der myfactory.BusinessWorld-Software aufgenommen werden können.

Damit Personalisierungsfunktionen, wie z.B. die auf einem Empfehlungssystem (Recommender System) basierende Anzeige von personalisierten Empfehlungen, für den Kunden und den Betreiber sinnvolle Ergebnisse liefern, braucht es konkretes Know-how. Für bbv war es sehr wichtig, dass durch das KTI-Projekt konkrete Algorithmen erarbeitet werden, die nicht nur sinnvolle Empfehlungen erzeugen, sondern auch programmiert werden können.

Im Projekt sollten die folgenden Ziele erreicht bzw. unterstützt werden:

- Identifizierung der relevantesten Personalisierungsfunktionen:
 - Mit welchen Personalisierungsfunktionen lässt sich die Wettbewerbsfähigkeit von bbv am Markt steigern?
 - Welche Personalisierungsfunktionen eignen sich für die Aufnahme in ein Standardmodul?
- Konkrete, lauffähige Algorithmen
 - Umsetzung von Algorithmen (durch bbv)
 - Erweiterung der myfactory-Software um die ausgewählten Funktionen.

- Positionierung am Markt
 - Erhöhung der Wettbewerbsfähigkeit bei Neuprojekten (generelle Positionierung von bbv als Spezialist für personalisierte Lösungen)
 - Erwerb von spezialisiertem Know-how im Bereich Personalisierung, um sich von den Mitbewerbern abzuheben (Alleinstellungsmerkmal)

3 IST-Situation: ERP-II-Software vor dem Personalisierungsprojekt

myfactory.BusinessWorld ("Das vernetzte Echtzeit-System für den Mittelstand") ist laut Anbieter eine universell einsetzbare, standortübergreifende und mehrsprachige Softwarelösung, die speziell für kleine und mittelgroße Unternehmen entwickelt wurde. Sie soll es den Firmen ermöglichen, komplexe Geschäftsprozesse zu vernetzen, sie in Echtzeit zu verarbeiten und alle Abläufe über die eigenen Unternehmensgrenzen hinaus abzubilden - z.B. auch solche Vorgänge, an denen externe Zulieferer, ausländische Kooperationspartner, weltweite Niederlassungen oder wechselnde Projektstandorte gleichzeitig beteiligt sind [Computerwoche 2006]. Die Software wird über Partner vertrieben und installiert.

3.1 Leistungsumfang der Software

myfactory.BusinessWorld besteht aus einem Framework von miteinander vernetzten Modulen. Dazu gehören unter anderem:

- ERP (Enterprise Resource Planning)
- PPS (Produktionsplanung- und steuerung)
- B2B (Business to Business)
- CRM (Customer Relationship Management)
- MIS (Management Information System)
- HRM (Human Resource Management)

myfactory.BusinessWorld basiert auf der Software-Architektur .NET von Microsoft. Dadurch lassen sich die Geschäftsprozesse, die Datenstrukturen und die Oberfläche des Systems an die individuellen Anforderungen der Anwenderunternehmen anpassen. Es ist möglich, die Datenstrukturen zu erweitern und z.B. zusätzliche Attribute zu bestehenden Datenbanktabellen hinzuzufügen. Bei einem neuen Release der Software können solche Ergänzungen dann an die Anwender weitergegeben werden. Bildschirmmasken können erweitert und um weitere Masken ergänzt werden. Aus den Masken heraus lassen sich neue Geschäftsprozesse aufrufen. Im Rahmen des Customizings können weitere Ergänzungen vorgenommen werden, wobei die Releasefähigkeit bestehen bleibt. Zugriffe werden über Benutzerrechte gesteuert.

myfactory.BusinessWorld wird als Webapplikation benutzt (Zugriff via Webbrowser) und bietet daher die Möglichkeit, unabhängig von Ort und Zeit auf das System zuzugreifen. Der Zugriff ist durch verschiedenste Zugangsgeräte und in verschiedenen Sprachen möglich.

myfactory.BusinessWorld wird in Unternehmen unterschiedlicher Größe (ein bis 500 Mitarbeitende) eingesetzt. Es gibt verschiedene Branchenlösungen (z.B. für die Textilbranche, Vereine, Entsorgung & Recycling, Gemeinden), die durch Partner realisiert werden. Der Kern der Software (Kernel) bleibt dabei identisch. Es werden nur Zusatzmodule hinzugefügt.

3.2 Elektronischer Produktkatalog

Die Artikelstammdaten werden in myfactory.BusinessWorld genau einer Artikelgruppe zugeordnet und 0:n (keiner bis beliebig vielen) Kataloggruppen. Diese Zuordnung kann dynamisch gepflegt werden. Die Darstellung der Artikeldaten im E-Shop ist damit unabhängig von der Abbildung der Daten im ERP-System; die Gruppierung der Artikel kann so im ERP-Modul (Ar-

tikelgruppe) anders sein als im Shop-Katalog (Kataloggruppe). Die Zuordnung zur Kataloggruppe wird im ERP-System unter „Webshopdetails“ erfasst. Im ERP-System werden auch die Preise erfasst, wobei je nach Kundengruppe (z.B. B2B- oder B2C-Kunden) unterschiedliche Preise hinterlegt werden können. Die folgenden Abbildungen (Abb. 3.1, Abb. 3.2) zeigen Beispiele eines Artikelkatalogs.

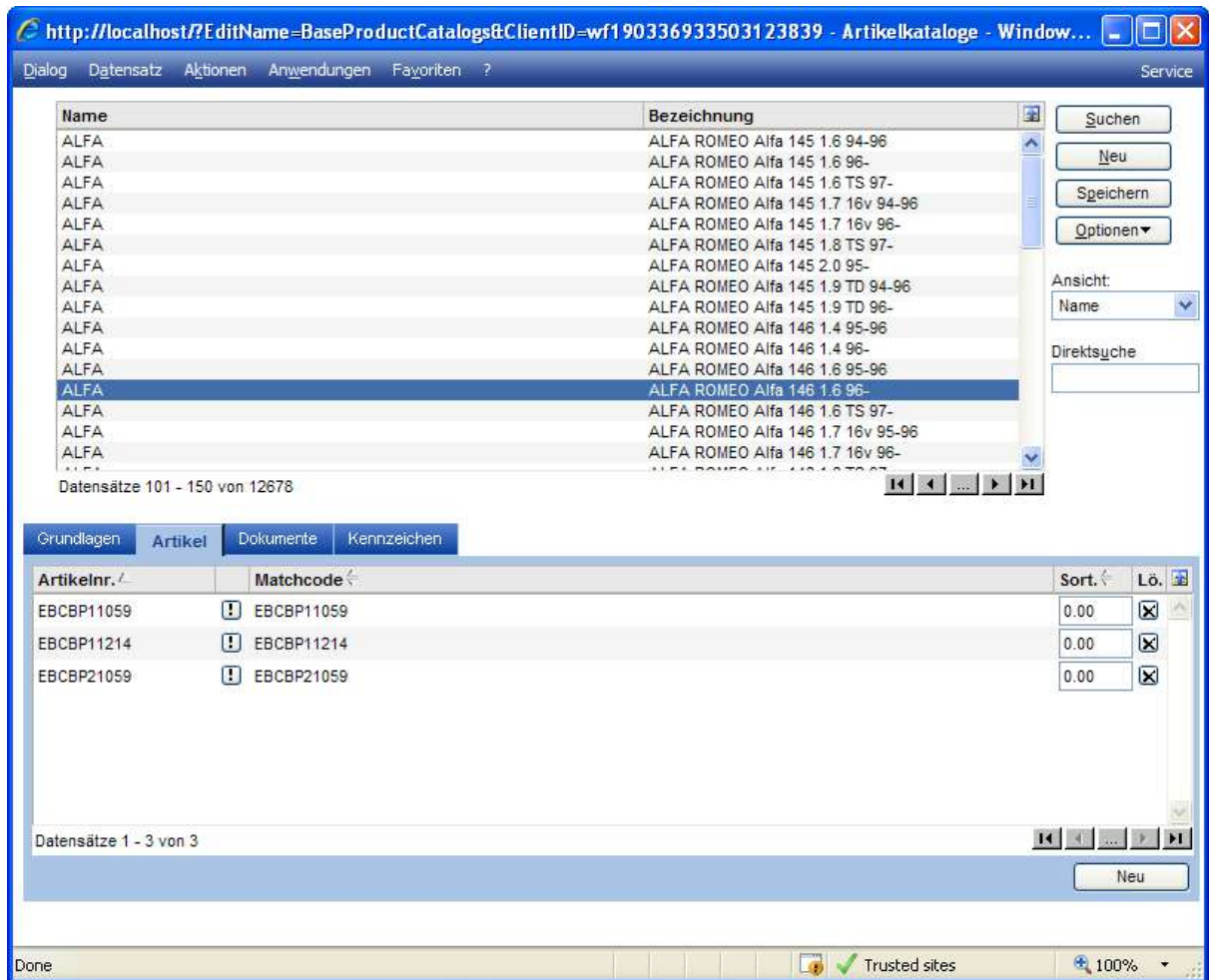


Abb. 3.1: Beispiel eines Artikelkatalogs in myfactory

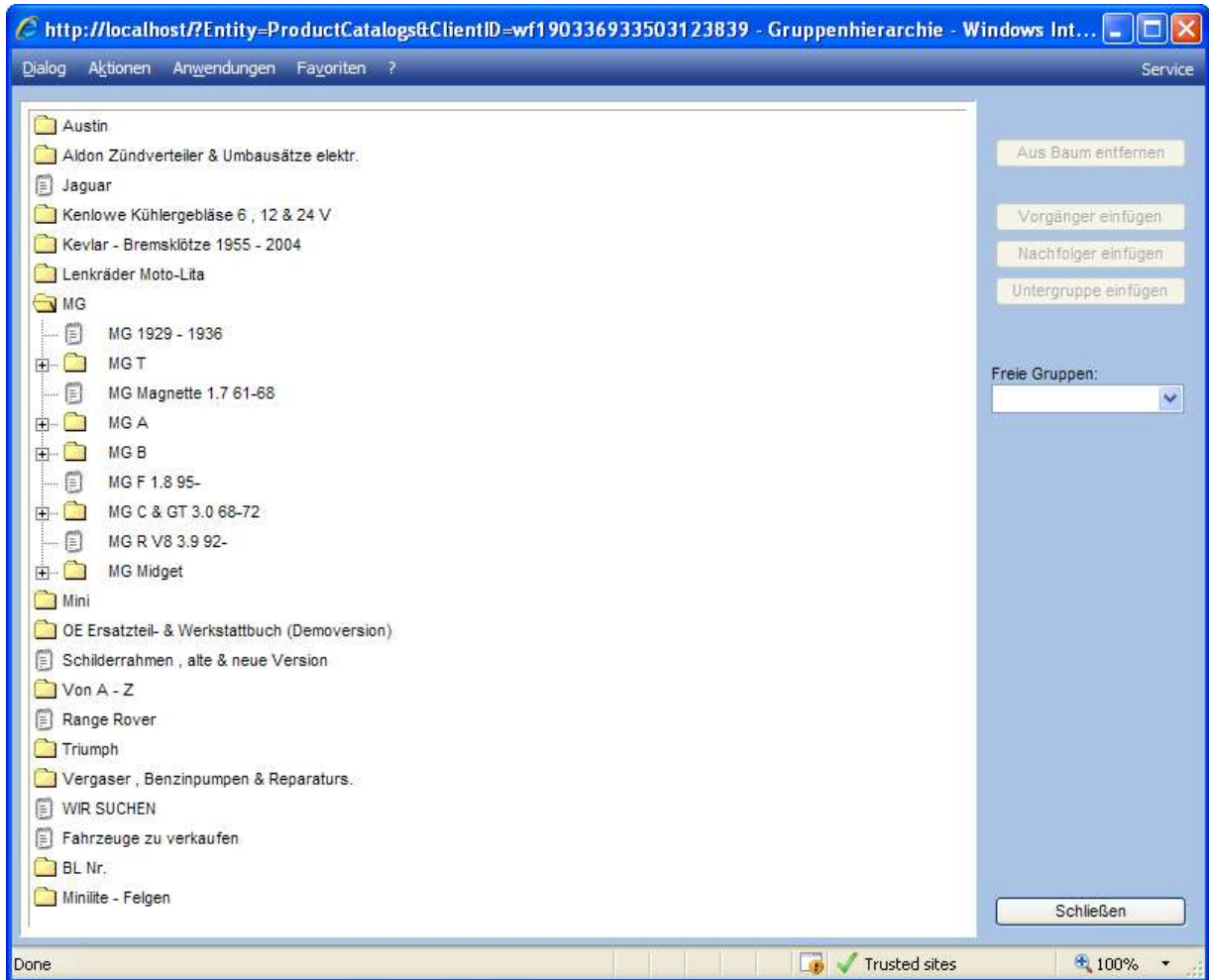


Abb. 3.2: Beispiel einer Gruppenhierarchie im Artikelkatalog

Beispielhaft werden die Daten eines myfactory-Anwenders im Projekt verwendet. Es handelt sich um die Firma BCCS (British Classic Car Spares) [BCCS 2008], die Ersatzteile und Fanartikel für klassische britische Automobile vertreibt. Die Daten umfassen über 4'000 Kunden und über 90'000 Produkte. Der Shop ist seit vier Jahren im Einsatz. Im Oktober 2006 lagen circa 17'000 Verkäufe vor, die von 1'000 Kunden getätigt wurden und sich auf 6'000 Produkte beziehen.

