

Personalisierte Suche in E-Commerce-Anwendungen unter Einbeziehung semantischer Aspekte

Paul Schalow¹, Florian Goertz¹,

¹ Humboldt Universität zu Berlin, Institut für Wirtschaftsinformatik
{schalow, goertzfl}@wiwi.hu-berlin.de

Abstract. Personalization helps customers to retrieve searched products, services and information more effectively. Currently no elaboration regarding personalized search in the e-commerce exists that also covers differentiated fields of application. This paper gives an overview of a characterization and conceptualization of personalized search classified by the phases of mining of customer related information, storing and using them for a personalizing algorithm. Related research is analyzed to show how personalized search can be enhanced with semantic elements. The findings are used to build up e-commerce scenarios that lead to differentiated recommendations of using personalized search.

Keywords: personalized search, semantic, ontology, e-commerce

1 Einleitung

Der elektronische Handel von Gütern und Dienstleistungen im Internet hat heutzutage einen enormen Stellenwert im weltweiten Handel eingenommen. Benutzer des Internets haben die Möglichkeit nahezu jedes Gut bzw. jede Dienstleistung über das Internet zu beziehen. Suchmaschinen wie Google, Bing oder Yahoo ermöglichen Benutzern des Internets gezielt nach Information suchen zu können. In E-Commerce Applikationen spielen integrierte Suchmaschinen ebenfalls eine immer wichtiger werdende Rolle. Betreiber von E-Commerce Anwendungen sind gehalten, ihr Produktportfolio dem potenziellen Kunden möglichst transparent zu gestalten und dem Benutzer passende Informationen zu liefern. Eine Herausforderung hierbei ist, dass Benutzer einer Suchmaschine typischerweise deren Matching-Prozess nicht kennen und daher selten in der Lage sind, optimale Suchkriterien zu verfassen, um ein optimales Suchergebnis respektive relevante Informationen zu erhalten. Zudem sind Suchanfragen in der Regel knapp formuliert, sollen aber dennoch unterschiedlichste Informationsbedarfe der Benutzer befriedigen[10]. Durch die Erweiterung von Suchmaschinen um die Fähigkeit Semantik und Kontext zu „verstehen“, können Suchergebnisse optimiert und auf die Interessen des Benutzers angepasst werden. Ontologie-gestützte und personalisierte Suchmechanismen dienen als Möglichkeiten zur Umsetzung dieser Fähigkeiten. Beide Konzepte werden bis zum heutigen Zeitpunkt nur spärlich verfolgt.¹ Mögliche Erklärungen sind fehlende Standards im Einsatz dieser Konzepte und der zum heutigen Zeitpunkt vorliegende geringe Reifegrad konkreter Umsetzungen. Das Ziel der vorliegenden Arbeit ist es, eine Übersicht der vorhandenen Konzepte der personalisierten Suche zu erstellen. Weiterhin soll untersucht werden, wie sich eine personalisierte Suche mit semantischen Mittel erweitern lässt. Darüber hinaus werden Empfehlungen für einen Einsatz im E-Commerce erarbeitet. Als Ziel- und Leitfragen dieser Arbeit werden somit im Weiteren folgende Forschungsfragen beantwortet:

Wie lässt sich eine personalisierte Suche für E-Commerce charakterisieren?

Existieren Ansätze einer personalisierten Suche, die mit semantischen Mitteln beeinflusst werden? Welche Aussagen und Empfehlungen lassen sich bezüglich des Einsatzes in E-Commerce Umgebungen identifizieren?

Um die Forschungsfragen zu beantworten, wird im Kapitel 2 untersucht, wie genannte Suchkonzepte in E-Commerce Systemen eingesetzt werden können und darüber hinaus, wie durch die Kombination beider, ein eventueller Mehrwert generiert werden kann. Des Weiteren werden die Eigenschaften untersucht, die mit der Verwendung von Suchmechanismen in einer E-Commerce Umgebung im Gegensatz zum World Wide Web verbunden sind. In Kapitel 3 werden mögliche Einflussfaktoren analysiert, welche den Einsatz konkreter Technologien rechtfertigen, um folglich ein Werkzeug zur Entscheidungsfindung zum effektiven Einsatz von Suchtechnologien zur Verfügung zu stellen.

¹ Diese Aussage erfolgt auf Basis einer Inspektion von Amazon und Ebay zum gegenwärtigen Zeitpunkt

Die Sichtung der gegenwärtigen Literatur brachte zum Ergebnis, dass Forschungen zur Beantwortung der gestellten Forschungsfrage derzeit nicht existieren. Vielmehr konzentrieren sich gegenwärtige Untersuchungen auf die Analyse und Verbesserung bestehender Konzepte zur persönlichen Suche oder zur ontologie-gestützten Suche im World Wide Web. [3, 9, 10] Konzeptübergreifende Analysen existieren derzeit nicht, darüberhinaus sind keine Untersuchungen mit dem engeren Blickwinkel einer im World Wide Web abgeschlossenen Applikation (wie einem Web-Shop) bekannt.

2 Adaptive Suche in E-Commerce-Anwendungen

2.1 Eingrenzung und Definition von E-Commerce

Das rasante Wachstum der Verfügbarkeit digitaler Informationen im Internet und die zunehmende Nutzung von Informations- und Kommunikationstechnologien durch Unternehmen, aber auch durch private Haushalte, haben bis in die heutige Zeit dazu geführt, dass neue Märkte und neue Geschäftsmodelle entstanden sind. E-Business und E-Commerce haben in den letzten Jahren immer mehr an Bedeutung gewonnen. Im Folgenden werden beide Begriffe voneinander abgegrenzt, um somit eine Definition für E-Commerce-Anwendungen geben zu können. Darüber hinaus wird aufgezeigt, wie sich E-Commerce-Anwendungen von Anwendungen im World Wide Web abgrenzen.

Die Begriffe E-Commerce und E-Business wurden in den Jahren ihrer Entstehung oft synonym verwendet. Ein Ansatz zur Definition und Differenzierung basiert auf den Ausführungen von Wirtz, die in der folgenden Arbeit Anwendung finden wird. E-Business ist demnach als die Anbahnung, vollständige Unterstützung, Abwicklung und Aufrechterhaltung von Leistungsaustauschprozessen über elektronische Netzwerke zu verstehen[24]. Der Begriff „E-Commerce“ hingegen wird durch Wirtz als Anbahnung, Aushandlung und Abwicklung von Aktivitäten, die im Zusammenhang mit dem Handel von Gütern oder Dienstleistungen über elektronische Netzwerke stehen, verstanden. E-Commerce grenzt sich somit durch die Fokussierung auf den Kauf und Verkauf von Gütern vom E-Business ab, kann laut Definition aber als ein Teil des E-Business verstanden werden. Die Definition unterteilt die Akteure im E-Commerce in Anbieter und Nachfrager² von Gütern bzw. Dienstleistungen. Akteure können Unternehmen, private Haushalte sowie öffentliche Institutionen sein.[24]

Eine E-Commerce-Anwendung wird demnach als Anwendungsprogramm bzw. Software verstanden, welches den Akteuren Funktionen zur Anbahnung, Aushandlung sowie Abwicklung von Handelstransaktionen zur Verfügung stellt. Eine konkrete Ausprägung dieser Anwendungsprogramme stellen vor allem E-Shops oder Web-Shops dar.

² Nachfrager werden im Rahmen dieser Arbeit auch als Kunden bzw. Nutzer bezeichnet

2.2 Personalisierte Suche

Übersicht

Im Fokus der personalisierten Suche steht im Gegensatz zu traditionellen Suchverfahren der Lösungsansatz die Suchergebnisse zu personalisieren. Jeder Form der Personalisierung liegt ein Modell eines Benutzers zugrunde[3]. Diese Modelle können in unterschiedlichen Formen erstellt und gespeichert werden[2]. Mit Hilfe des Nutzermodells können personalisierte Suchen durchgeführt werden. Dabei müssen andersgeartete Ansätze der Umsetzung einer personalisierten Darstellung der Suchergebnisse in Betracht gezogen werden. Die Personalisierung der Suche unterliegt somit einem Prozess und umfasst im Wesentlichen die folgenden Phasen[2]:

- Erfassen der Nutzerdaten,
- Erstellen eines Nutzerprofils,
- Personalisierung der Suchergebnisse.

Erfassen von Nutzerdaten

Die erste Phase des Personalisierungsprozesses dient der Erfassung von Nutzerdaten. Wesentliche Nutzerdaten im Sinne einer Personalisierung umfassen die Merkmale Wissen, Interesse, Ziele, Hintergrund und individuelle Wesenszüge und dienen der Bildung und Modellierung eines Nutzerprofils, wobei das Nutzerinteresse die wichtigste Rolle einnimmt[1]. Eine notwendige Voraussetzung zur Erfassung von Nutzerdaten ist die Fähigkeit, einen Benutzer bzw. Benutzergruppen eindeutig zu identifizieren[2].