Die Verkäufe werden in myfactory.BusinessWorld in sogenannten Belegen erfasst:

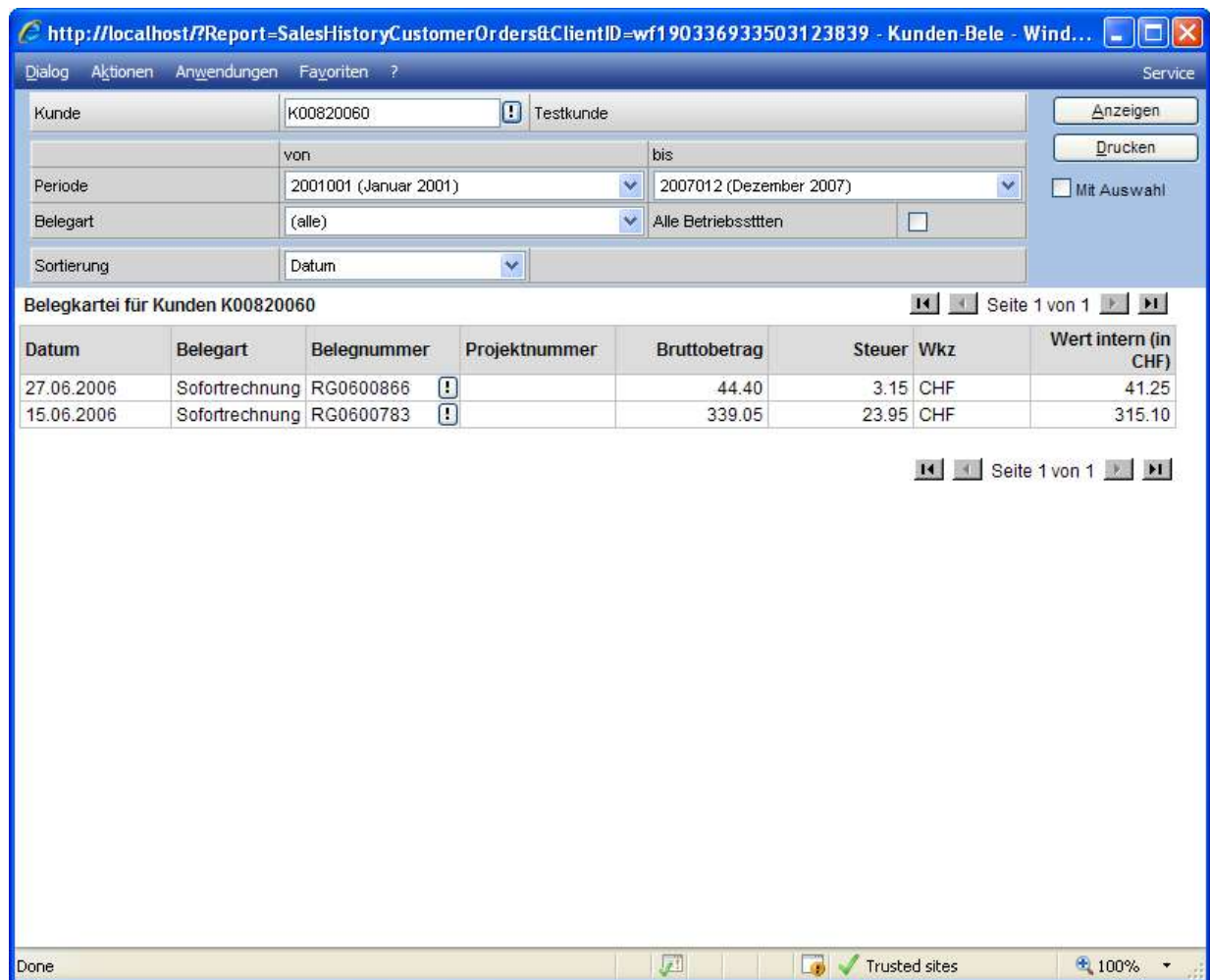


Abb. 3.3: Beispiel der Erfassung eines Verkaufs

3.3 Technologie

3.3.1 Software-Lösung/Programmierung

myfactory.BusinessWorld ist aus dem Bedürfnis entstanden, ein bestehendes, komplexes, technisch rückständiges und kaum noch wartbares ERP-System komplett zu überholen. Das neue Produkt wurde mit neuesten Technologien von Grund auf neu erstellt, um den veränderten Kundenanforderungen und Ablaufprozessen optimal zu genügen.

3.3.2 IT-Architektur

myfactory.BusinessWorld ist eine vollumfänglich webbasierte Lösung. Die gesamte Applikation ist auf einem zentralen Webserver installiert. Eine Installation an den Arbeitsplätzen der

Benutzer entfällt. Aus Sicherheitsgründen ist der Datenbankserver üblicherweise physisch vom Webserver getrennt und durch eine Firewall abgesichert.

Die technische Plattform von myfactory.BusinessWorld basiert auf den folgenden Produkten:

- Microsoft .NET 2.0
- Microsoft Internet Information Services (IIS) Webserver
- Microsoft SQL Server 2000/2005

4 Soll-Konzept: ERP-II-Software mit Personalisierungsfunktionalität

Dieses Kapitel beschreibt nicht nur das fertige Soll-Konzept, sondern führt auch die Schritte auf, die bei der Entwicklung des Soll-Konzepts durchlaufen wurden. Dazu mussten zunächst die relevanten Personalisierungsfunktionen identifiziert und ausgewählt werden (4.1). Anschliessend wurden sie einer Nutzen- (4.2) und Aufwandsbewertung (4.3) unterzogen, um die Wirtschaftlichkeit der Funktionen beurteilen zu können.

4.1 Auswahl der Personalisierungsfunktionen

4.1.1 Diskussion der Personalisierungslandkarte

Die Identifikation geeigneter Personalisierungsfunktionen für die myfactory-Shoplösung startete mit einer Analyse der so genannten Personalisierungslandkarte. Die Personalisierungslandkarte ist eine Mindmap, die wesentliche Personalisierungsfunktionen strukturiert darstellt. Sie wurde im Rahmen des Projekts PersECA II entwickelt und basiert auf langjährigen Erfahrungen, die Mitarbeitende der FHNW in mehreren Personalisierungsprojekten sammeln konnten [Alioski et al. 2008].

Die in der Personalisierungslandkarte aufgeführten Funktionen setzen meist die Anmeldung des (potenziellen) Kunden und damit dessen eindeutige Identifizierung voraus. Davon kann jedoch speziell im B2C-Bereich, auf den sich - wie unten beschrieben - das Projekt mit bbv konzentriert, nicht ausgegangen werden. Die Benutzer geben sich oft erst in einer möglichst späten Phase des Transaktionsprozesses zu erkennen. Die Lösung in der myfactory.BusinessWorld-Software muss also auch die Situation berücksichtigen, dass der Kunde sich noch nicht angemeldet hat.

4.1.2 Schwerpunkt: Personalisierung während der Informationsphase

Die Informationsphase ist die erste der drei typischen Phasen im E-Commerce: Informations-, Vereinbarungs- und Abwicklungsphase. Während der Informationsphase informiert sich ein Kunde über die Produkte im Katalog. Die Phase endet, wenn der Kunde beschliesst einen gefüllten Warenkorb zu kaufen.

Für das Projekt wurden im Anschluss an die Diskussion der Personalisierungslandkarte die beiden folgenden generellen Bereiche identifiziert und diskutiert:

- Individuelle Produktempfehlungen während der Informationsphase im E-Shop
- Unterstützung des Customer Relationship Management (CRM) und des Marketing, z.B. durch personalisierte Newsletter

Zwar verspräche der CRM/Marketing-Teil grosses Potenzial, jedoch wäre dies mit einem sehr grossen Aufwand für die Beschaffung von Daten verbunden. Entsprechende Daten aus dem Kunden- und Marketingbereich waren für bbv zu diesem Zeitpunkt nicht zu beschaffen, da kein Pilotkunde am Projekt beteiligt war.

Aus diesem Grund wurde beschlossen, sich auf die Entwicklung eines Empfehlungssystems für die Informationsphase im E-Shop zu konzentrieren. Während der Informationsphase sollen

dem Kunden persönliche Empfehlungen aufgezeigt werden. Dies kann in verschiedenen Bereichen des E-Shops genutzt werden, z.B.:

- Einstiegsseite: Anzeige von empfohlenen Produkten für den jeweiligen Kunden
- Produktsuche: Sortierung der gefundenen Produkte nicht allein an Hand der Produktbezeichnung, sondern auf Grund von berechneten Empfehlungen
- Füllen des Warenkorbs: Berücksichtigung des Inhalts des Warenkorbs zur Anzeige entsprechender Produktempfehlungen

4.1.3 Notwendige Erweiterungen der Kundenprofildaten

In der myfactory-Software musste die Aufzeichnung der Kundendaten erweitert werden, damit diese zur Berechnung von Empfehlungen mit herangezogen werden können. So mussten die Kundenstamm-Standardfelder durch weitere Felder ergänzt werden, wie z.B. Clickstream-Aufzeichnung.

Clickstreams¹ werden in der ursprünglichen Version des myfactory-E-Shops nicht aufgezeichnet. Eine Aufzeichnung des Clickstreams ist aber notwendig, um die Personalisierungsfunktionen mit genügend Daten zu beliefern (siehe 5.3.1.1).

4.2 Erwarteter Nutzen

Der erwartete Nutzen der zusätzlichen Personalisierungsfunktionen in der myfactory.BusinessWorld-Software umfasst unter anderem die folgenden Punkte:

- Passendere Angebote für Kunden
- Cross- und Up-Selling
- Umsatzsteigerung
- stärkere Kundenbindung
- höhere Kundenzufriedenheit

4.3 Erwarteter Aufwand und Risiken

Der Aufwand für die Realisierung konnte für bbv dadurch im Rahmen gehalten werden, dass die Umsetzung der erarbeiteten Personalisierungsfunktionen im Rahmen des Projekts in enger Zusammenarbeit zwischen bbv und den Hochschulpartnern (FHNW und Universität Fribourg) erfolgte.

Das Risiko resultierte daraus, dass kein Pilotkunde am Projekt beteiligt war. Die eigentliche Anwendung im laufenden Betrieb bei einem Anwender von myfactory.BusinessWorld wird daher erst nach Projektabschluss erfolgen. Dadurch können auftretende Schwierigkeiten im produktiven Betrieb nicht während der Projektlaufzeit angegangen werden.

¹ Clickstream: Sammlung und Auswertung des Verhaltens von Besuchern auf Websites

5 Detailkonzept und Umsetzung

5.1 Allgemeine Rahmenbedingungen

Für das Projekt wurden Rahmenbedingungen festgelegt, um das Vorhaben einzugrenzen. Die erste Rahmenbedingung bestand darin, dass sich das Projekt auf den B2C-Bereich konzentrieren sollte, da nur Beispieldaten aus dem B2C-Bereich zur Verfügung standen.

Zweitens wurde davon ausgegangen, dass die Shopbetreiber (also diejenigen Firmen, die die myfactory-Software inklusive E-Shop einsetzen) aus dem KMU-Bereich stammen. Für diese Firmen ist es in der Regel zu aufwändig, explizit Daten zu pflegen, die für die Empfehlung von Produkten – je nach angewendetem Verfahren – benötigt werden (insbesondere detaillierte Produkteigenschaften für das unten kurz beschriebene Content-based Filtering). Aufgrund dieser Rahmenbedingung wurden im Projekt Möglichkeiten für die Voreinstellung von allgemeinen Parametern beschrieben und implementiert (siehe 5.4).

Drittens ist davon auszugehen, dass Kunden zunächst anonym auftreten. Das heißt, dass sie im Allgemeinen nicht bereit sind, ihre Identität, Vorlieben, Hobbys etc. preiszugeben, bevor sie überhaupt mit der Suche nach Produkten beginnen. Die Preisgabe der Identität erfolgt im B2C-E-Shop meist erst während des letzten Schrittes des Einkaufsprozesses, wenn es um die Abwicklung der Bezahlung und Lieferung geht. Für ein Empfehlungssystem bedeutet das, dass Empfehlungen auch funktionieren müssen, wenn über die Identität des Kunden noch nichts bekannt ist. Oftmals wird hierbei mit Cookies² gearbeitet, jedoch kann nicht davon ausgegangen werden, dass alle Kunden Cookies zulassen. Die ursprüngliche Version der myfactory-Software verfügte nicht über die Möglichkeit, Cookies aufzuzeichnen. Diese Funktion wurde im Rahmen des Projekts implementiert.

5.2 Auswertungsmethoden und Algorithmen für Empfehlungssysteme

Grundsätzlich lassen sich verschiedene Verfahren unterscheiden. Dies sind im Wesentlichen:

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

Die Verfahren werden im Folgenden kurz beschrieben. Vorweg sei angemerkt, dass im beschriebenen Projekt das Item-based Collaborative Filtering verwendet wurde.

² Ein Cookie ist eine kleine Datei, die von einem Server meist in einem speziellen Dateiverzeichnis auf einem Clientcomputer gespeichert wird. Sie dient dem Informationsaustausch zwischen Client und Server. Hierdurch kann ein Benutzer bei einem abermaligen Besuch des Shops mit hoher Wahrscheinlichkeit identifiziert werden, ohne dass er sich aktiv anmeldet.

5.2.1 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 unstrukturierter Produktbeschreibungen extrahieren. Mathematische Verfahren wie die Inverse Document Frequency (IDF³) 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.

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.

Aus den genannten Gründen scheidet dieses Verfahren für die Umsetzung im Projekt aus. Damit ist jedoch nicht ausgeschlossen, dass eine entsprechende Funktionalität später hinzugefügt werden kann.

5.2.2 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

³ Die "Inverse Document Frequency" ist ein Mass für die Bedeutung eines Begriffs. Diese Methode kommt beim "Information Retrieval" und "Text Mining" zum Einsatz.

- 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.

5.2.2.1 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. 5.1).

Tab. 5.1: Einfache Bewertungsmatrix für User-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	0.5	\emptyset	...
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:

$$\begin{aligned} (1*1) + (0*1) + (0*0) + (1*0) + (-1*-0.3) \\ = 1 + 0.3 = 1.3 \end{aligned}$$

Der *Nenner* errechnet sich wie folgt:

Die erste Summe des Nenners lautet:

$$1 + 0 + 0 + 1 + 1 = 3$$

Die Wurzel daraus ergibt den Wert 1.73

Die zweite Summe des Nenners ergibt:

$$1 + 1 + 0 + 0 + 0.09 = 2.09$$

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. 5.2: 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 (bzw. 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").
- Aus den genannten Gründen wurde im Projekt gegen das User-based Collaborative Filtering entschieden. Stattdessen wurde das im Folgenden beschriebene Verfahren angewandt.