Nach der erfolgreichen Identifikation und Zuordnung des Nutzerprofils müssen die Nutzerdaten erfasst und gesammelt werden. Information über den Benutzer können dabei

- Explizit oder
- Implizit

erfasst werden[2, 3].

Identifikation von Benutzern

Zur Identifikation von Benutzern bedarf es technischer Umsetzungen, die im Folgenden dargestellt werden. Die Eignung der jeweiligen Umsetzung innerhalb von E-Commerce-Anwendungen wird darüber hinaus diskutiert.

Cookies – Ein Cookie ist ein Programmcode, der auf dem Computer des Benutzers (Client) gespeichert wird. Das Cookie selber enthält Daten, die im Wesentlichen dazu beitragen, einen Benutzer auch über einen längeren Zeitraum eindeutig zu identifizieren. Eine Einrichtung des Cookies erfolgt in der Regel für Benutzer unbemerkt, was den Aufwand zur Nutzung erheblich minimiert. Eine Installation zusätzlicher Client-seitiger Komponenten ist nicht notwendig. Der Einsatz von Cookies ist speziell in E-

Commerce-Anwendungen verbreitet. Cookies werden demnach eine Eignung unterstellt. Nachteile bei dem Einsatz von Cookies können entstehen, wenn

- Ein Benutzer mehrere Computer benutzt.
- Mehrere Benutzer einen Computer (bei gleichem Nutzerprofil auf diesem Computer) benutzen.
- Cookies durch den Benutzer gelöscht bzw. durch administrative Restriktionen unterbunden werden.

Als Folge dessen kann eine eindeutige Identifizierung bei Eintritt dieser Szenarien nicht mehr sicher gestellt werden.

Session IDs – Eine Session ID ist ein eindeutiger Identifier, der einem Benutzer während einer Session zugeordnet wird. Der Aufwand zur Einrichtung einer Session-ID ist aus Sicht eines Benutzers transparent und stellt somit keine Barriere dar. Der Nachteil, bei der Verwendung von Sessions-IDs ist die Gültigkeit für nur eine Session. Bei Verlassen des Systems durch den Benutzer wird die ID verworfen.

Logins – Durch die Vergabe von Benutzernamen und Kennwort können sich Benutzer an einem System anmelden und eindeutig identifizieren. Der wesentliche Vorteil eines Logins besteht in der Möglichkeit der eindeutigen Identifizierung auch nach einem langen Zeitraum. Der Aufwand der Anmeldung und vor allem der erstmaligen Registrierung stellt eine Hürde für den Benutzer dar. Des Weiteren bedarf der Login weiterer Techniken (z.B. Cookies, Session-IDs) um den Benutzer während der laufenden Session zu identifizieren.

Software Agents – Ein Software Agent ist ein Programm, welches Client-seitig ausgeführt wird und dadurch eine Identifikation ermöglicht. Vorteil bei der Verwendung von Software Agents ist die eindeutige Zuordnung von Benutzer über einen langen Zeitraum. Defizite sind analog zu Cookies. Darüber hinaus stellt die Installation eines Software Agents einen Aufwand und somit eine Hürde für die Verwendung dar.

IP-Adresse – Die IP-Adresse kann einen Benutzer innerhalb des Internets identifizieren. Die IP-Adresse stellt allerdings ein sehr ungenaues Mittel zur Identifikation dar. Gründe dafür sind vor allem, dass viele User eine IP-Adresse (zeitgleich) nutzen können sowie die Tatsache, dass sich die IP-Adresse stets ändern kann.

Table 1. Bewertung von Technologien zur Identifizierung von Benutzern

Technologie	Usability ³	Effektivität der Identifikation ⁴	Eignung für E-Commerce-Anwendungen ⁵
Cookie	hoch	hoch	gut geeignet – Effektiver in Kombination mit Login.
Session ID	hoch	niedrig	bedingt geeignet – Effektiver in Kombination mit Login.
Login	mittel	hoch	geeignet – bedarf zur vollständigen Identifizierung einer weiteren Technologie.
Software Agent	niedrig	hoch	weniger geeignet – der Aufwand der Installation stellt eine Hürde dar.
IP	-	sehr niedrig	weniger geeignet – Die Effektivität ist sehr gering.

In Tabelle 1 werden beschriebene Technologien zur Identifikation von Benutzern und ihrer Eignung im E-Commerce bewertet. Im Fokus stehen dabei E-Commerce-Anwendungen gemäß Abschnitt 2.1. Sonderformen (z.B. aus dem Bereich B2B) finden an dieser Stelle keine Betrachtung und müssen gegebenenfalls gesondert bewertet werden.

Explizite Erfassung

Die explizite Erfassung basiert auf Angaben von Information von und über den Benutzer. Grundsätzlich können die Ansätze und korrespondierender Umsetzungen expliziter Informationserfassung durch die zeitliche Einordnung in einen Suchprozess unterschieden werden.

Durch Angabe von Information durch den Benutzer ist es möglich ein Profil zu erstellen. Ein mögliches Werkzeug zur Erfassung dieser Informationen ist ein Fragebogen, indem in geeigneter Weise relevante Interessen abgefragt und erfasst werden.

Eine weitere Methode zur expliziten Informationserfassung stellt das Feedbackverfahren dar. Der Benutzer wird mit der Aufgabe vertraut, die Suchergebnisse und somit

³ Usability ist ein wesentlicher Indikator zur Erfolgsbewertung einer Technologie im E-Commerce. Als maßgeblicher Einflussfaktor dient der Aufwand zur Benutzbarkeit einer Technologie. Usability wird im Rahmen dieser Arbeit auf den Aufwand reduziert, wobei ein hoher Aufwand eine geringe Usability impliziert.

⁴ Die Effektivität beschreibt den Grad der Zielerreichung. Eine hohe Effektivität impliziert demnach einen hohen Grad der (hier zu bewertenden) Identifikation.

⁵ Der Benutzung einer E-Commerce-Anwendung wird unterstellt, dass die Einrichtung und Konfiguration eines zusätzlichen Systems mit hohen (nicht monetären) Kosten verbunden sind.

die Effektivität der Personalisierung zu bewerten. Eine Bewertung erfolgt in der Regel durch die Einschätzung der Relevanz einer Information, beispielsweise durch eine Bewertung einer angezeigten Webpage durch den Benutzer.

Weitere - von einer Suche unabhängigen - Prozesse eines Systems können als Informationslieferanten dienen. Denkbar wäre die Erfassung demografischer Daten (Postleitzahl, Ort, Alter etc.) während eines Verkaufsprozesses, welche zur weiteren Auswertung dienen können. Interessen können in der Regel nicht direkt abgeleitet werden, jedoch können durch beispielsweise Mining-Verfahren Aussagen über eine Person bzw. Personengruppe gemacht werden.

Implizite Erfassung

Im Gegensatz zur expliziten Erfassung der Benutzerinformationen können durch die Beobachtung des Nutzerverhaltens Aussagen über benutzerspezifische Interessen getroffen werden. Zur Auswertung dienen Technologien, die primär durch den Benutzer besuchten Webpages auswerten und anhand dessen, korrespondierende Interessen ableiten. Folgende Umsetzungen finden verstärkt Anwendung:[2]

Browser Cache – Der Cache eines Browser (auf Seiten des Nutzers) speichert alle in der Vergangenheit besuchten Webpages. Die Speicherung erfolgt aus Sicht des Benutzers transparent. Die Auswertungen des Browser Caches, welche dezentral auf Seiten des Nutzers positioniert sind, müssen manuell an das zentrale Suchsystem übermittelt werden. Ein automatischer Zugriff ist in der Regel nicht umsetzbar. Dies stellt zugleich den wesentlichen Nachteil dieser Technologie dar. Der Versand der Daten ist für den Benutzer mit einem hohen Aufwand verbunden.

Browser Agent – Der Browser Agent ist ein auf Nutzerseite installiertes Programm und kann die Historie besuchter Webpages und prinzipiell darüber hinaus zusätzliche Informationen (Verweildauer, Click-Through-Rate, Käufe, etc.) erfassen. Die Installation und der damit verbundene Aufwand stellt ein Nachteil der Anwendung dar.

Desktop Agent – Eine weiterführende Form des Browser Agents ist ein Desktop Agent. Dieser ist prinzipiell in der Lage alle auf dem Computer des Nutzers gespeicherten (relevanten) Dokumente auszuwerten. Diese Umsetzung stellt theoretisch die stärkste und umfangreichste Form der impliziten Informationserfassung dar. Auch hier gilt ein verstärkter Aufwand durch die Installation. Obwohl die Thematik „Privacy“ nicht Thema der gegenwärtigen Erörterung ist, kann die Aussage nicht verwehrt werden, dass Desktop Agents einen enormen Eingriff in die Privatsphäre verursachen können.

Webserver Log – Im Gegensatz zu den bisherigen Umsetzungen werden Webserver-Logs zentral (auf Server-Seite) ausgeführt, eine Installation auf dem Client entfällt. Die Verwendung von Webserver-Logs verhält sich für den Benutzer somit transparent. Mittels Logfiles lassen sich neben besuchten Webpages eine Vielzahl weiterer

Informationen sammeln.⁶ Daten, die durch Web-Server Logs erhoben werden können, sind unter anderem:

- Die Verweildauer auf einer Seite⁷,
- Herkunftsort,
- Browser,
- Gerät (des Clients),
- Such-Queries,
- Uhrzeit der Seitenaufrufe.

Wie der Liste zu entnehmen ist, eignet sich ein Webserver-Log auch um demographische Information über den Benutzer zu erfahren.

Search Log – Ein Search Log ist eine Logdatei, die Suchanfragen innerhalb eines Suchsystems fortlaufend dokumentiert. Eine Zuordnung einer Suchanfrage zu einem Benutzer ist prinzipiell denkbar. Das „Logging“ findet anbieterseitig und somit transparent für den Benutzer statt. Search Logs liefern keine zusätzlichen Informationen darüber, welche Seiten der Nutzer als relevant einstufen könnte.

Table 2. Bewertung der Technologien zur impliziten Erfassung von Nutzerinformationen

Technologie	Usability ⁸	Effektivität der Erfassung ⁹	Eignung für E-Commerce-Anwendungen ¹⁰
Browser Cache	Sehr niedrig	Mittel	Nicht geeignet
Browser Agent	Niedrig	Hoch	Nicht geeignet
Desktop Agent	Sehr niedrig	Sehr hoch	Nicht geeignet
Webserver-Log	Hoch	Hoch	Sehr gut geeignet ¹¹
Search-Log	Hoch	niedrig	Weniger Geeignet

In Tabelle 2 werden beschriebene Technologien zur impliziten Erfassung von Nutzerinformationen hinsichtlich ihrer Eignung im E-Commerce bewertet. Browser Cache, Browser Agent und Desktop Agent wurden als nicht geeignet klassifiziert. Ausschlaggebend für diese Klassifizierung ist der mit der Technologie verbundene Einrichtungsaufwand. Es wird angenommen, dass ein Nutzer nicht gewillt ist, den Instal-

⁶ Zur Auswertung dieser Daten kann Software eingesetzt werden, welche Webserver-Logs übersichtlich und grafisch darstellen kann. Bei dieser Software handelt es sich vor allem um Web-Analytics.

⁷ Die Verweildauer kann u.a. aufgrund der Zustandslosigkeit des HTTP-Protokolls nicht immer genau gemessen werden.

⁸ Usability wurde im letzten Abschnitt bereits definiert.

⁹ Effektivität wurde im letzten Abschnitt bereits definiert.

¹⁰ Der Benutzung einer E-Commerce-Anwendung wird unterstellt, dass die Einrichtung und Konfiguration eines zusätzlichen Systems mit hohen (nicht monetären) Kosten verbunden sind.

¹¹ Gilt nur in Verwendung mit Web-Analytics.

lationsaufwand zu tragen und dass der Nutzen einer Personalisierung nicht die Kosten, verursacht durch Sicherheits- und Privacy-Bedenken, übersteigt.

Bewertung expliziter und impliziter Erfassung von Nutzerinformationen

Die Effektivität beider Ansätze zur Erfassung von Nutzerinformationen wurde in Studien verglichen [4, 5]. Die Studien brachten zum Ergebnis, dass sich beide Ansätze hinsichtlich ihrer Effektivität geringfügig unterscheiden. Als wesentlicher Nachteil und hemmender Einflussfaktor der expliziten Erfassung wird die fehlende Bereitschaft eines Benutzers zur expliziten Beantwortung bzw. zur Vergabe von Feedbacks genannt[3]. Eine Studie für den Bereich E-Commerce widerlegt diese Aussage in einzelnen Aspekten[6]. Probanden wurden innerhalb einer E-Commerce Umgebung Fragen durch das System (präziser durch Dialoge) gestellt. Bei den Fragen handelte es sich um kaufrelevante Fragen, allgemeine Fragen, die keine Relevanz zum Kaufprozess besitzen und darüber hinaus, personenbezogene Fragen. Das Ergebnis zeigte wider Erwartens eine hohe Resonanz. Folgende Aussagen beruhen daher auf der Annahme, dass in E-Commerce Umgebungen explizite Fragen in geeigneter Form einen essentiellen Beitrag zur Erfassung von Benutzerinformationen leisten können.

Des Weiteren wird angenommen, dass auf Basis einer expliziten Datenabfrage genauere Aussagen, mit einem höheren Wahrheitsgehalt erhalten werden können.

Begrenzt wird die explizite Abfrage von Daten (z.B. durch Fragebögen) durch den Umfang und die Komplexität. Eine Eignung findet daher primär in homogenen Suchräumen (mit einer geringen Anzahl von Interessensgebieten) statt. In stark heterogenen Räumen (mit einer hohen Anzahl an Interessensgebieten) verliert die explizite Datenerhebung an Eignung. Ein weiterer Nachteil ist die Schwierigkeit zur Erfassung von Kurzzeitinteressen des Benutzers, da die Umsetzung zur expliziten Erfassung dieser als nicht praktikabel erscheint.

Das Schlussfolgern auf Nutzerinteressen anhand implizit erfasster Daten birgt eine gewisse Unschärfe. Darüber hinaus stellt das Sammeln von negativem Feedback eine Grenze impliziter Technologien dar[2]. Mit wachsender Anzahl an Informationen über den Benutzer steigt die Effektivität der Personalisierung. Stärken einer umfangreichen Datensammlung sind vor allem im heterogenen Raum (mit einer hohen Anzahl an Interessensgebieten) zu erwarten. Bei Verwendung von demografischen Daten bzw. Schlussfolgerung von technischen Daten auf Benutzergruppen (beispielsweise wird anhand des Webserver-Log erkannt, dass der Nutzer ein iPhone benutzt) können implizite Verfahren zur Nutzerdatenerfassung optimal eingesetzt werden. Ferner bilden implizite und explizite Verfahren keine disjunkte Menge, d.h. beide Verfahren können durchaus hybrid eingesetzt werden.

Speicherung der Benutzerdaten

Je nach Anwendungsgebiet existieren unterschiedlich mächtige Repräsentationsformen für benutzerspezifische Interessen.

Interessenvektoren

Benutzerprofile werden generell in Vektoren von Schlüsselwörtern repräsentiert – auch als Schlüsselwortprofil bezeichnet. Jedes Schlüsselwort bildet ein Interesse des jeweiligen Benutzers ab. Diese können wiederum gewichtet sein um sie zu priorisieren. [2, 3, 17]

Taxonomien und Ontologien

Interessenvektoren besitzen einen Mangel an Repräsentationsfähigkeit für Zusammenhänge zwischen Synonymen und Homonymen. Semantische Beziehungen und demzufolge Ableitungen für weitere noch nicht definierter Interessen sind nicht darstellbar. Weitere Modellierungsarten des Benutzerprofils, insbesondere Taxonomien und Ontologien, bei denen semantische Beziehungen und Metainformationen verwendet werden, werden im Abschnitt 2.3 intensiver erläutert.

Personalisierung der Suche

Die Personalisierung der Suche bildet die 3. Phase des eingangs genannten Personalisierungsprozesses. In der ersten und zweiten Phase wurden Benutzerdaten erfasst und gespeichert. Das Nutzerprofil wird im jeden Fall verwendet um die eigentliche Personalisierung der Suche durchzuführen.

Inhaltsbasierte und kollaborative Suche

Die personalisierte Suche untergliedert Ansätze zur Integration von Interessensprofilen in

- Inhaltsbasierte (Content-based) und
- Kollaborative Ansätze.[3]

Bei der inhaltsbasierten Suche wird angenommen, dass jeder Nutzer eigenständig handelt. Die einzige Quelle der Personalisierung stellt das Interessensprofil des Nutzers dar. Dieses wird im Zuge der Personalisierung mit den Inhalten von Suchobjekten verglichen und über eine Methode bzw. einen Algorithmus die Relevanz des Objektes bestimmt.

Kollaborative Suchen basieren auf der Annahme dass Nutzer mit gleichen Interessen die gleichen Bedürfnisse haben. Eine Analyse der Historie über das Kaufverhalten einer Gruppe von Nutzern kann Rückschlüsse über das Verhalten auf den einzelnen Nutzer geben. Ein wesentlicher Vorteil dieses Ansatzes ist, dass kein Wissen über das Suchobjekt bestehen muss. Ein Nachteil stellt das sogenannte „Cold-Start“ Problem dar, welches vor allem bei neuen und kurzlebigen Suchobjekten auftaucht.[3]

Implementierung und Präsentation

Folgende Ansätze zur Implementierung der Personalisierung werden diskutiert. [8]

Re-Ranking Mittels Re-Ranking werden Suchergebnisse einer herkömmlichen Suche neu sortiert. Nachteil dieses Ansatzes ist die Abhängigkeit von der Qualität der herkömmlichen Suche. Ergebnisse aus der herkömmlichen Suche werden nicht entfernt.