5.2.2.2 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; Leimstoll et al. 2008]. 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 so genannten Offline-Prozess [Leimstoll et al. 2008].

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 Formel⁴:

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

Dabei bedeuten:

- $\text{sim}(i, j)$: Ähnlichkeit zwischen den beiden Produkten i und j
- $\text{Freq}(ij)$: Häufigkeit, mit der die Produkte i und j gemeinsam gekauft wurden
- $\text{Freq}(i)$: Häufigkeit, mit der Produkt i gekauft wurde
- α : Gewichtungparameter – 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. 5.3: 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	
...				

⁴ 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.

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

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

Die folgende Tabelle zeigt die Ähnlichkeitsmatrix der Produkte (Tab. 5.4):

Tab. 5.4: Ähnlichkeitsmatrix für Item-based Collaborative Filtering

	Produkt 1	Produkt 2	Produkt 3	Produkt 4	Produkt 5	...
Produkt 1		0.5	∅	0.35	0.58	
Produkt 2			∅	∅	0.58	
Produkt 3				0.7	0.58	
Produkt 4					0.58	
Produkt 5						
...						

Diese Matrix kann wie folgt gespeichert werden (Tab. 5.5):

Tab. 5.5: Speicherung der Ähnlichkeitsmatrix für Item-based Collaborative Filtering

P1#	P2#	Similarity
Produkt 1	Produkt 2	0.5
Produkt 1	Produkt 3	∅
Produkt 1	Produkt 4	0.35
Produkt 1	Produkt 5	0.58
Produkt 2	Produkt 1	0.5
Produkt 2	Produkt 3	∅
Produkt 2	Produkt 4	∅
Produkt 2	Produkt 5	0.58
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
...

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.

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.

Die individuelle Empfehlung für einen Kunden erfolgt dann dynamisch im Moment des Besuchs des E-Shops. Mit folgendem SQL⁵-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 5.4 Parametrisierung).

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 laufzeitintensiv 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).

Im Projekt mit bbv wurde das Item-based Collaborative Filtering umgesetzt. Erste Tests innerhalb des BCCS-Shops (siehe 3.2) zeigten, dass hierdurch sinnvolle Empfehlungen generiert werden.

⁵ SQL (Structured Query Language) ist eine standardisierte Datenbanksprache zur Definition, Abfrage und Manipulation von Daten in relationalen Datenbanken.

5.2.3 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.

Für das Projekt mit bbv kamen diese Ansätze nicht in Frage, da - wie oben erwähnt - nicht davon ausgegangen werden kann, dass die Shopbetreiber bereit sind, die notwendigen Informationen über die Produkteigenschaften für das Content-based Filtering zu erfassen, geschweige denn zu pflegen.

5.2.4 Aktuelle Forschungsthemen

Einige Forschungsfelder im Bereich der Empfehlungssysteme sind:

- Einbezug von Kontextinformationen bei der Berechnung der Empfehlungen (z.B. des aktuellen Standorts oder der Tageszeit), um deren Qualität zu erhöhen [vgl. z.B. Kwon/Keun Shin 2008 und Krüger et al. 2007]
- Messung der Güte einer Bewertungsmatrix: Wie viele Bewertungen benötigt eine Bewertungsmatrix, um gute Empfehlungen zu liefern [vgl. z.B. Lee et al. 2007]?
- Optimierung der Performance (Laufzeit- und Speicherverhalten des Algorithmus) insbesondere unter Miteinbezug aktuellster Daten [vgl. 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 [vgl. Castagnos/Boyer 2006 und Kobsa 2007]

Mit diesen Themen hat sich das Projekt nicht weiter beschäftigt, da die Umsetzung einer Lösung und nicht die Klärung dieser grundlegenden Probleme im Vordergrund stand.

5.2.5 Qualitätssicherung

Wie kann die Qualität der Empfehlungen, die mit einem Empfehlungssystem errechnet werden, getestet werden? Hierzu existieren im Wesentlichen zwei Verfahren:

Testverfahren 1

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. 5.6: Beispiel einer Bewertungsmatrix zur Qualitätssicherung

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

Hierdurch ändert sich der Bewertungsvektor des Kunden a. Nun wird die Ähnlichkeitsmatrix, wie unter 5.2.2.2 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.

Testverfahren 2

Im laufenden Betrieb wird überprüft, wie die Kunden auf die berechneten Empfehlungen reagieren. Dieses Verfahren ist aufwändiger und nur bei entsprechender Beteiligung sowohl auf Shopbetreiber- als auch auf Kundenseite durchführbar.

5.3 Ablauf der Berechnung von Empfehlungen

5.3.1 Input-Profile

In der MS-SQL-Server-Datenbank von myfactory werden die für die Personalisierung relevanten Daten in separaten Tabellen gespeichert. Diese Tabellen werden im Folgenden beschrieben.

5.3.1.1 Aufzeichnung des Clickstreams

Um den Füllungsgrad der Bewertungsmatrix zu erhöhen, werden die Clickstreams der Kunden aufgezeichnet. Dies war in der ursprünglichen Version von myfactory.BusinessWorld nicht möglich und wurde im Rahmen des Projekts implementiert.

Immer, wenn ein Kunde Details zu Produkten aufruft, wird in der folgenden Tabelle ein Datensatz hinzugefügt:

Table Clickstream (SequenceNo, InputDate, SessionID, ProductID)

Es wird jeweils die Session-ID, die Artikelnummer (ProductID) sowie ein Zeitstempel (InputDate) aufgezeichnet. Um den Kunden zu identifizieren, erfolgt in einer separaten Tabelle die Zuordnung von SessionID und CustomerID. Wie das funktioniert, wird unter 5.5.2 (Einsatz von

Cookies) beschrieben. Auf diese Weise kann die Identität des Kunden ermittelt werden, sobald sich dieser einmal im Shop angemeldet hat.

5.3.1.2 Füllen der Rating-Matrix

Hat ein Kunde ein Produkt gekauft, so wird in der Rating-Matrix die Bewertung 1 eingetragen. Hat er es nur angeklickt (Details angeschaut), so wird eine Bewertung von z.B. 0.3 eingetragen. Dieser Wert kann vom Shopadministrator verändert werden. Die Ratingwerte liegen insgesamt immer zwischen 0 und 1.

Die entsprechende Tabelle sieht wie folgt aus:

```
Table Ratingmatrix (CustomerID, ProductID, Rating)
```

Der Clickstream kommt nur zum Tragen, wenn der Kunde das Produkt noch nicht gekauft hat. Zudem wird die auf dem Clickstream basierende Bewertung für die Berechnung der Ähnlichkeitsmatrix nur dann berücksichtigt, wenn der Kunde das Produkt in den letzten x Tagen y -mal angeschaut hat. x und y lassen sich ebenfalls vom Shopadministrator konfigurieren.

Die Prozedur sieht wie folgt aus:

```
CREATE Procedure dbo.bbvAddImpressionsToRatingMatrix
(
    @MinImpressions int = 3
    ,@LastImpressionWithinDays int = 7
    ,@RatingForImpression float = 0.3)

```

5.3.2 Berechnung der Ähnlichkeitsmatrix

Mit den Werten der Rating-Matrix wird der unter 5.2.2.2 beschriebene Algorithmus von Deshpande/Karypis [2004] ausgeführt. Die Ergebnisse werden in der Tabelle ProductCorrelation (Ähnlichkeitsmatrix) gespeichert. DesKar bezeichnet dabei die Ähnlichkeit zwischen Produkt 1 und Produkt 2:

```
Table ProductCorrelation (Product1, Product2, DesKar)
```

5.3.3 Output-Profile

Die folgende Tabelle enthält alle Empfehlungen für einen spezifischen Kunden. Die Liste ist absteigend sortiert nach der Ähnlichkeit zu den von diesem Kunden bereits gekauften Produkten (bzw. angeklickten, falls der Kunde noch keine Produkte gekauft hat). (DesKarRank beschreibt die Rangordnung der Produkte):

```
Table Recommendations (CustomerID, ProductID, DesKar, DesKarRank)
```

Um das wiederholte Empfehlen eines Produkts zu verhindern, wird in der folgenden Tabelle gespeichert, wann ein Produkt einem Kunden das letzte Mal empfohlen wurde. Es wird dann n Mal unterdrückt. n ist für den Shopadministrator konfigurierbar. Wie lange ein Produkt unterdrückt werden soll, steht im Attribut *hide*. Bei jeder Empfehlung wird *hide* um 1 vermindert. Erreicht es den Wert 0, so wird es aus dieser Tabelle gelöscht und das Produkt kann dem entsprechenden Kunden wieder empfohlen werden. Die Tabelle hat folgende Struktur:

```
Table AlreadyRecommended (CustomerID, ProductID, LastRecommended, Hide)
```

5.4 E-Shop-Klassifizierung und -Parametrisierung für ein Standard-Empfehlungssystem

Da es sich bei dem beschriebenen Personalisierungsmodul um eine Standardsoftware für ein Empfehlungssystem handelt, die nach der Installation beim Anwender an dessen Anforderungen angepasst werden muss, ist es sinnvoll, Voreinstellungen vorzunehmen. Hierdurch erleichtert sich die spätere Arbeit der Administratoren beim Einrichten des E-Shops. Im Folgenden wird beschrieben, wie diese Voreinstellungen aussehen.

5.4.1 Fragestellung

Ein Beispiel für einen Parameter, der bei der Installation eines E-Shops voreingestellt werden soll, ist die Frage, ob die von einem Kunden bereits gekauften Produkte diesem Kunden empfohlen (z.B. Butter) oder nicht empfohlen (z.B. Fernseher) werden sollen. Die Idee besteht nun darin, E-Shops generell zu kategorisieren, um eine Reihe solcher Parameter entsprechend der Shopkategorie vorab bestimmen und einstellen zu können. Auf diese Weise könnte der Shop vorkonfiguriert werden. Der Shopadministrator müsste diese Voreinstellungen anschließend nur noch für einzelne Produkte oder Produktgruppen anpassen.

In einer Reihe von Workshops wurden im Projekt geeignete Kategorien definiert, und die Grundeinstellung der Parameter wurde festgelegt. Das Vorgehen wird im folgenden Abschnitt kurz dargelegt. Die dem Empfehlungssystem zugrunde liegenden Algorithmen (siehe 5.2.2.2) sind von der Parametrisierung nicht betroffen und können unabhängig davon fest implementiert werden.

5.4.2 Klassifizierung von E-Shops

Ein erster Ansatzpunkt für die Kategorisierung von E-Shops ergibt sich aus der *klassischen Einteilung von Gütern in den Wirtschaftswissenschaften*:

B2C-Bereich

- Konsumgüter (primär für den privaten Konsum, Endverbraucher)
 - Verbrauchsgüter des täglichen Bedarfs (engl. „convenience goods“)
 - relativ geringer Preis
 - z.B. Nahrungsmittel, Waschmittel, Medikamente, etc.
 - hochwertige Güter (engl. „Shopping Goods“)
 - hier werden typischerweise vor dem Kauf Vergleiche angestellt
 - z.B. Fahrzeuge, Immobilien, Möbel, TV-Geräte, etc.
 - Luxusgüter (engl. „Specialty Goods“)
 - beschränkte Anzahl alternativer Marken
 - geringere Bedeutung des Preises
 - hochpreisig
 - z.B. Luxusuhren, teures Porzellan, spezielle Automobile etc.

- Dienstleistungen
 - z.B. Reisen, Bankprodukte, Online-Beratung etc.

B2B-Bereich

- MRO-Güter (Maintenance, Repair and Operations): indirekte Bedarfe in der Produktion eines Industriebetriebes
- Investitionsgüter

Ein zweiter Ansatzpunkt wäre die *Kategorisierung nach Branchen*. Risch (2007) beschreibt z.B. in einer nicht-repräsentativen Studie die folgende Verteilung von E-Shops nach Branchen:

- 18 % Kleidung und Mode → saisonale Abhängigkeit
- 13 % IT, Elektronik → aktuell, markengebunden, Selbstdeklaration der Präferenzen
- 12 % Reisen & Tourismus
- 11 % Nahrung, Tabak
- 8 % Infos und News / Medien / Diverses
- 7 % Bürobedarf / Geschenkartikel
- 4 % Bau, Handwerk
- 3 % Automobile

Abb. 5.1 zeigt als dritten Ansatzpunkt die Ergebnisse einer Studie über das Einkaufsverhalten im Internet. Man sieht dabei, dass Produktkategorien und Branchen miteinander vermischt sind. Eine geeignete Kategorisierung scheint auf diesem Wege schwierig realisierbar zu sein.