Filtering Durch das Filtern von Suchergebnissen ist es möglich, nicht relevante Dokumente dem Nutzer vorzuenthalten, welches positiven Einfluss auf die Precision einer Suche haben kann.¹² Eine Studie brachte zum Ergebnis, dass Filtering in der Lage ist, einen Teil irrelevanter Dokumente zu filtern, jedoch wird auch ein nahezu gleichgroßer Anteil relevanter Dokumente irrtümlicherweise ausgeschlossen. [9]

Query Expansion Durch die Verwendung von Query Expansion ist es möglich, eine Anfrage einer Suche zu erweitern. Die Notwendigkeit ist unter anderem dadurch begründet, dass ein Großteil aller Suchanfragen kurz und mehrdeutig ist[10]. Es wird argumentiert, dass durch die Erweiterung einer kurzen Anfrage (Ein-Wort-Anfrage), der Suchraum eingeschränkt wird, Suchergebnisse werden nach der Erweiterung der Anfrage nicht mehr angezeigt[8].

2.3 Semantische Einflüsse

Bei klassischen Suchmethoden des Information Retrieval werden semantische Aspekte außen vorgelassen und Dokumente als Sequenz von Wörtern behandelt. Im Allgemeinen wird von einem Benutzer eines Suchdienstes eine geringe Anzahl an Suchwörtern eingegeben. Diese beschreiben die gesuchten Informationen mit einem Vokabular, das dem Nutzer geläufig ist und welches im Resultat erwartet wird. Hier jedoch besteht die Diskrepanz. In den durchsuchten Dokumenten kann ein anderes Vokabular benutzt werden, welches die gleiche Semantik beschreibt. Somit entgehen bei diesem klassischen Suchansatz, der Schlüsselwortsuche, dem Nutzer einige eventuell entscheidende Suchergebnisse. [12, 13]

Mittels Nutzung von Taxonomien oder Ontologien wird versucht, den Mangel an semantischer Ausdruckfähigkeit auszugleichen. Durch den Einsatz von Semantiken im E-Commerce wird versucht, dem Kunden eine gezielte Suche nach Produkten und Dienstleistungen in natürlicher Sprache zu ermöglichen.

Bezüglich der Auswahl an semantischen Technologien wird sich nur auf Taxonomien sowie Ontologien und die Ontologie-basierte Suche konzentriert, da diese im Zusammenhang mit Personalisierung in der untersuchten Literatur identifiziert werden konnten. Diese semantischen Mittel werden im ersten Schritt beschrieben. Im zweiten Schritt wird aufgezeigt, wie diese Mittel in Kombination mit Personalisierung

¹² Die Anzahl der gefundenen Dokumente nähert sich der Anzahl relevanter Dokumente, d.h. die Precision steigt.

auftreten. Im letzten Schritt werden Implikationen durch die Benutzung der beschriebenen semantischen Mittel bei gleichzeitiger Integration einer personalisierten Suche auf den E-Commerce Bereich erläutert und diskutiert.

Semantische Basiselemente

Taxonomien

Taxonomien sind in einem hierarchischen Konzept organisiert. Jeder Klasse wird eine Oberklasse zugeordnet, wodurch eine Baumstruktur entsteht. Somit werden Generalisierungen und Spezialisierungen zwischen den einzelnen Klassen ermöglicht[2].

Ontologien

Eine Instanz einer Ontologie ist ein formal semantisches Modell einer Anwendungsdomäne. Als eine Implikation der Formalisierung kann Mehrdeutigkeit vermieden werden. Die Repräsentation erfolgt über eine standardisierte Terminologie. Dazu gehören Beziehungen und gegebenenfalls Ableitungsregeln zwischen den dort definierten Begriffen. Das gemeinsam genutzte Vokabular ist in der Regel in Form einer Taxonomie gegeben, die als Ausgangselemente Klassen, Relationen, Funktionen und Axiome enthält. Ontologien an sich sind üblicherweise in Taxonomien, in Baumstrukturen mit mehrfacher Vererbung und disjunkten Unterkategorien organisiert. Kategorien, also Klassen einer Ontologie können mit anderen Kategorien über Relationen verknüpft oder mit Attributen detailliert beschrieben werden. Ontologien beschreiben für einen Wissensbereich nicht nur Kategorisierungen, sondern auch Regeln. Somit lassen sich über ontologische Regelmechanismen auch implizite Verknüpfungen folgern. Beispielhaft sei die Web Ontology Language (OWL) erwähnt - eine semantische Auszeichnungssprache zum Veröffentlichen und Austauschen von Ontologien. [12, 13]

Ontologie-basierte Suche

Im Folgenden wird die „Concept-Search“, eine Ontologie-basierte Suche als Beispiel einer konkreten Ausprägung einer semantischen Suche schemenhaft erläutert und mittels eines beispielhaften Ablaufs ergänzt.

Bei der „Concept Search“ wird über in einer Ontologie definierte sogenannte Konzepte versucht, die Problematik aufgrund von Synonymen und Homonymen zu lösen, bei der zu wenig oder zu viele relevante Suchergebnisse bestimmt werden. Konzepte repräsentieren abstrakte Themen, die über Beziehungen zu verschiedenen Unterkonzepten verbunden sind. Somit ist es möglich, über den Weg des Konzepts Synonyme zu Unterkonzepten zu bestimmen. [11]

Der Ergebnisraum wird spezifiziert indem Dokumente, die als Ergebnis der Suche auftreten dürfen mit einer Referenz-Ontologie assoziiert werden. Somit besitzt jedes Dokument einen Satz semantischer Annotationen, die die Semantik beschreiben, welche wiederum für die semantische Suche benötigt werden. Dieser Vorgang kann

sehr präzise manuell durchgeführt werden oder eben automatisiert. Hierbei werden Websites oder andere Ergebnisdokumente „gespidert“. (S.2)[9]

Bei der üblichen Ontologie-basierten Suche können eingegebene Begriffe in ihrer Anzahl automatisch vervielfacht werden. Dies erfolgt abhängig von der Modellierung und Komplexität des Suchthemas. Es erfolgt eine Suche nach logischen Kombinationen der Eingabebegriffe, sowie Unterbegriffen (Subkonzepte) und Synonyme. Somit wird die klassische schlüsselwortbasierte Suche über eine Inferenzmaschine und die zugrunde liegende Modellierung der Ontologie erweitert und verfeinert. Es erfolgt eine automatische Erweiterung der Suchphrase mit beispielsweise Synonymen, welches eine intuitivere Suche ermöglicht. [12]

Verwendete Referenz-Ontologien

In der untersuchten Literatur wurde beispielsweise die Magellan Ontologie verwendet. Diese besteht aus 4400 Knoten. Dabei ist die Semantik der Kanten nicht genauestens spezifiziert. In der Regel folgen sie dem Spezialisierungskonzept. Diese Ontologie ist zum jetzigen Zeitpunkt unter der originalen Website¹³ nicht mehr verfügbar. [9]

Häufig wird auch die im Open Directory Project (ODP) spezifizierte Taxonomie eingesetzt. ODP repräsentiert ein freies von Freiwilligen gepflegtes Webverzeichnis. Bei dieser Taxonomie besteht die Besonderheit darin, dass jeder Knoten mehrere Eltern- und Kindknoten haben darf. [14]

Hybride Verwendung von Personalisierter Suche und semantischen Mitteln

In diesem Kapitel wird aufgezeigt, wie semantische Aspekte Teil der Personalisierung sein können. Dazu wird erläutert, wie das Potenzial der semantischen Suche durch Personalisierung erweitert werden kann sowie der Umkehrschluss wie die semantische Suche oder Teilverfahren respektive Bestandteile in die Verfahren der personalisierten Suche unterstützend eingreifen können. Es wird dargelegt, welche Möglichkeiten aktuelle Studien untersuchen.

Zur Beantwortung der Fragen konnten folgende Kombinationen definiert werden:

1. Für eine personalisierte Suche wird eine Ontologie für die Benutzermodellierung verwendet.
2. Eine Ontologie-basierte Suche wird mit einer Personalisierung erweitert.

Bezüglich der zweiten Kombination zeigen im Allgemeinen die untersuchten Quellen, dass die Ontologie-basierte Suche auf zwei Arten durch Personalisierung erweitert wird. Im gängigen Fall fungiert die Personalisierung als Wrapper für einen klassischen Suchdienst. Das heißt die Personalisierung setzt auf den Ergebnissen der semantischen Suche auf [vgl. Abbildung 1 (a)].

In anderen Fällen wird eine Art "Verzahnung" oder "Verflechtung" von semantischer Suche und Personalisierung angestrebt [vgl. Abbildung 1 (b)]. In diesem Fall greift die Personalisierung innerhalb des Vorgangs der semantischen Suche ein. Wie

¹³ <http://magellan.mckinley.com>

bereits angedeutet, kann in beiden Fällen und wird in den untersuchten Artikeln für das Benutzerprofil entweder der klassische Interessenvektor oder eine Ontologie verwendet. Im Folgenden werden Beispiele für beide Herangehensweisen erläutert.

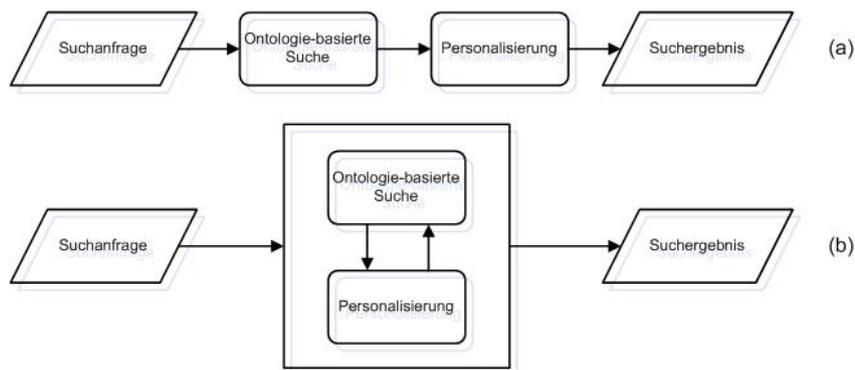


Fig. 1. Möglichkeiten der Erweiterung einer Ontologie-basierten Suche mittels Personalisierung

Das Benutzerprofil mit semantischen Erweiterungen

Das Mengen- respektive Vektor-Modell stellt die einfachste Form der Modellierung dar. Jedoch wird dieses Konzept ohne eine interne Struktur charakterisiert, welches keine Verbindungen zwischen den Konzepten (Kategorien) aufweist. Somit lassen sich keine Aussagen über neue Merkmale auf Basis schon modellierter Merkmale treffen.[18] Weiterhin besteht bei dieser Darstellungsart der Modellierung der Benutzerinteressen auch hier der Konflikt in des unzureichenden Ausdrucks aufgrund Synonymen, also Wörter, die die selbe Bedeutung jedoch eine unterschiedliche Schreibweise besitzen, und Homonymen, Ausdrücke mit gleichen Schreibweisen, jedoch anderer Bedeutung. Beispielsweise könnte ein Interessenvektor die Kategorie 'Auto' gewichten, aber durch die fehlende Beziehung zu 'Kraftfahrzeug' wäre eine derartige Modellierung unzureichend und Dokumente, die sich auf ‚Kraftfahrzeug‘ anstelle von ‚Auto‘ beziehen, würden womöglich aus der Ergebnisliste ausgeschlossen, beziehungsweise weniger priorisiert werden. [14]

Eine Erweiterung und somit ein Versuch der Beseitigung der genannten Einschränkungen existiert in der Modellierung der Benutzerinteressen in Taxonomien [vgl. Abbildung 2 (a)] oder Ontologien [vgl. Abbildung 2 (b)] [18]. Hierbei werden besuchte Knoten gespeichert und gegebenenfalls gewichtet, beispielsweise über die Anzahl der Besuche (Klicks) von Dokumenten, die mit jener Kategorie beschrieben werden. Somit gehört jedes Interessenthema einer Kategorie der Ontologie an [9]. Es existiert eine personenbezogene Zuordnung bei der ein Kontext (Suchquery) zu einem

Knoten in jener Ontologie assoziiert werden kann, wobei jeder Kontext auch mehrere Zuordnungen (Knoten) in dieser Ontologie besitzen darf. [14]

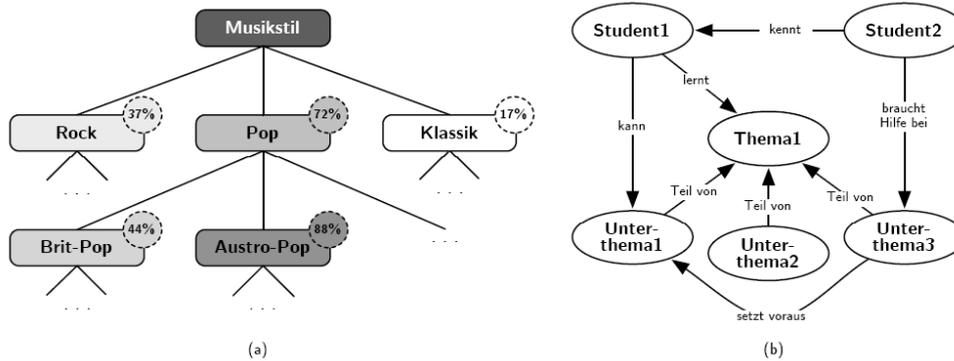


Fig. 2. Benutzerprofil modelliert als Taxonomie (a) und Ontologie (b) (nach (S.17)[18])

[2] bezeichnet diese Art des Benutzerprofils als „Concept Profile“. Wobei hier verschiedene Abstufungen beschrieben werden. So werden auch Benutzerprofile erläutert, bei denen nur die abstrakten generellen Konzepte als Interesse gespeichert werden.

Durch die Benutzung von Ontologien können somit über die Beziehungen ähnliche Interessen oder auch Desinteressen beispielsweise über Generalisierung und Spezialisierung genauer bestimmt werden.

Verflechtung von Ontologie-basierter und personalisierter Suche

Im Folgenden wird eine Möglichkeit der Verflechtung Ontologie-basierter und personalisierter Suche beschrieben, bei der eine Ontologie für die Benutzermodellierung zugrunde liegt. Das Benutzerprofil wird implizit und dynamisch adaptiert.

Angelehnt an [14] wird allgemein, so bald ein neuer Kontext besteht, das heißt ein Such-Query von einem Benutzer eingegeben wurde, versucht diesen Kontext mit einem oder mehreren Knoten des Benutzerprofils, der benutzerbezogenen Ontologie zu assoziieren. Wenn der Such-Query anhand von Knoten des Benutzerprofils charakterisiert werden kann, werden die mit diesem Knoten assoziierten Dokumente in die Ergebnisliste übernommen und die Knoten stärker gewichtet. Im anderen Ausgangspunkt, dem Fall das kein Knoten direkt mit dem Such-Query assoziiert werden kann, wird überprüft ob ein anderer Knoten, der sich noch nicht im Benutzerprofil befindet in einer nahen Beziehung steht. Dementsprechend wird je nach Algorithmus ausgehend von Knoten der Ontologie des Benutzerprofils eine Ebene weiter, beispielsweise deren Eltern- oder Kindknoten untersucht. Sollten diese mit dem Suchquery in Verbindung gebracht werden können, werden diese Knoten stärker gewichtet, die Interes-

sen werden spezifiziert sowie die assoziierten Dokumente in das Ergebnis übernommen beziehungsweise stärker gewichtet.

Das folgende Beispiel wird anhand der abgebildeten Ontologie beschrieben.

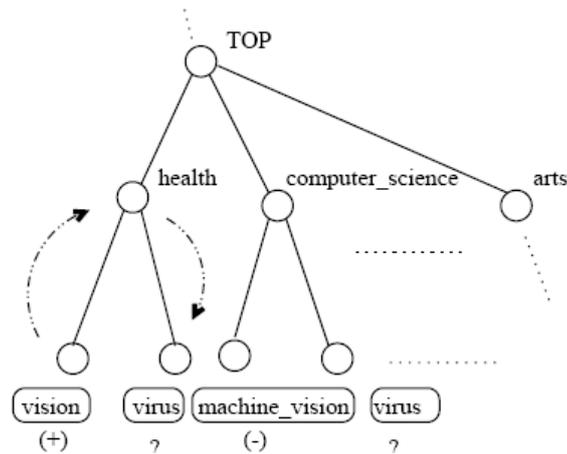


Fig. 3. Beispielhafter Veränderungsvorgang einer Ontologie für das Benutzerprofil [14]

Im ersten Suchvorgang sucht der Benutzer nach "vision" und erhält aufgrund seines noch anfänglichen, nicht bestimmten Benutzerprofils eine komplette Ergebnisliste, die sowohl Dokumente aus dem Bereich "health" als auch aus dem Bereich "computer_science" enthält. Von den Ergebnissen wählt er nur jene aus dem Bereich "health" aus. Infolgedessen wird in der Ontologie des Benutzerprofils der Knoten "vision" als Kindknoten von "health" stärker gewichtet. In einem nächsten Suchvorgang sucht der Nutzer nach "virus". In diesem Falle existieren bereits Knoten im Benutzerprofil, das heißt Gewichtungen nach den Interessen des Benutzers. Dementsprechend wird versucht den Suchquery einem gewichteten Knoten zu assoziieren. Da nur "vision" gewichtet ist, kann dies nicht durchgeführt werden. Somit wird nun ausgehend vom gewichteten Knoten "vision" über dessen Elternknoten "health" nach weiteren Eltern- oder Kindknoten gesucht, die den Suchquery beschreiben. Als Ergebnis wird der Kindknoten "virus" unter "health" bestimmt, dessen assoziierten Dokumente in die Ergebnisliste übernommen bzw. stärker gewichtet. Außerdem wird dieser Knoten im Benutzerprofil eine stärkere Gewichtung zugewiesen. Als Essenz lässt sich zusammenfassen, dass somit im zweiten Vorgang Dokumente aus dem Bereich "health" und nicht etwa "computer_science" aufgrund des ersten Vorgangs und damit bestimmter Interessenkategorien stärker im Ergebnis repräsentiert werden.