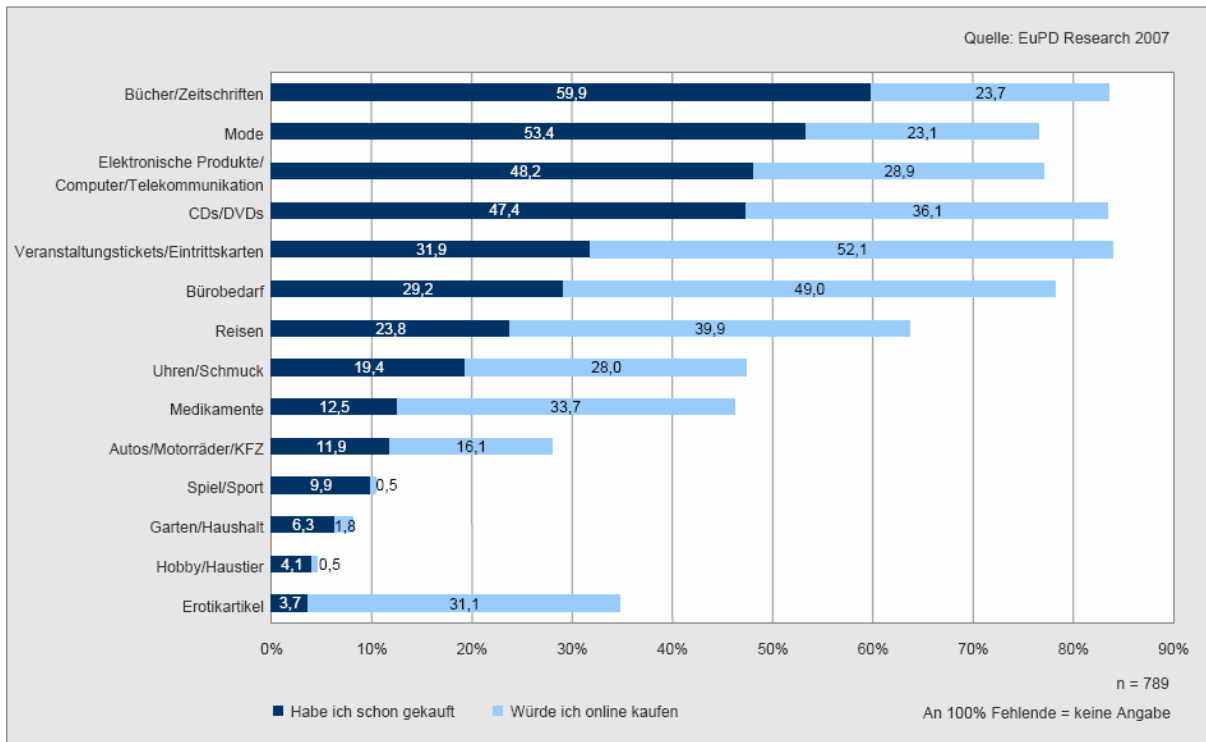


Abb. 5.1: Im Internet gekaufte Warengruppen [PayPal 2007]

5.4.3 Zusammenführung von Shop-Kategorien und Empfehlungs-Parametern

Generell stellt sich die Frage, ob die Parameter für die Empfehlungen anhand von Branchen oder Produkten ermittelt werden sollten. Da eine Kategorisierung nach „Branche“ sehr grob wäre, erwies es sich als sinnvoll, die Parameter für Produktkategorien festzulegen. Mit Hilfe einer Mindmap wurden E-Shop-Kategorien und Parameter einander zugeordnet (Abb. 5.2).

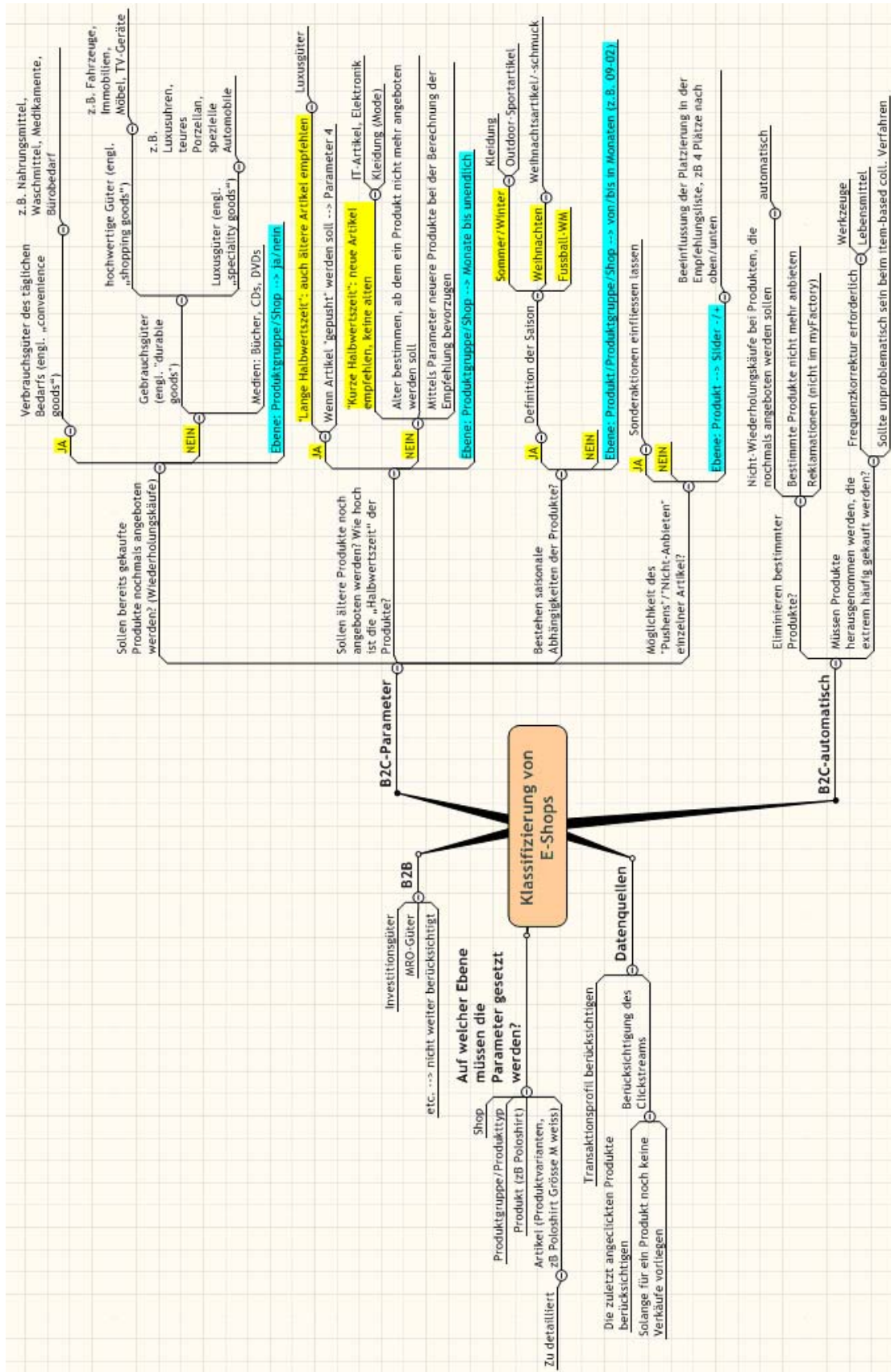


Abb. 5.2: Mindmap zur Zuordnung von E-Shop-Kategorien und Parametern

Als Parameter im B2C-Bereich wurden die folgenden vier wesentlichen Fragen identifiziert:

1. Sollen bereits gekaufte Produkte angeboten werden?
2. Sollen ältere Produkte noch angeboten werden? Wie hoch ist die „Halbwertszeit“ der Produkte?
3. Bestehen saisonale Abhängigkeiten der Produkte?
4. Sollen einzelne Artikel speziell empfohlen oder nicht empfohlen werden? Der Shopbetreiber sollte die Möglichkeit haben, einzelne Produkte zu "pushen" beziehungsweise gezielt aus den Empfehlungen herauszuhalten ("nicht-pushen").

Die Parameter beziehen sich auf das Filtern und nicht auf die Berechnung der Ähnlichkeitsmatrix im Item-based Collaborative Filtering. Das Prinzip ist also, zuerst die Ähnlichkeitsmatrix mittels des Item-based Collaborative Filtering Algorithmus unabhängig von den Parametern zu berechnen (wie unter 5.2.2.2 beschrieben) und anschliessend gemäss der Parametereinstellungen zu filtern.

Jeder Parameter lässt sich entweder für den ganzen Shop, auf Produktgruppenebene, Produktebene oder Artikelebene einstellen. Bei der Vorkonfiguration des Systems werden die Parameter auf Shopebene voreingestellt (Default-Werte). Sie können anschliessend vom Administrator auf den verschiedenen Ebenen überschrieben werden.

Als Datenquellen (links unten in der Mindmap in Abb. 5.2) werden Transaktionen (Käufe) immer berücksichtigt. Clickstreams werden berücksichtigt, wenn für ein Produkt noch keine Käufe vorliegen. Dies wird fest im Algorithmus implementiert und ist nicht Teil der Parametereinstellungen durch den Shopadministrator (siehe 5.3.1.2).

5.4.4 Parametrisierung

Es scheint keine einheitliche Systematik zur Kategorisierung von Produkten zu geben. In der Literatur tauchen zwar dieselben Begriffe immer wieder auf. Je nach Autor sind diese aber unterschiedlich zugeordnet oder um zusätzliche Begriffe ergänzt. Auch der Versuch, die vier zentralen Fragen (5.4.3) Branchen zuzuordnen, scheitert daran, dass es innerhalb der einzelnen Branchen wiederum sehr unterschiedliche Produkte gibt, so dass hier eine einheitliche Voreinstellung der Parameter wenig Sinn macht.

Daher wurde eine neue Kategorisierung erarbeitet, die im B2C-Bereich für Konsumgüter (materielle Güter) wie folgt aussieht:

- Verbrauchsgüter des täglichen Bedarfs:
 - „Convenience Goods“: vorwiegend geringer Preis (z.B. Nahrungsmittel, Waschmittel, Medikamente)
 - seltener hoher Preis (z.B. Premium-Nahrungsmittel, Wein, Kosmetika, Bio)
- Gebrauchsgüter:
 - geringwertig (z.B. Küchenartikel, Wäsche, Deko-Artikel)
 - hochwertig (engl. „Shopping Goods“): typischerweise werden vor dem Kauf Vergleiche angestellt, z.B. Fahrzeuge, Immobilien, Möbel, TV-Geräte
- Luxusgüter (hochpreisige Güter, engl. „Specialty Goods“): beschränkte Anzahl alternativer Marken, geringe Bedeutung des Preises, z.B. Luxusuhren, teures Porzellan, spezielle Automobile

Die folgende Tab. 5.7 zeigt die Vorkonfiguration der Parameter in Abhängigkeit von der Güterart:

Tab. 5.7: Vorkonfiguration der Parameter

Cluster	Parameter	Verbrauchsgüter des täglichen Bedarfs		Gebrauchsgüter		Luxusgüter („specialty goods“)
		geringer Preis („convenience goods“) z.B. Nahrungsmittel, Waschmittel, Medikamente	höherer Preis z.B. Premium-Nahrungsmittel, Wein, Kosmetika, Bio	geringwertig z.B. z.B. Küchenartikel, Wäsche, Deko-Artikel	hochwertig z.B. Fahrzeuge, Immobilien, Möbel, TV-Geräte	
Sollen bereits gekaufte Produkte nochmals angeboten werden? (Wiederholungskäufe)	Shop	√	√	—	—	—
	Produktgruppe/-typ	√	√	—	—	—
	Produkt (zB Poloshirt)	√	√	—	—	—
	Artikel (zB Poloshirt Grösse M weiss)	√	√	—	—	—
Sollen ältere Produkte noch angeboten werden? Wie hoch ist die "Halbwertszeit" der Produkte?	Shop	√	√	√	√	√
	Produktgruppe/-typ	√	√	√	√	√
	Produkt (zB Poloshirt)	√	√	√	√	√
	Artikel (zB Poloshirt Grösse M weiss)	√	√	√	√	√
Bestehen saisonale Abhängigkeiten der Produkte?	Shop	—	—	—	—	—
	Produktgruppe/-typ	—	—	—	—	—
	Produkt (zB Poloshirt Lacoste)	√	√	√	√	√
	Artikel (zB Poloshirt Lacoste Grösse M weiss)	√	√	√	√	√
Möglichkeit des "Pushens"/"Nicht-Anbieten" einzelner Artikel?	Shop	—	—	—	—	—
	Produktgruppe/-typ	—	—	—	—	—
	Produkt (zB Poloshirt)	√	√	√	√	√
	Artikel (zB Poloshirt Grösse M weiss)	√	√	√	√	√

Auf der linken Seite sieht man die vier Parameter, die jeweils für den gesamten Shop oder auf Produktgruppen-, Produkt- oder Artikelebene individuell eingestellt werden können. Oben ist die vorgeschlagene Kategorisierung dargestellt. Ein "√" bedeutet, dass der Parameter aktiviert ist ("ja"), ein "-" bedeutet, dass der Parameter deaktiviert ist ("nein").

Bei "Convenience Goods" sollen z.B. die von einem Kunden bereits gekauften Produkte diesem standardmässig wieder empfohlen werden: Kauft ein Kunde regelmässig ein bestimmtes Waschmittel, so ist es wahrscheinlich, dass er dieses bei einem späteren Einkauf erneut kaufen möchte.

Anders verhält es sich bei hochwertigen Gebrauchsgütern. Hat ein Kunde z.B. einen Fernseher gekauft, ist es unwahrscheinlich, dass er in nächster Zeit wieder einen Fernseher kaufen wird. Diesem Kunden sollte also in den Empfehlungen kein Fernseher vorgeschlagen werden.

5.5 Programmierung und Umsetzung

5.5.1 Umsetzung des Item-based Collaborative Filtering

Wie erwähnt, wurde im Projekt das Item-based Collaborative Filtering mit dem Algorithmus nach Deshpande und Karypis [2004] implementiert. Als Beispielshop diente der mit myfactory-Software aufgebaute und betriebene BCCS-Shop [BCCS 2008]. Der Prozess vom Bilden der Bewertungsmatrix bis zur Berechnung der Ähnlichkeitsmatrix inklusive Ranking dauert auf einem Dell Latitude D820 Laptop rund eineinhalb Stunden (bei circa 1.45 Millionen Produktkorrelationen und 6.6 Millionen Empfehlungen).

Als Alphawert (siehe 5.2.2.2) wurde 0.5 gewählt.

5.5.2 Implementierung von Cookies

Die myfactory-Software musste um Cookies erweitert werden, damit nicht-eingeloggte Kunden identifiziert werden können. Dabei wird die Session-ID von myfactory in das Cookie geschrieben.