Personalisierung als Wrapper um eine Ontologie-basierte Suche

Das Wort „Wrapper“ bezeichnet aus dem Englischen und verstärkt geläufig im technischen Raum eine Hülle oder einen Adapter. Dementsprechend wird in diesem Kon-

text die Personalisierung "aufgesetzt". Das heißt in den meisten Quellen werden Open World Suchdienste wie beispielsweise Google erweitert, indem ein Suchergebnis über Re-Ranking oder Filtering personalisiert werden. Es werden nur die Fälle betrachtet, die entweder eine Ontologie als Benutzerprofil benutzen oder eine zugrundeliegende Ontologie-basierte Suche mittels Personalisierung erweitern.

[9] benutzt eine wie im Abschnitt 2.3 erläuterte Ontologie-basierte Suche. Es wird mittels verschiedener extensiver Personalisierungsverfahren geprüft, wie sich das Suchverfahren optimieren lässt. Dabei wird ein hierarchisches automatisch bestimmtes sowie dynamisches Benutzerprofil verwendet. Hierarchisch entspricht dabei einer Ontologie mit 4400 Knoten im Kontrast zu einer einfachen Liste an Interessen. Diese wird automatisch generiert indem dem Benutzer "über die Schulter geguckt wird", also anstelle des expliziten Angebens von Interessen ein implizites Verfahren benutzt wird. Dafür werden besuchte Webseiten des Benutzers analysiert. Dieses Profil wird dynamisch angepasst. Somit wird auf Veränderungen in den Interessen beispielsweise den Short Term Interests, also den kurzfristigen Interessen, eingegangen. Wobei als Interessenkriterium die Annahme repräsentiert, dass je öfter eine Kategorie ausgewählt wird desto stärker das Interesse darin sein muss. Als Personalisierungsverfahren werden Re-Ranking und Filterung, also Neuplatzierung der relevanteren Themen und Ausblendung der nicht relevanten eingesetzt. Als Ergebnis wird konstatiert, dass die Verbesserung gegenüber der zugrundeliegenden semantischen Suche moderat, aber wahrnehmbar ist. Sie variiert aufgrund der gewählten Funktion zur Bestimmung der Interessen sowie der Personalisierungsmethode (Re-Ranking, Filtering) an sich. Gleichzeitig wird betont, dass eine automatische Generierung von großen hierarchischen Benutzerprofilen möglich ist.

Erkenntnisse und Implikationen für den Einsatz im E-Commerce

Generell lässt sich konstatieren, dass sich die untersuchte Literatur auf den Open World Approach bezieht. Weitestgehend werden vorhandene Suchmaschinen personalisiert. Ein Großteil der Probleme, die beispielsweise in [9] erwähnt werden, entsteht gerade durch diesen Open World Ansatz. Beispiele dafür sind die Aktualität der Dokumente in der Datenbank, ein zu starkes Anwachsen der Anzahl der Dokument oder Anderssprachigkeit der Ergebnisse. All diese Probleme dürften im E-Commerce in der Regel nicht auftreten oder zumindest aufgrund der Abgeschlossenheit des Systems stärker kontrollierbar sein. Im Folgenden wird versucht untersuchte semantische Mittel, die in Kombination mit Personalisierung auftreten, differenziert für den E-Commerce-Bereich zu untersuchen.

Differenzierung der Kombination von Ontologie-basierter Suche und Personalisierung

Im ersten Schritt lässt sich keine Aussage treffen, ob die Kombination von Ontologie-basierter und personalisierter Suche abhängig von einer Eingangsgröße zu einer differenzierten Empfehlung für den Einsatz jener führt. Die untersuchte Literatur lieferte

keine generelle Aussage, wann beispielsweise eine Ontologie-basierte Suche die Personalisierung eher einschränkt respektive negativ beeinflusst oder sie eher fördert. Anhand der erläuterten Untersuchungen lässt sich keine Differenzierung ableiten. Mögliche Ursachen dafür sind eventuell, dass die Auswirkungen einer Ontologie-basierten Suche noch schwierig messbar zu machend sind. Die untersuchte Literatur zeigt auf, dass es schon allein sehr schwierig ist die Auswirkungen der bloßen Personalisierung für verschiedene Szenarien zu bestimmen.

Differenzierung der Kombination einer Personalisierten Suche zuzüglich eines als Ontologie modellierten Benutzerprofils

Somit wird sich im Folgenden nur auf Empfehlungen für die Integration von Ontologien für die Modellierung von Kundeninteressen konzentriert. Wie im Abschnitt 2.3 erläutert, lassen sich Referenz-Ontologien dahingehend einschränken, dass Kundeninteressen gewichtet gespeichert werden. Verglichen mit einem Interessenvektor ist dieser Vorgang sehr spezifisch und genau. Dementsprechend sollten jene Art (Taxonomien und Ontologien) Anwendung in Domänen finden, in denen die Priorität auf eine sehr genaue Modellierung des Kunden liegt. Dies können beispielsweise kritische Domänen sein. Angenommen es existiert ein großer heterogener Ergebnisraum respektive eine breite Produktdatenbank, dann sind Ontologien in ihrem Einsatz empfehlenswert, da eine hohe Wahrscheinlichkeit beispielsweise an Synonymen über verschiedene Produktkategorien besteht.

3 E-Commerce Szenarien und Empfehlungen für den Einsatz von Personalisierung

3.1 Überblick

„ [...] personalization has different effectiveness on different queries, users, and search contexts.” [10]

Zu diesem zitierten Resultat gelangen einige untersuchte Quellen [10, 15]. Wie in Abschnitt 2.3 beschrieben, beziehen sich diese in der Regel auf den Open World Approach. Für den E-Commerce Bereich abgeleitet, kann das heißen, dass auch hier verschiedene Szenarien zu einer unterschiedlichen Sinnhaftigkeit und Effektivität bis hin zu einer Nicht-Empfehlung der Nutzung einer personalisierten Suche führen. Beispielsweise ist personalisierte Suche nicht für jeden Kundentyp geeignet oder jede Art von Produkt bzw. deren Beschreibung.

Somit werden in diesem Kapitel verschiedene E-Commerce Szenarien aufgebaut und adaptiert sowie eine mögliche Empfehlung diskutiert. Adaptieren lassen sich dafür beispielsweise Artikel, die sich explizit auf den Open World Approach beziehen. So finden bei diesem Ansatz für die Modellierung der Benutzerinteressen eher allgemeine Interessen Verwendung. Für den E-Commerce Bereich lassen sich weitere Erkenntnisse wie das Kaufverhalten als Interesse eines Kunden spezifizieren. Die

untersuchte Literatur diskutiert bezüglich einer personalisierten Suche nicht E-Commerce typische Charakterisierungen wie beispielsweise den Kundentyp. Dafür existieren verwandte Untersuchungen, die sich auf andere Arten der Personalisierung beispielsweise Recommender-Systeme beziehen. Diese lassen sich für den Bereich der Personalisierten Suche adaptieren. Somit besteht das Ziel dieses Kapitels darin, auf Basis gemachter Erkenntnisse und Diskussionen aus benachbarten Bereichen Dimensionen zu definieren um personalisierte Suchsysteme zu charakterisieren und zu organisieren.

Zur Erstellung von Szenarien werden die Bereiche

- Kundenbeziehung
- Domäne

untersucht.

Dabei wird sich an das allgemeine Schema der Personalisierung, wie in Abschnitt 2.2 beschrieben, orientiert. Somit wird abhängig von differenzierbaren Einflussfaktoren eine Empfehlung oder Nicht-Empfehlung für die Teilprozesse

- Erfassen der Nutzerdaten,
- Erstellen eines Nutzerprofils und
- Personalisierung der Suchergebnisse

gegeben. Zusätzlich sei erwähnt, dass in bestimmten Szenarien keine Empfehlung für einen Teilprozess oder gar sämtliche Prozesse der Personalisierung getroffen wird. Dies begründet sich darin, dass sich der untersuchte Einflussfaktor nicht oder nur teilweise differenzierend auf die Personalisierung auswirkt. Somit kann die Wahl des jeweiligen Teilprozesses einer Personalisierung von anderen Einflussfaktoren abhängig gemacht oder frei gewählt werden.

3.2 Domäne

Die Domäne wird nach [20] charakterisiert und adaptiert:

- Beschaffenheit des Ergebnisraums,
- Lebensdauer von Ergebnissen,
- Einfluss der Userinteressen,
- Criticality,
- Kundentyp.

Erweitert wird diese Definition um den Typ des angebotenen Guts, dem Produkttyp.

Table 3. Domäne

	Ausprägung	Erfassung	Speicherung	Personalisierung
Kudentyp	Spezialisierte Kunden	explizit	Ontologie	content-based; kein Expansion
	Nicht spezialisierte Kunden	-	-	-
Produkttyp	Erfahrungsgüter	-	Vektor	grundsätzlich nicht empfohlen; kein Filtering; kollaborativ
	Suchgüter	-	-	-
Lebensdauer von Produkten	kurz	-	-	nicht kollaborativ
	lang	-	-	-
Criticality	kritische Güter	explizit	Ontologie	Prüfen, ob Personalisierung eingesetzt werden soll; kein Filtering
	unkritische Güter	implizit, explizit	-	-
Dauer von Kundeninteressen	langfristig	explizit	-	-
	kurzfristig	implizit, explizit (Feedback)	-	-
Suchraum	homogen	explizit	-	-
	heterogen	implizit	Ontologie	-

Kudentyp

Bei der Differenzierung eines Kundentyps lassen sich

- spezialisierte Kunden (Domänenexperten) und
- nicht spezialisierte Kunden

unterscheiden[20]. Domänenexperten sind Kunden, die den jeweiligen E-Commerce Shop entweder während der Ausübung ihres Berufs nutzen oder mit einem breiten Hintergrundwissen Produkte suchen und kaufen. Eigenschaften des Domänenexperten sind:

- konkrete und sehr genaue Suchanfragen,
- Bedarf an akkuraten Suchergebnissen,
- Vermeidung von Unschärfe und nicht-relevanten Ergebnissen (Noise-free).

Nicht spezialisierte Kunden zeichnen sich in der Regel vor allem durch

- unspezifische Suchanfragen und
- eine weniger exakte Vorstellung über das Suchobjekt

aus.

Tabelle 3 liefert einen Überblick empfohlener Suchansätze. Experten wird aufgrund der Spezialisierung ein weitgehend homogener Suchraum unterstellt (Kunde ist Experte für Fotografie bzw. Kameras). Explizite Erfassungen von Nutzerinteressen finden im homogenen Raum eine höhere Genauigkeit, wogegen eine implizite Erfassung mit Ungenauigkeiten (Noise) verbunden sein kann. Ontologien können durch ihre Fähigkeit des Erfassens komplexer Sachverhalte eingesetzt werden um ein umfangreicheres und genaueres Nutzerprofil zu erstellen. Durch die unterstellte Genauigkeit der Suchanfragen, wird eine Erweiterung der Suchanfrage durch den Suchalgorithmus nicht empfohlen. Dem Domänenexperten wird es auch weniger interessieren, was die Mehrheit der Kunden denkt. Das schließt beim Personalisierungstyp wiederum den kollaborativen Ansatz aus.

Die Erfassung von Benutzerdaten für nicht spezialisierte Kunden kann einerseits explizit oder andererseits implizit erfolgen. Eine Empfehlung kann ohne die Berücksichtigung weiterer Dimensionen der Domäne (z.B. Beschaffenheit des Ergebnisraums) oder weiterer Bereiche (z.B. Art der Kundenbeziehung) nicht erfolgen. Aufgrund der unexakten Vorstellung über Suchobjekt wird an dieser Stelle der kollaborative Ansatz gewählt. Kunden mit gleichen Interessensprofilen wird unterstellt, dass gleiche Kaufinteressen vorhanden sind.

Produkttyp

Produkte respektive angebotene Güter können in

- Suchgüter (Search Goods) und
- Erfahrungsgüter (Experience Goods)

unterteilt werden[22].

Bei Suchgütern sind dem potentiellen Käufer Eigenschaften wie Farbe und Form, sowie die Qualität des Produktes bekannt. Die Kaufentscheidung wird auf Basis eines Vergleichs der gewünschten Eigenschaften mit den Produktspezifikationen, beispielsweise eines Fernsehers oder eines Fahrrads, getroffen. Im Umgang mit Suchgütern kann eine personalisierte Suche uneingeschränkt empfohlen werden.

Bei Erfahrungsgütern hängt die Kaufentscheidung von subjektiven Merkmalen des Produktes, wie Geruch oder Geschmack, ab. Somit kann eine Einschätzung, wenn vorher keine Erfahrung vorhanden war, erst nach dem Kauf gemacht werden. Aufgrund dieser subjektiven Meinung eines Erfahrungsguts ist es im Gegensatz zu den Suchgütern sehr schwierig zu beschreiben also auch schwierig in einer Taxonomie oder Ontologie in Beziehung zu anderen Eigenschaften zu setzen. Aufgrund dieser schwierigen Bestimmung besteht ein potentielles Risiko beim Filtering etwas „Richtiges“ auszublenden, deswegen wird dieser Algorithmus explizit aus der Empfehlung ausgenommen. Der Einsatz von personalisierter Suche kann aufgrund fehlender Kate-

gorisierung des Produktes nicht vorgenommen werden. Eine Ausnahme besteht im kollaborativen Ansatz, da dieser nicht auf der Produktbeschreibung aufbaut.

In jedem Fall besteht ein gewisser Grad an Risiko bei Erfahrungsgütern zu personalisieren respektive sie zu empfehlen durch beispielsweise eine höhere Platzierung in den Suchergebnissen. Wenn ein Kunde diese Empfehlung annimmt, hat er ein hohes Vertrauen in den Anbieter. Dieses Vertrauen kann verloren gehen, dadurch, dass die subjektive Wahrnehmung des Kunden das empfohlene Produkt nicht gefällt.

Lebensdauer von Produkten

Bei der Lebensdauer von Produkten wird im Folgenden zwischen

- kurzlebigen Produkten und
- langlebigen Produkten

unterschieden.

Anhand der Lebensdauer von Produkten kann eine Differenzierung in der Umsetzung der Personalisierung gemacht werden. Kurze Lebenszyklen schließen eine kollaborative Meinungs- und Interessenbildung über ein Produkt nahezu aus. Der kollaborative Ansatz zum Abgleich der Nutzerinteressen mit den Interessen anderer Nutzer ist bei diesen Produkten nicht zielführend, weil eine Datenbasis nicht bzw. nur über einen kurzen Zeitraum vorhanden ist.[22] Neu eingeführte Produkte stehen vor den gleiche Herausforderung.

Criticality

Die Criticality erfasst die Kosten einer fehlerhaften Personalisierung beispielsweise durch den Ausschluss relevanter Items aus dem Suchergebnis. E-Commerce Anbieter bieten möglicherweise Produkte an, deren Ausschluss bzw. Nichtauffindbarkeit als kritisch einzustufen wäre. Exemplarisch soll das Beispiel eines Anbieters für Medikamente dienen. Wenn dem Nutzer anhand seines Nutzerprofils, in dem Daten über seinen Gesundheitszustand erfasst sind, wichtige Medikamente vorenthalten werden, wird dies als kritisch eingestuft.

Folgendermaßen werden die Güter in

- Kritische Güter und
- Unkritische Güter

kategorisiert.

Das Ziel der Personalisierung für kritische Güter muss das Ausschließen von Unsicherheiten beinhalten. Durch die Angabe von expliziten Nutzerinteressen über z.B. Fragebögen können diese Unsicherheiten weitgehend ausgeschlossen werden. Das Ausschließen von Suchergebnissen durch Filtering (gemäß 2.2) wird nicht empfohlen. Das Risiko falsch eingestufte Objekte dem Nutzer vorzuenthalten, sollte vermieden werden. Das Speichern innerhalb einer Ontologie (vgl. Ab-

schnitt 2.3) ermöglicht es zusätzlich, die Genauigkeit des Interessensprofils zu erhöhen.

Grundsätzlich sollte nur dann eine Personalisierung eingesetzt werden, wenn sichergestellt werden kann, dass Unsicherheiten im ausreichenden Maße ausgeschlossen werden können.

Die Verwendung der personalisierten Suche bei unkritischen Gütern kann uneingeschränkt empfohlen werden. Die niedrige Relevanz korrekt personalisierter Ergebnisse verursacht minimale Kosten bei einer Fehleinschätzung durch die Personalisierung.