Zwei wesentliche Probleme ergeben sich bei der Arbeit mit Cookies:

1. Cookies können vom Client verändert werden. Die Information, die im Cookie gespeichert wird, darf dem Client nicht erlauben, sich eine falsche Identität zu erschleichen.
2. Cookies können vom Client abgelehnt werden. Statistiken, wie viele Benutzer Cookies deaktiviert haben, liegen nicht vor. Man schätzt jedoch, dass es circa 25 % der Benutzer sind.

Zur Lösung dieser Probleme wurde ein Ansatz über die Session-ID gewählt. Das funktioniert folgendermassen:

- Die Session-ID wird vom Webserver für jede Sitzung (engl. Session) automatisch erzeugt. Die Session-ID ist eine sehr grosse Zahl. Ein Brute-Force-Angriff⁶ auf die Session-ID würde zu lange dauern, um erfolgreich zu sein.
- Die Session-ID kann vom Server im Cookie gespeichert und an den Client gesendet werden.
- Der Server muss sich die Session-ID ebenfalls merken, um bei einer späteren Sitzung den Benutzer wieder zu identifizieren. Hierdurch kann das Verhalten des Kunden auch über mehrere Sitzungen hinweg verfolgt werden (Clickstream-Aufzeichnung).

Meldet sich ein bisher unbekannter Kunde im E-Shop an (Registrierung und Authentifizierung), so ermittelt der Server eine spezifische User-ID für diesen Kunden sowie eine Session-ID. Anschliessend wird ein Cookie mit der Session-ID an den Client gesendet und dort auf der Festplatte gespeichert. Gleichzeitig speichert der Server die Session-ID und die User-ID in einer Datenbanktabelle auf dem Server.

Kommt der Kunde vom gleichen Clientcomputer aus zu einem späteren Zeitpunkt auf den E-Shop zurück, so wird dabei das auf dem Clientcomputer gespeicherte Cookie an den Server

⁶ Ein Brute-Force-Angriff probiert alle Möglichkeiten aus, um eine Lösung zu finden.

gesandt. Über die darin enthaltene Session-ID kann der Server nun die zugehörige User-ID ermitteln (durch Nachschlagen in der entsprechenden Datenbanktabelle). Dadurch kann der Server auf die Identität des Kunden schliessen, ohne dass sich dieser explizit am System angemeldet hat.

Der Ansatz löst das erste oben genannte Problem, da sich der Kunde aufgrund der Session-ID keine falsche Identität erschleichen kann (durch explizite Manipulation des Cookies). Bestehen bleibt hingegen das Problem, dass der Kunde auf seinem Rechner Cookies ablehnen kann. In diesem Fall kann der Kunde vor der eigentlichen Anmeldung im Shop nicht erkannt werden.

5.5.3 Implementierung der Clickstream-Aufzeichnung

Die Clickstream-Aufzeichnung wurde für den myfactory-Standard implementiert. Im realen E-Shop von BCCS erfolgte keine Installation der Clickstream-Aufzeichnung, weil die myfactory-Version des BCCS-Shops bereits drei Jahre alt ist und daher zuerst ein Upgrade auf die neue Programm-Version durchgeführt werden müsste. Dies wird vom Betreiber derzeit nicht gewünscht.

5.5.4 Empfehlungen in der myfactory.BusinessWorld-Software

Das Empfehlungssystem wurde wie beschrieben in einem Zusatzmodul zur myfactory.BusinessWorld-Software implementiert (siehe Abb. 5.3). Mithilfe so genannter "Design-Tags" (Markierungen, die das Aussehen und das Verhalten des Shops beeinflussen) kann der Shopadministrator den Shop an eigene Bedürfnisse anpassen (z.B. Positionierung der Empfehlungen, Anzahl der angezeigten Empfehlungen). Dieses Prinzip des Customizing wurde auch für das Empfehlungssystem realisiert. So bestehen z.B. folgende Anpassungsmöglichkeiten:

- Es kann eingestellt werden, wie viele Produkte empfohlen werden sollen.
- Ein einmal empfohlenes Produkt kann bei den nächsten Empfehlungen unterdrückt werden. Der Administrator kann einstellen, wie häufig diese Unterdrückung stattfindet.
- Die Empfehlungen können wahlweise mit einem Produktbild und dem aktuellen Verkaufspreis ergänzt werden.
- Die Position der Empfehlungen auf der Startseite lässt sich anpassen. Die folgenden Abbildungen veranschaulichen die Implementierung.

Die personalisierten Empfehlungen für einen Kunden i basieren entweder auf den in der Vergangenheit getätigten Käufen des Kunden i oder - falls keine Käufe vorhanden sind - auf den Produkten im aktuellen Warenkorb. Die Lösung arbeitet mit Cookies, um die Kunden zu identifizieren, die sich noch nicht angemeldet haben. Der Benutzer muss allerdings mit dem aktuellen Clientcomputer mindestens einmal angemeldet gewesen sein, da ansonsten kein Cookie vorhanden ist (siehe 5.5.2).

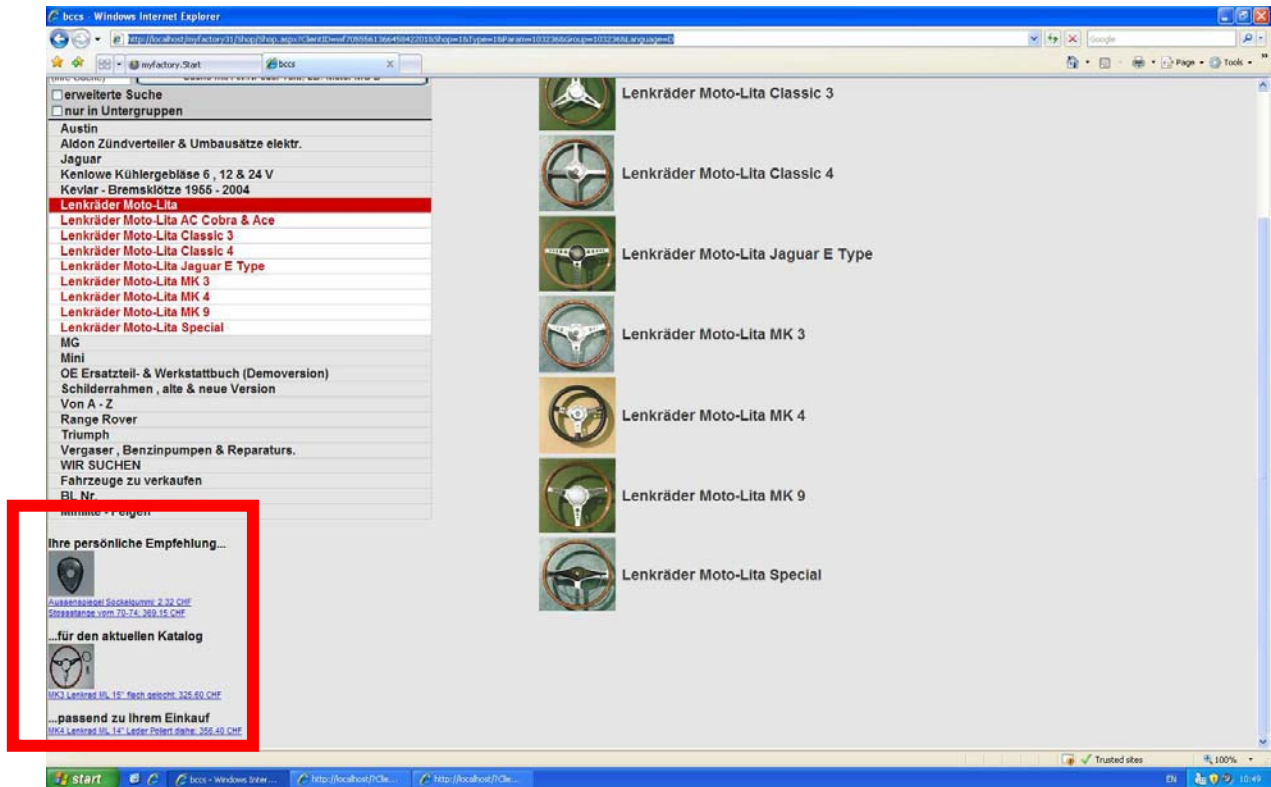


Abb. 5.3: Persönliche Empfehlungen

Das Empfehlungssystem ist zur Installation im myfactory.BusinessWorld verfügbar (Abb. 5.4). Natürlich lässt sich diese Funktionalität auch für andere Business Software implementieren, also nicht nur für die myfactory.BusinessWorld-Software, sondern auch in andere Produkte.

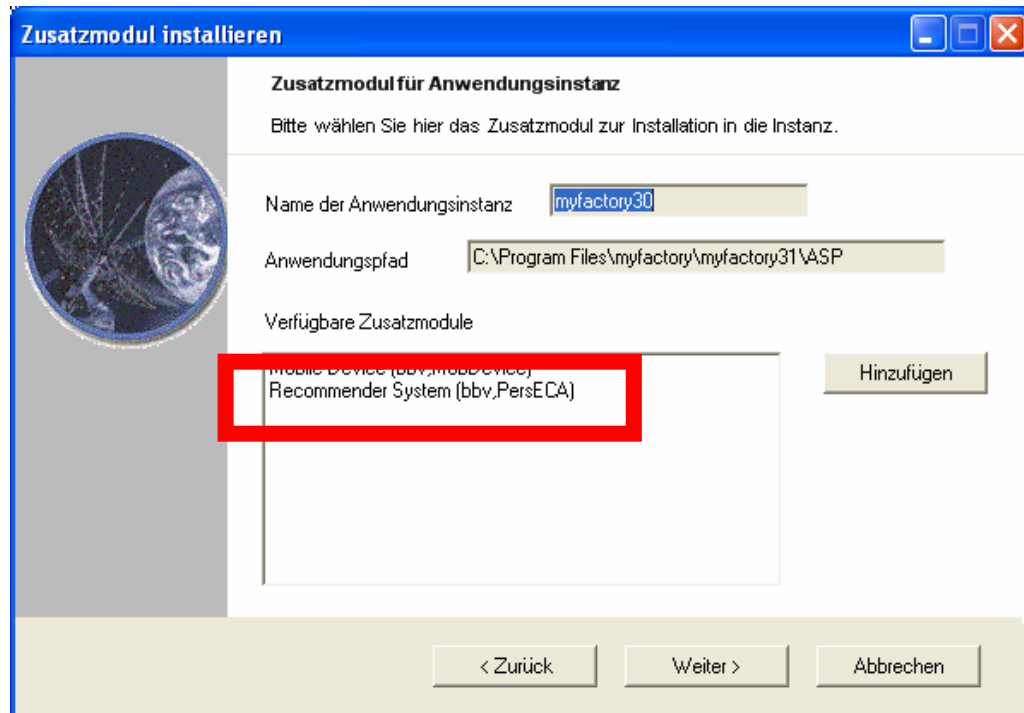


Abb. 5.4: Zusatzmodul Empfehlungssystem

5.5.5 Implementierung der Parametereinstellungen bei der Konfiguration eines E-Shops

Die Parametrisierung wurde entsprechend der in 5.4 dargestellten Diskussion implementiert. Die Parameter werden – mit Ausnahme des Alpha-Wertes und der Bewertung der Artikel im Clickstream – erst nach Berechnung der Empfehlungsliste angewandt, um z.B. bereits gekaufte Produkte herauszufiltern (falls der entsprechende Parameter so eingestellt ist).

Die Konfiguration erfolgt anhand der erarbeiteten Kategorisierung (5.4.4), wobei der Administrator individuelle Anpassungen vornehmen kann (d.h. die voreingestellten Werte können z.B. auf Produkt- oder Produktgruppenebene überschrieben werden). Die folgenden Abbildungen zeigen die Umsetzung der Parameter und damit gleichzeitig die Schritte, die bei der Konfiguration des Shops für den Anwender durchlaufen werden.



Abb. 5.5: Kategorisierung des E-Shops

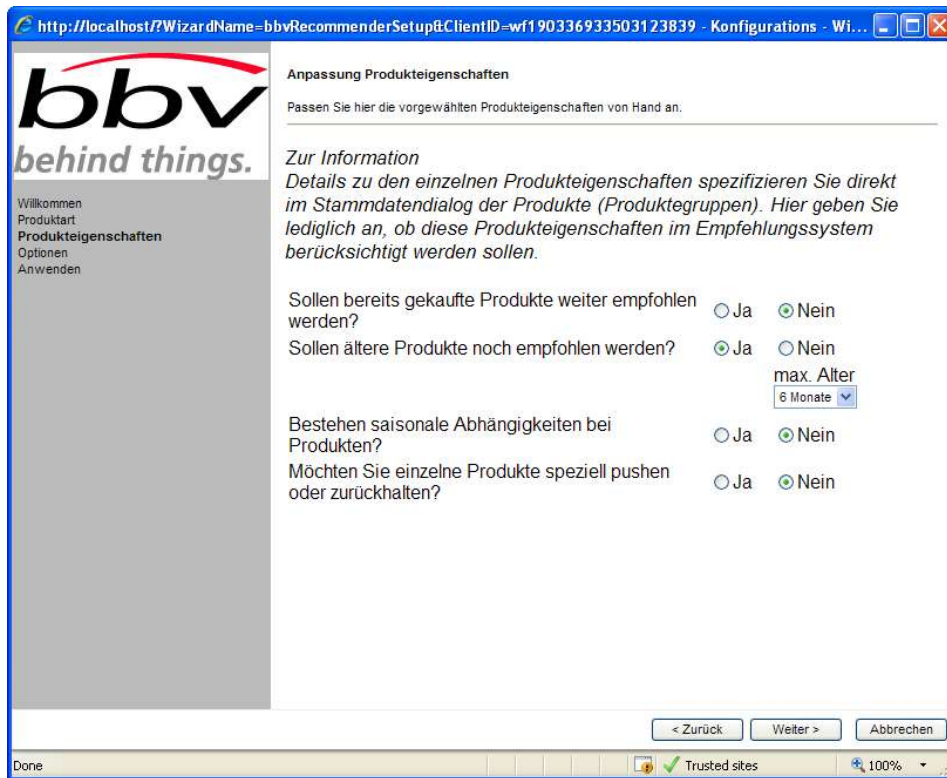


Abb. 5.6: Anpassung der Parameter für die Filterung der Empfehlungen

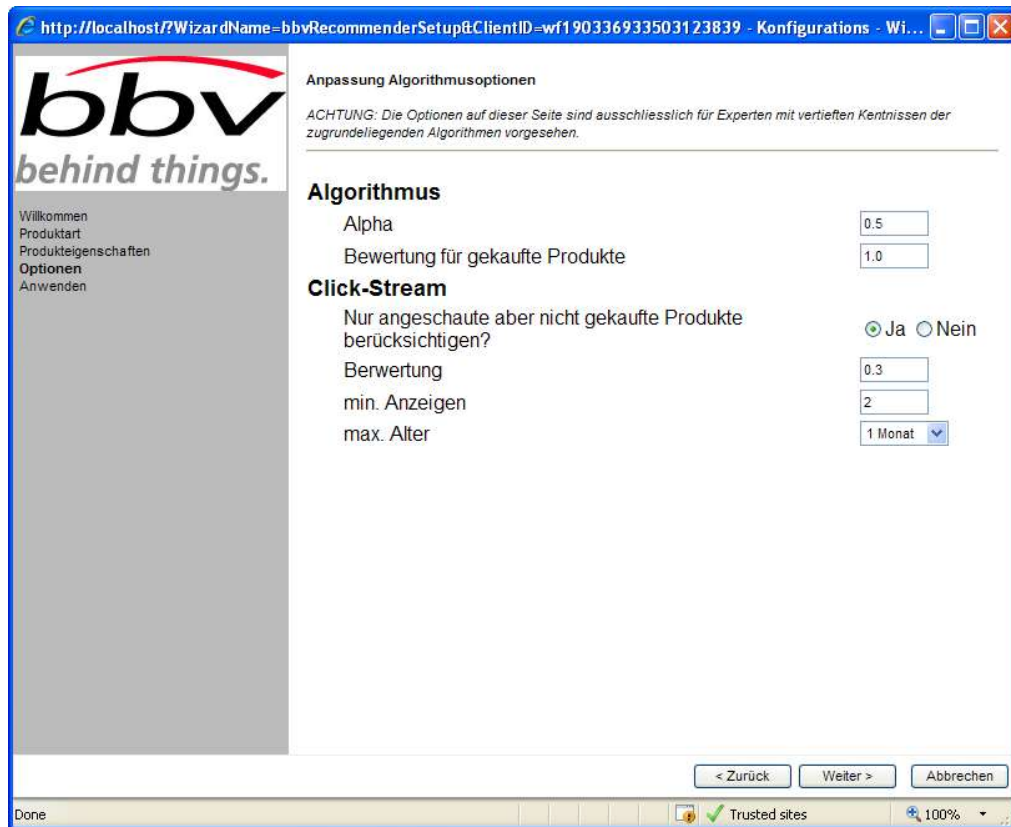


Abb. 5.7: Einstellung weiterer Parameter für das Empfehlungssystem

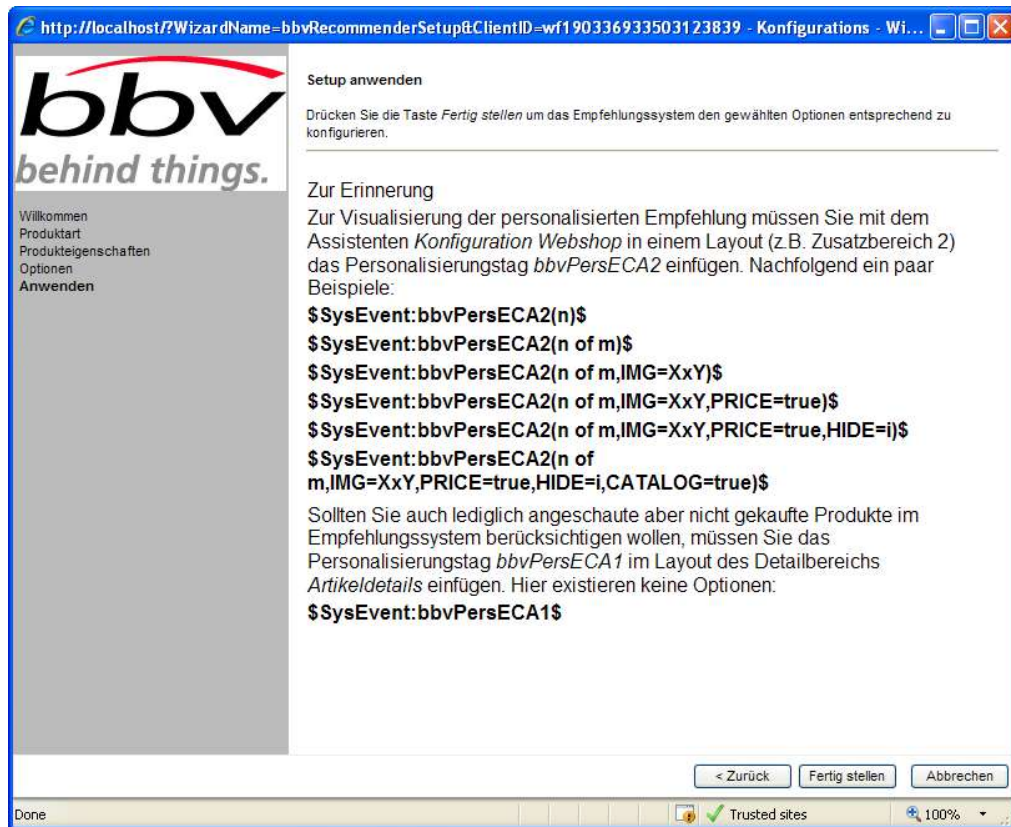


Abb. 5.8: Bestätigung der Konfiguration des Empfehlungssystems

Die Parameter *Saisonale Abhängigkeit* und *Nicht-Empfehlen* lassen sich in myfactory.BusinessWorld über so genannte *Kennzeichen* auf verschiedenen Ebenen (Produktgruppe, einzelnes Produkt, einzelner Artikel) überschreiben (Abb. 5.10). Gleiches gilt für die Push-Funktion:

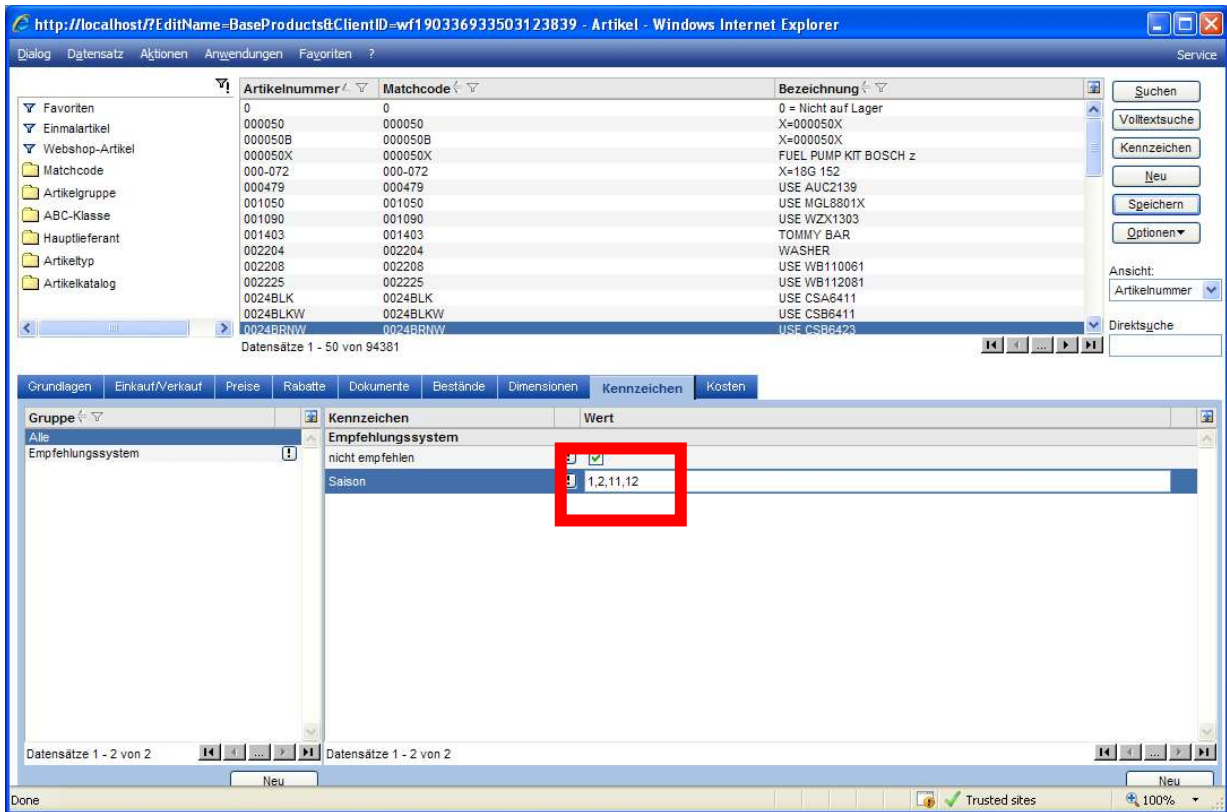


Abb. 5.10: Feinsteuerung auf Artekelebene

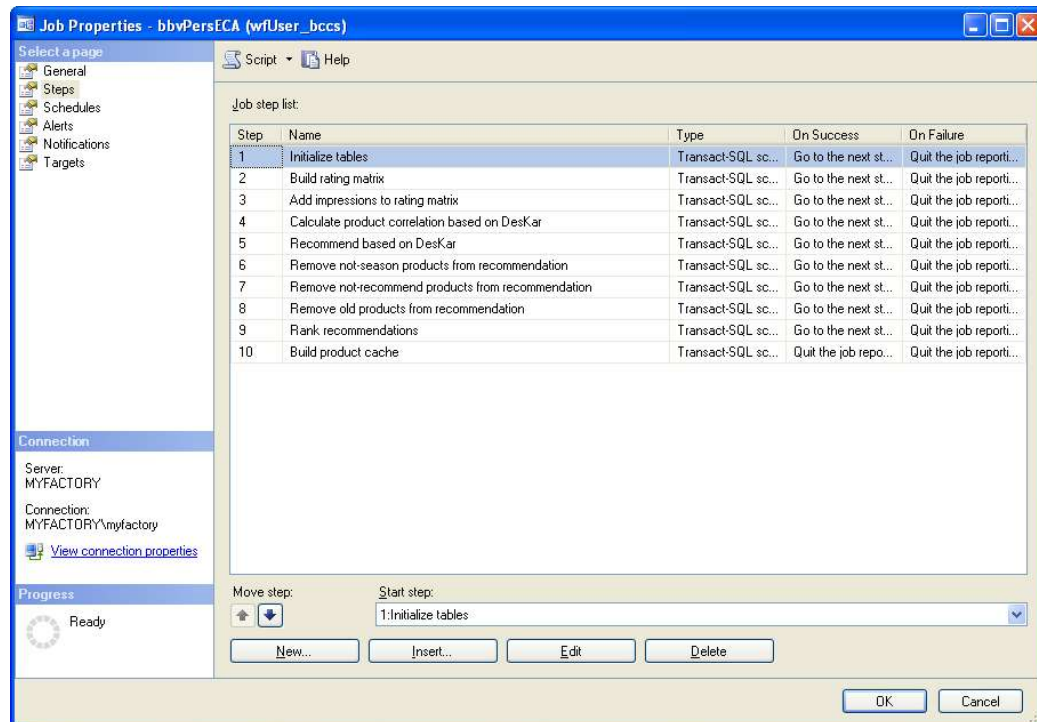


Abb. 5.11: Zusammenstellung der Datenbankaufträge (Jobs)

5.5.6 Weitere Implementierungen

Weitere durchgeführte Arbeiten umfassen:

- Bild in Empfehlung mit aufnehmen
- Preis in Empfehlung mit aufnehmen
- Empfehlung unter Berücksichtigung des aktuellen Clickstreams berechnen
- Empfehlung aufgrund der Warenkorbanalyse berechnen
- Ein-/Ausschalten von Wiederholungskäufen (werden Produkte, die bereits gekauft wurden, nochmals empfohlen oder nicht?)
- Ausführungsplan für Berechnung (wann werden die Berechnungen der Bewertungsmatrix durchgeführt, z.B. Nachts zu einer bestimmten Uhrzeit)
- Aufnehmen in myfactory-Verkaufsliste

5.5.7 Entwicklungsumgebung

Um die Entwicklungsumgebung einfach und effizient von einem zum anderen PC transferieren oder vervielfältigen zu können, wurde die ganze Umgebung (inklusive myfactory-Server-Installation) auf einem Microsoft Virtual PC installiert. Dies erlaubt z.B. zur Präsentation bei potenziellen Shopanwendern ein einfaches Kopieren der ganzen Lösung auf einen Verkaufslaptop.

Die gesamte Entwicklungsumgebung befindet sich somit auf einem Virtual PC Image, das die folgenden Komponenten beinhaltet:

- Windows XP
- Microsoft Internet Information Server
- .NET Framework 2.0
- SQL Server 2005 Developer Edition inklusive SQL Server Management Studio (SSMS)
- Visual Studio 2005
- myfactory.BusinessWorld 3.2

5.5.8 Entwicklungssoftware

Sämtliche Algorithmen und Datenbankzugriffe wurden in T-SQL als Stored Procedure (SP) geschrieben. Diese SPs werden in so genannten myfactory System Events (in C# geschrieben) verwendet. Die System Events selber werden vom myfactory Framework aufgerufen. Damit dies funktioniert, müssen die geschriebenen System Events in myfactory registriert werden. Diese Registrierung erfolgt via Datenbankeintrag in einer Tabelle

Folgende System Events und Stored Procedures wurden verwendet (in den folgenden Tabellen (Tab. 5.8 und Tab. 5.9) ist vor allem die rechte Spalte interessant, die die Umsetzung der oben beschriebenen Abläufe aufzeigt).

Tab. 5.8: System Events

EventPublisher	EventName	EventType	Beschreibung
Webshop	Login	2	Zuordnung SessionID zu Benutzer
CMSAddin	Login	2	Wie oben, aber für das Informationsportal von myfactory für Geschäftspartner und Kunden
Webshop	Detail	2	Clickstream aufzeichnen
CMS	StatusSet	2	Nach dem Setzen eines Statuswertes (Status 4 = Artikel für Detailseite). Verwendung: Clickstream aufzeichnen
Webshop	ClientCreated	2	Von myfactory eigens für dieses Projekt eingeführt: Nach Erstellen des Kundeninformationsobjekts bei einem neuen Benutzer (wird nicht bei der Vorschau verwendet). Verwendung: Speichern eines Cookies mit der SessionID
CMS	ClientCreated	2	Von myfactory eigens für dieses Projekt eingeführt: analog zu Webshop/ClientCreated, aber für das Informationsportal
Webshop	Global	2	Visualisierung der Empfehlung

Tab. 5.9: Stored Procedures

Stored Procedure	Parameter	Beschreibung
bbvFillRatingMatrix	@RatingForSale float = 1.0	Rating-Matrix auf Basis der getätigten Käufe abfüllen
bbvAddImpressionsToRatingMatrix	@MinImpressions int = 3 @LastImpressionWithinDays int = 7 @RatingForImpression float = 0.3	„Rating“ aus Clickstream der Rating-Matrix hinzufügen
bbvCalcDesKar	@Alpha as float = 0.5	Berechnung des Item-based Collaborative Filtering nach Deshpande und Karypis
bbvRecommendDesKar	@RecommendSoldProducts bit = 0	Berechnung der persönlichen Empfehlung
bbvRemoveNotSeasonProducts	-	Entfernen von Produkten aus der Empfehlungsliste, die zurzeit keine Saison haben
bbvRemoveNotRecommendProducts	-	Entfernen von Produkten aus der Empfehlungsliste, die der Shopadministrator nicht empfehlen will
bbvRemoveOldProducts	@MaxProductAge int = 360	Entfernen von Produkten aus der Empfehlungsliste, die zu alt sind
bbvCalcDesKarRanking	-	Ranking der Empfehlungen pro Benutzer
bbvFillProductCache	-	Abfüllen eines Produktcaches für eine effizientere Visualisierung der Empfehlung
bbvSaveSession	@CurClientID nvarchar(50), @PrevClientID nvarchar(50)	Cookie speichern
bbvIdentifySession	@ClientID nvarchar(50), @CustomerID int	Verbindung Benutzer zu Cookie (SessionID) ablegen
bbvSaveImpression	@ClientID nvarchar(50), @ProductID int	Clickstream abspeichern
bbvGetRecommendation	@ClientID nvarchar(50), @TopN int = 1, @Random int = 0, @Hide int = 0	Persönliche Empfehlung abfragen

Stored Procedure	Parameter	Beschreibung
bbvGetRecommendationUseCatalog	@ClientID nvarchar(50), @TopN int = 1, @Random int = 0, @Hide int = 0, @Catalog nvarchar(35) = N"	Persönliche Empfehlung innerhalb eines bestimmten Produktkatalogs abfragen
bbvGetShoppingCartRecommendation	@ClientID nvarchar(50), @TopN int = 1, @Random int = 0	Persönliche Empfehlung nur basierend auf dem aktuellen Shopping Cart abfragen

5.6 Feinabstimmung und potenzielle Verbesserungen

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

Im Rahmen des Projekts wurden verschiedene Alpha-Werte für den Item-based-Collaborative Filtering-Algorithmus nach Deshpande und Karypis [2004] getestet. Die Ergebnisse weichen kaum voneinander ab. Die Güte der Empfehlungen liegt immer bei ungefähr 15 %, d.h. 15 % der tatsächlich gekauften Produkte werden empfohlen, wenn man den Kauf zu Testzwecken vorab aus den Daten eliminiert (siehe 5.2.5). Im Vergleich zu den in verwandten Projekten erzielten Testergebnissen, ist diese Trefferquote sehr niedrig. Mit gut abgestimmten Empfehlungssystemen lassen sich durchaus Trefferquoten in Höhe von 60 % und mehr realisieren [Leimstoll/Alioski 2008; Quade et al. 2008]. Weshalb die Tests mit den Daten von BCCS nicht besser ausfallen, lässt sich nicht eindeutig begründen. Eine plausible Erklärung könnte auf der Heterogenität und der Modellabhängigkeit des Kundenbedarfs basieren (Ersatzteile für Oldtimer). Der Bedarf an einzelnen Artikeln hängt primär vom Zustand der Kundenfahrzeuge ab und ist möglicherweise so sehr zufallsgesteuert, dass auf der Basis von Kundenvergleichen (Collaborative Filtering) keine guten Empfehlungen berechnet werden können. Eine zweite Erklärung liefert die schlechte Datenbasis. Im Vergleich zur Vielzahl der im BCCS-Shop angebotenen Produkte und der vielen Kunden lagen nur wenige Transaktionsdaten vor.

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.

5.6.2 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.

Zur Identifizierung der schlechten Scouts muss aus der vorhandenen Bewertungsmatrix – wie schon bei der Hitrate-Berechnung (siehe 5.2.5) – ein Testset extrahiert werden. Auf den übrigen Daten (dem Trainingsset) werden anschliessend die Produktähnlichkeiten berechnet.

Der Scout-Ansatz setzt 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. Da aus dem BCCS-Shop nur relativ wenige Daten zur Verfügung stehen, ist hier die Anwendung des Scout-Ansatzes nicht sinnvoll.

5.6.3 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 (5.2.5), welche Datenbasis zu den besten Ergebnissen führt.

5.7 Vermarktung

Mit der Verfügbarkeit des Personalisierungsmoduls für myfactory.BusinessWorld, lässt sich die Funktionsweise der in diesem Projekt erarbeiteten Methoden, Konzepte und Abläufe gegenüber potenziellen Shopanwendern anschaulich demonstrieren. Hierdurch kann bbv ihr Know-how über personalisierte Empfehlungssysteme unterstreichen. Über das Know-how kann es gelingen, auch Shopbetreiber mit Fremdsystemen anzusprechen, um für diese eine individuelle Lösung zu entwickeln.

Das personalisierte Empfehlungssystem wurde in die Verkaufsliste von myfactory aufgenommen (Abb. 5.12). Zusätzlich wurde ein Produktflyer erstellt (Abb. 5.13).

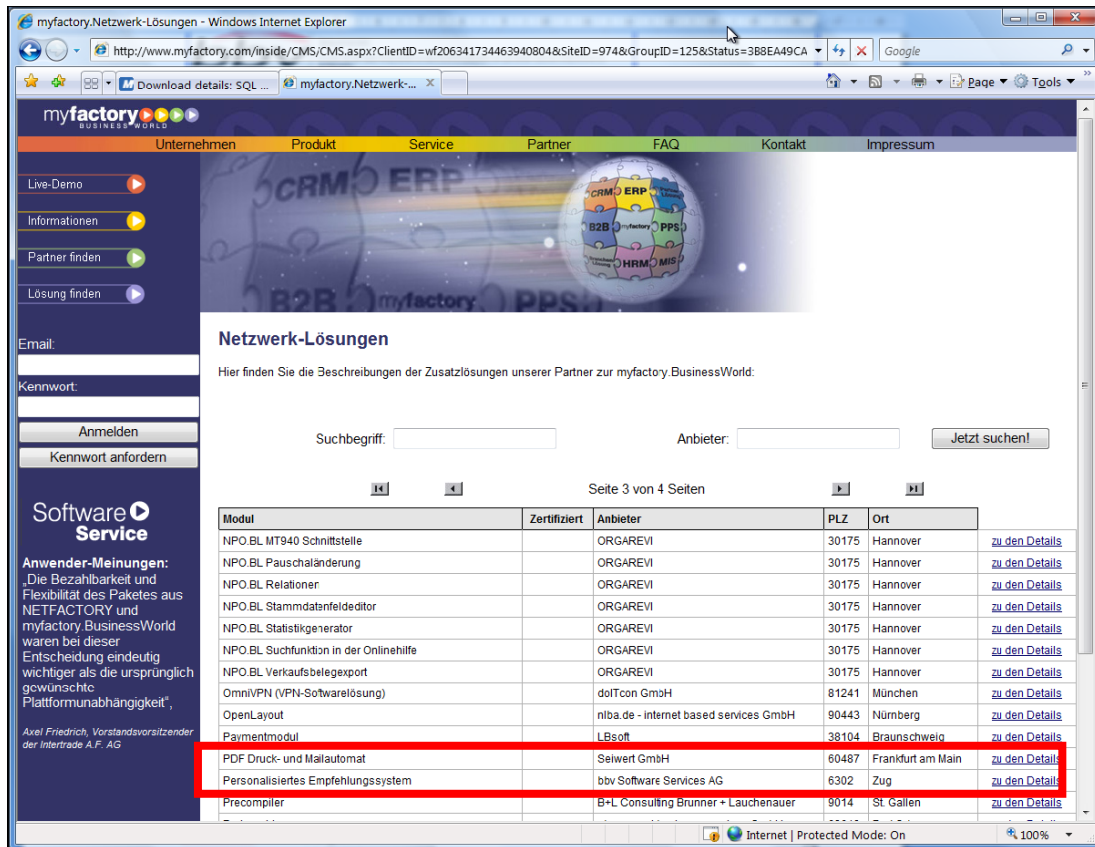


Abb. 5.12: Personalisierungsmodul in der myfactory Modulliste



FACTSHEET



Mit dem Zusatzmodul *Personalisiertes Empfehlungssystem* der bbv Software Services AG bringen Sie Ihren myfactory Webshop auf Vordermann. Erhöhen Sie die Kundenbindung und steigern Sie Ihren Umsatz dank persönlichen Einkaufsempfehlungen – für jeden Shop-Benutzer individuell zusammengestellt.

Eins, Zwei, Drei

Und schon haben Sie Ihr persönliches Empfehlungssystem in Betrieb. Die Installation ist äusserst einfach, bedient Sie sich doch dem Standard Installationswizard von myfactory. Eben so schnell geht die Konfiguration dank den mitgelieferten Vorlagen für diverse Shop Typen von statten. Platzieren Sie jetzt noch im Layout des Webshops Ihre Empfehlung und schon sind Sie fertig. Herzliche Gratulation zu Ihrem persönlichen Empfehlungssystem!

Features

Nur die wichtigsten Features in einer kurzen Übersicht zusammengestellt:

✓ Berücksichtigung aller bereits getätigten Einkäufe	✓ Auswahl wahlweise nur innerhalb der aktuellen Katalogposition
✓ Berücksichtigung des Click-Streams	✓ Ausschliessen einzelner Artikel
✓ Berücksichtigung saisonaler Artikel	✓ Warenkorbanalyse
✓ Freie Gestaltung der Empfehlung	✓ Einfachste Konfiguration
✓ via Webshop Konfigurator	✓ via Empfehlungssystem Konfigurator
✓ mit oder ohne Artikelbild	✓ mit vordefinierten Shop Typen
✓ mit oder ohne Preis	✓ individuell Parametrierbar
✓ beliebig viele Artikel	✓ mit Kennzeichen auf den Artikeln
✓ Zufällige Auswahl aus Best-of	✓ Unterdrücken bereits empfohlener Artikel

Basis

Unser Zusatzmodul basiert auf einer Standard Installation von myfactory.BusinessWorld ab Version 3.0 und dem Standard Webshop Modul der myfactory. Die Upgradekompatibilität zu künftigen Versionen von myfactory.BusinessWorld ist gewährleistet.

Unser Angebot

Gerne zeigen wir Ihnen, wie auch Sie unser Zusatzmodul für Ihren Shop gewinnbringend einsetzen können. Rufen Sie uns an und verlangen Sie Ihre ganz persönliche Offerte. Herr Roman Mayer ist gerne für Sie da (info@bbv.ch oder +41 41 429 01 11).



info@bbv.ch www.bbv.ch

Abb. 5.13: Produktflyer zum Empfehlungssystem zu myfactory

6 Projektmanagement

Dieses Kapitel beschreibt abschliessend, wie das in dieser Fallstudie beschriebene Teilprojekt mit bbv in das Gesamtprojekt PersECA II integriert war. Dabei wird auf den Ablauf im Gesamtprojekt (6.1) sowie im Detail auf den Ablauf im Teilprojekt eingegangen (6.2). Anschliessend werden die im Projektablauf gesammelten Erfahrungen beschrieben (6.3).

6.1 Ablauf des Gesamtprojekts

Abb. 6.1 zeigt die Projektorganisation des Gesamtprojekts PersECA II. Daraus ist auch die Eingliederung des Teilprojekts mit bbv ersichtlich. Die Teilprojekte wurden zeitversetzt durchgeführt, damit die jeweils gewonnenen Erkenntnisse aufbereitet und von Teilprojekt zu Teilprojekt weitergegeben werden konnten.

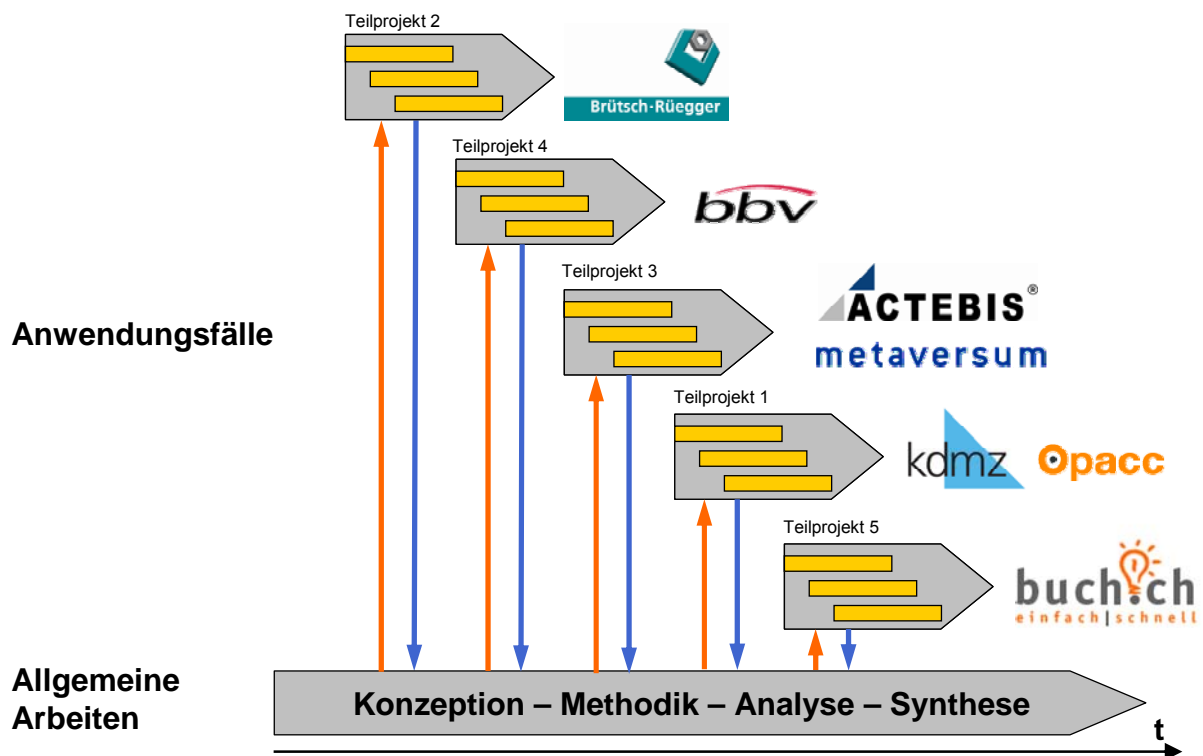


Abb. 6.1: Das Gesamtprojekt PersECA II

6.2 Vorgehensweise im Teilprojekt mit bbv

Die PersECA-II-Teilprojekte liefen nach einem einheitlichen Grundmuster ab (Tab. 6.1), das an die besonderen Gegebenheiten der jeweiligen Teilprojekte angepasst wurde. Im Teilprojekt mit bbv waren Modifikationen der Vorgehensweise notwendig (Tab. 6.2), weil kein Pilotkunde zur Verfügung stand, mit dem kundenspezifische Personalisierungsfunktionen erarbeitet werden konnten.

Tab. 6.1: Teilschritte im PersECA-II-Projekt

Hauptmeilensteine	Generelle Teilschritte
Personalisierungsfunktionen evaluieren und selektieren	1 Personalisierungsideen erarbeiten
	2 Anforderungen und Potenziale analysieren
	3 Optionen für Personalisierung prüfen
	4 Personalisierungsfunktionen priorisieren und auswählen
Profile und Technologien erarbeiten	5 Detailkonzept für Funktionen und Profile erarbeiten
	6 Lösungskonzept für Data Processing erarbeiten
Ergebnisse verwerten und dokumentieren	7 Anwendung und Vorgehensweise beschreiben (Fallstudie)
	8 Prototypische Umsetzung der ausgewählten Funktionen
Laufende Arbeiten	9 Projektleitung und Koordination

Die Termine in Tab. 6.2 geben einen groben Hinweis auf die Dauer der einzelnen Schritte im Teilprojekt. Die detaillierten Arbeitsschritte sind jeweils der entsprechenden Tätigkeit im generellen Ablauf (Tab. 6.1) zugeordnet. Dies wird mit der Schattierung der rechten Tabellenspalte verdeutlicht. Aus dem Detailablauf wird ersichtlich, dass im Teilprojekt mit bbv der Fokus auf der Umsetzung lag.

Tab. 6.2: Detailablauf des Teilprojekts mit bbv

Detaillierter, tatsächlicher Ablauf der einzelnen Arbeitsschritte im Teilprojekt mit bbv	Termine	Zugehöriger genereller Teilschritt
Hochschulinternes Kick-off Meeting	24.2.2006	
Kick-off-Meeting mit Wirtschaftspartner: Vorgehensweise und "Deliverables" im Projekt mit bbv abgestimmt	30.5.2006	1 Personalisierungsideen erarbeiten
Grobe Personalisierungsideen erarbeitet		
Optionen anhand der Personalisierungslandkarte identifiziert	29.6.2006	1 Personalisierungsideen erarbeiten
Informationsphase als Schwerpunkt festgelegt		
Ist-Aufnahme: myfactory-Software, E-Shops, Artikel- und Kundenstammdatenstrukturen	25.7.2006	2 Anforderungen und Potenziale analysieren
Literatur zur Gestaltung von Kundenprofilen analysiert [Bohrer/Holland, 2000]		3 Optionen für Personalisierung prüfen
Personalisierungsfunktionen priorisiert	25.7.2006	4 Personalisierungsfunktionen priorisieren und auswählen
Mit Uni Fribourg Algorithmen für Empfehlungssysteme erarbeitet	12.9.2006	3 Optionen für Personalisierung prüfen
Funktionen verfeinert und eingegrenzt	12.9.2006	4 Personalisierungsfunktionen priorisieren und auswählen
		5 Detailkonzept für Funktionen und Profile erarbeiten

Detaillierter, tatsächlicher Ablauf der einzelnen Arbeitsschritte im Teilprojekt mit bbv	Termine	Zugehöriger genereller Teilschritt
Experimente mit dem User-based-Collaborative-Filtering-Algorithmus durchgeführt	17.10.2006	5 Detailkonzept für Funktionen und Profile erarbeiten
		6 Lösungskonzept für Data Processing erarbeiten
		8 Prototypische Umsetzung der ausgewählten Funktionen
Item-based-Collaborative-Filtering-Algorithmus nach Deshpande/Karypis [2004] prototypisch umgesetzt	21.11.2006	5 Detailkonzept für Funktionen und Profile erarbeiten
		6 Lösungskonzept für Data Processing erarbeiten
		8 Prototypische Umsetzung der ausgewählten Funktionen
E-Shops zur Steuerung von Empfehlungen klassifiziert	26.1.2007	5 Detailkonzept für Funktionen und Profile erarbeiten
Cookies und Aufzeichnung von Clickstreams in myfactory implementiert	16.2.2007	8 Prototypische Umsetzung der ausgewählten Funktionen
E-Shop-Kategorien und Parameter zugeordnet (5.4)	16.2.2007	6 Lösungskonzept für Data Processing erarbeiten
Empfehlungssystem und Visualisierung in MyFactory prototypisch umgesetzt	30.3.2007	8 Prototypische Umsetzung der ausgewählten Funktionen
Berücksichtigung des Clickstreams in der Berechnung der Empfehlungen implementiert	15.6.2007	8 Prototypische Umsetzung der ausgewählten Funktionen
Parametrisierung implementiert (5.4.4)	13.9.2007	8 Prototypische Umsetzung der ausgewählten Funktionen
Input- und Output-Profile (5.3.1) dokumentiert	13.9.2007	7 Anwendung und Vorgehensweise beschreiben
Berücksichtigung der Parametrisierung vollständig implementiert	5.11.2007	8 Prototypische Umsetzung der ausgewählten Funktionen
Massnahmen zur Vermarktung der personalisierten Lösung entwickelt (5.7)		
Abschluss der Fallstudie	August 2008	7 Anwendung und Vorgehensweise beschreiben

6.3 Lessons Learned

6.3.1 Schwierigkeiten bei der Erarbeitung und Implementierung

Das grösste Problem im Teilprojekt mit bbv war, dass es keinen Pilotkunden gab, der sich am Projekt beteiligte. Daher war es nicht möglich, die Personalisierungsfunktionen gemeinsam mit einem Pilotkunden zu identifizieren und dann gezielt für diesen zu implementieren. Auf der anderen Seite ergab sich hierdurch die Chance, die generellen Möglichkeiten der Personalisierung zu analysieren und daraus Funktionen zu identifizieren, sie sich als Standard in einer ERP-II-Software implementieren liessen.

Ein weiteres Problem war die Datenlage. Die Daten, die aus einem Shop zur Verfügung standen, waren nicht sehr umfassend. Dennoch genügten die Daten, um die nötigen Grundlagen für die Umsetzung in der Business Software zu schaffen.

6.3.2 Begünstigende Faktoren

Die konkrete Umsetzung der ausgewählten Personalisierungsalgorithmen war nur möglich durch die enge und vertrauensvolle Zusammenarbeit zwischen den beteiligten Projektpartnern. Hervorzuheben sind insbesondere Henrik Stormer von der Universität Fribourg und Urs Gehrig von bbv. Die klare Zuordnung von Herrn Gehrig zu diesem Projekt und auch sein zeitlicher und beherzter Einsatz sind der massgebliche Grund dafür, dass die Projektziele erreicht werden konnten.

7 Zusammenfassung und Schlussfolgerungen

Die vorliegende Fallstudie beschreibt das PersECA-II-Teilprojekt mit der bbv Software Services AG. In diesem Teilprojekt ging es um die Identifikation, den Entwurf und die Implementierung von standardisierten Personalisierungsfunktionen für eine ERP-II-Lösung (myfactory.BusinessWorld). Am Projekt waren die folgenden Parteien beteiligt:

- bbv Software Services AG, Luzern
- Universität Fribourg, Department of Informatics, Information Systems Research Group
- Fachhochschule Nordwestschweiz, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik

Von Anfang an war bbv sehr an der Umsetzung interessiert. Durch die enge Zusammenarbeit mit der Information Systems Research Group der Universität Fribourg konnten konkrete Datenstrukturen und Algorithmen entwickelt und implementiert werden. Dies musste auf einer allgemeingültigen Ebene erfolgen, da kein Anwendungspartner im Projekt dabei war. Hierdurch ergab sich die Notwendigkeit, grundlegende Überlegungen zur Kategorisierung und Parametrisierung von E-Shops anzustellen und neben den eigentlichen Personalisierungsfunktionen auch die Parametrisierung zu implementieren.

Es ist gelungen, ein standardisiertes Empfehlungssystem zu entwickeln, das mit Hilfe einer Parametrisierung an die spezifischen Bedingungen einzelner Anwender angepasst werden kann. Nach Abwägung von Aufwand und Nutzen erweist sich der Collaborative-Filtering-Mechanismus als ein effizientes und effektives Instrument, um ein Empfehlungssystem zu realisieren. Transaktionsdaten reichen aus, um mit dem Deshpande/Karypis-Algorithmus gute Empfehlungen berechnen zu können. Allerdings müssen die berechneten Empfehlungen anschliessend anhand unterschiedlicher Kriterien gefiltert werden. Dieser Optimierungsprozess ist von grosser Bedeutung, um wirklich gute und sinnvolle Empfehlungen zu erzeugen.

Abschliessend kann festgestellt werden, dass das Projekt erfolgreich abgeschlossen werden konnte und bezüglich konkreter Umsetzung sehr viel erreicht wurde. Entscheidend war dabei die sehr gute Zusammenarbeit der beteiligten Organisationen und Mitarbeitenden. Als wichtig erwies sich eine konsequente Orientierung an der Ablaufplanung, auch wenn es bei den Projektpartnern betriebsbedingt immer wieder zu Unterbrechungen im Projekt kam. Entscheidend ist in solchen Fällen, dass die Projektleitung für die Wiederaufnahme im Projekt sorgt. Ein entscheidender Erfolgsfaktor bestand darin, dass der Wirtschaftspartner sich für die konzipierte Lösung begeisterte und diese konsequent umsetzte.

Literaturverzeichnis

- 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.
- bbv (2008): www.bbv.ch [Zugriff: Mai 2008].
- BCCS (2008): www.bccs.ch [Zugriff: Mai 2008].
- 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.
- Bohrer, Kathy; Holland, Bobby (2000): Customer profile exchange (CPExchange) specification, Version 1.0, <http://www.idealliance.org/cpexchange>, [Zugriff: August 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.
- Computerwoche (2006): Product Guide – myfactory.BusinessWorld 3.0, [<http://www.computerwoche.de/index.cfm?pid=355&pk=3188>, Zugriff: Mai 2008].
- Deshpande, Mukund; Karypis, George (2004): Item-based top-n recommendation algorithms, in: ACM Transactions on Information Systems, 22(1), 2004, 143–177.
- 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; 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.

- Leimstoll, Uwe; Stormer, Henrik; Schneider, Raoul; Quade, Michael; Pülz, Michael (2008): Profile und Technologien der Personalisierung, Basel: Fachhochschule Nordwestschweiz, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik, Arbeitsberichte der Hochschule für Wirtschaft FHNW, Arbeitsbericht Nr. 37 des Competence Center E-Business Basel, 2008.
- 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.
- myfactory (2008): [<http://www.myfactory.com/inside/CMS/CMS.aspx?ClientID=wf434354122004556358&SiteID=875&GroupID=0&Status=3187AF93A5A7ABA3AA94A495&Language=D>], Zugriff: Mai 2008].
- PayPal (2007): eCommerce 2007, EuPDRResearch 09/2007, Berichtsband, PayPal Deutschland GmbH, Europarc-Dreilingen [https://www.paypal-deutschland.de/presse/media/0/12028044428310/ecommerce_2007.pdf], Zugriff: Mai 2008].
- 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.
- 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.
- Risch, Daniel (2007): Kundenprofile im E-Commerce, Basel: Fachhochschule Nordwestschweiz FHNW, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik IWI, Arbeitsbericht E-Business Nr. 29, 2007.
- 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.

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 Brütsch/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.

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.

Leimstoll, Uwe; Stormer, Henrik; Schneider, Raoul; Quade, Michael; Pülz, Michael (2008): Profile und Technologien der Personalisierung, Basel: Fachhochschule Nordwestschweiz, Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik, Arbeitsberichte der Hochschule für Wirtschaft FHNW, Arbeitsbericht Nr. 37 des Competence Center E-Business Basel, 2008.

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