Dauer von Kundeninteressen

Exemplarisch soll folgendes Beispiel Veranschaulichung dienen. Ein Nutzer, der vor dem Kauf einer Kamera steht, besitzt kurzfristig ein Interesse. Nach dem Kauf ist dieses Interesse nicht mehr vorhanden, da davon auszugehen ist, dass der Nutzer sich in naher Zukunft keine weitere Kamera kaufen wird. Ist der Nutzer aber Fotograf, kann es durchaus möglich sein, dass Interessen langfristiger Natur sind. Wie an diesem Beispiel zu erkennen ist, müssen zwischen kurzfristigen und die langfristigen Interessen unterschieden werden. Eine Differenzierung kann sich als schwierig erweisen.

Kurzfristige Interessen können explizit schwer erfasst werden, da die Angabe von kurzfristigen Interessen nicht mit den "eigentlichen" Interessen eines Nutzers gleichzusetzen sind. Eine denkbare Umsetzung einer explizierten Datenerfassung ist das Feedback-Verfahren. Eine implizite Erfassung von Nutzerdaten stellt ebenfalls eine Herausforderung für kurzlebige Interessen dar. Die implizite Erfassung von Nutzerdaten minimiert jedoch den Aufwand der Erfassung (für den Nutzer) und stellt demnach eine präferierte Methode zur Erfassung von Nutzerdaten dar.

Langfristige Interessen eines Nutzers können explizit über Fragebögen erfasst werden. Im Gegensatz zur impliziten Erfassung können die Interessen mit einem höheren Wahrheitsgehalt erfasst werden. Unschärfen im Interessensprofil, welche durch ein implizites Erfassen von kurzfristigen Interessen entstehen könnten, können durch eine Beschränkung auf explizite Erfassung von Daten vermieden werden.

Beschaffung des Ergebnisraums

Der Ergebnisraum stellt im Rahmen dieser Arbeit gleichzeitig die Menge aller angebotenen Güter dar. Ein heterogener Suchraum ist geprägt durch eine Vielzahl von Güterkategorien. Ein Beispiel stellt der Online Shop Amazon dar. Eine spezialisierte E-Commerce Umgebung (Web-Shop für Fotografiebedarf) verfügt in der Regel über einen homogenen Suchraum. Wie in 2.2 beschrieben, wird die explizite Erfassung von Benutzerdaten durch die Komplexität des Suchraums begrenzt. Im stark heterogenen Raum wird diese Form der Erfassung nicht empfohlen. Es wurde zudem herausgestellt, dass die implizite Erfassung von Nutzerdaten im heterogenen Suchraum durch ihren Automatismus effektiv eingesetzt werden kann. Im homogenen Raum, sollte eine explizite Erfassung von Nutzerdaten präferiert werden, da dieser ein höherer Wahrheitsgehalt unterstellt wird. Die Untersuchung von Ontologien brachte zum Ergebnis, dass der Einsatz im heterogenen Suchraum eine höhere Effektivität vorweist.

3.3 Kundenbeziehung

Die Kategorie „Kundenbeziehung“ lässt sich in

- Dauer und Tiefe der Kundenbeziehung und
- Kaufverhalten des Kunden

untergliedern.

Table 4. Kundenbeziehung

	Ausprägung	Erfassung	Speicherung	Personalisierung
Dauer und Tiefe	kurzfristig, oberflächlich	implizit	-	-
	langfristig, tiefgründig	implizit und explizit	-	-
Kaufverhalten	z.B. Produkt-loyal, Marken-loyal, Kategorie-loyal	implizit	gewichtete Taxonomie	content-based; Re-Ranking, Filtering

Dauer und Tiefe der Kundenbeziehung

Die Dauer und Tiefe der Kundenbeziehung kann in E-Commerce Systemen variieren. Sie lässt sich charakterisieren in

- kurzfristig und oberflächlich (transaktionsorientiert) oder
- langfristig und tiefgründig.

Transaktionsorientierte Systeme setzen dabei auf eine eher kurzfristige, oberflächliche Beziehung. Hierbei kommen teilweise sogar anonyme Abwicklungen zu Stande. Somit lässt sich auch keine tiefgründige Historie zu einem spezifischen Kunden aufbauen. Dementsprechend empfiehlt es sich die aktuelle Session implizit zu untersuchen und darauf basierend zu personalisieren.

Im Kontrast zu der kurzfristigen Beziehung steht eine langfristige, tiefgründige. Hierbei kann auf ein langfristiges Kunden-Profil zugegriffen werden, welches über implizite oder explizite sowie kombinatv vervollständigt werden kann. [20]

Kaufverhalten

Ein speziell hervorzuhebendes Kaufverhalten kann nach [21] in

- Produkt-loyal,
- Marken-loyal,
- Hybride respektive Kombinationen aus beidem

klassifiziert werden.

Zusätzlich denkbar wäre die Bestimmung einer Art Klassifikations-Loyalität indem überprüft wird, ob der Kunde sich an ein allgemeines Schema anpasst, indem er bei-

spielsweise immer wieder Produkte kauft, die verstärkt von anderen Kunden gekauft werden.

Das Verhalten des Kunden kann implizit analysiert werden, indem eine Produkt-Taxonomie mit Gewichten beispielsweise die Anzahl der Käufe eines spezifischen Produkts assoziiert werden. Anhand dessen kann eine Suche dann in der Hinsicht personalisiert werden, dass von dem Kunden bevorzugte Produkte sowie Marken eine höhere Priorisierung im Suchergebnis finden. Dabei könnten jene Interessen beispielsweise höher gerankt werden oder nicht interessante Produkt oder Marken gefiltert werden.

4 Schlussbetrachtung

Im vorliegenden Paper wurden Ansätze der personalisierten Suche hinsichtlich der Einsetzbarkeit im E-Commerce untersucht. Technische als auch konzeptionelle Umsetzungsvarianten der personalisierten Suche konnten kategorisiert werden. Als Grundlage zur Kategorisierung der Suchmechanismen dienten Ansätze der personalisierten Suche aus gegenwärtigen Studien, als auch Techniken, die bereits vielfältig im Bereich Internet eingesetzt werden. Durch eine Adaption dieser auf das E-Commerce wurde die Anwendbarkeit diskutiert und Empfehlungen gegeben.

Des Weiteren wurden Wege identifiziert, wie semantische Mittel Einfluss auf eine personalisierte Suche nehmen können. Einerseits existieren Beispiele, bei der die "Concept-Search" als Repräsentant einer Ontologie-basierten Suche mit einer Personalisierung erweitert wird. Andererseits werden Ontologien für die Modellierung von Interessenprofilen einer personalisierten Suche verwendet. Es konnten differenzierte Empfehlungen für den Einsatz Letzteres im E-Commerce herausgestellt werden.

Schließlich wurden mögliche Szenarien für E-Commerce Anwendungen definiert. Zur Erstellung dieser Szenarien dienten Forschungen aus weiteren Gebieten des E-Commerce vor allem aus dem Bereich der „Recommender Systems“. Diese Szenarien konnten weitgehend an den untersuchten Kontext adaptiert werden. Anhand der durchgeführten Untersuchungen bezüglich der personalisierten Suche und den Einsatzmöglichkeiten von Ontologien, wurden Empfehlungen für Szenarien gegeben.

Gemachte Untersuchungen basieren auf Forschungen über personalisierte Suche und aus anliegenden Themenbereichen. Empfehlungen und durchgeführte Adaptionen wurden deduktiv vom gegenwärtigen Forschungsstand abgeleitet. Die vorliegenden Untersuchungsergebnisse wurden nicht in Experimenten überprüft.

Quellen

1. Brusilovsky, Millan: User Models for Adaptive Hypermedia and Adaptive Educational Systems, pp. 2-53. (2007)
2. Gauch, Speretta, Chandramouli, Micarelli: User Profiles for Personalized Information Access, pp. 55-89. (2007)
3. Micarelli: Personalized Search on the World Wide Web (2007)
4. Quiroga, Mostafa: Empirical evaluation of explicit versus implicit acquisition of user profiles in information filtering systems, In: D. H. Kraft (Ed.), Proceedings of the 63rd annual meeting of the American Society for Information Science and Technology, Medford, NJ: Information Today 37, pp. 4-13. (2000)
5. White, Jose, Ruthven: Comparing explicit and implicit feedback techniques for Web retrieval, pp.1-5. (2001)
6. Spiekermann S., Grossklags J., Berendt B: E-Privacy in 2nd Generation E-Commerce: Privacy Preferences versus actual Behavior (2005)
7. Sieg, Mobasher, Burke: Web Search Personalization with Ontological User Profiles (2007)
8. Keenoy, Levene: Personalisation of Web Search (2005)
9. Pretschner, Gauch: Ontology Based Personalized Search (1999)
10. Dou, Z., R Song, R., Wen, J.-R.: A Large-scale Evaluation and Analysis of Personalized Search Strategies (2007)
11. Kiryakov, Atanas K., Simov, Kiril Iv.: Ontologically Supported Semantic Matching (1999)
12. Studer, R., Schnurr, H., Nierlich, A. - Ontoprise Whitepaper Series, Karlsruhe, Semantisches Knowledge Retrieval (2001)
13. Otto, M.: Ontologien zur semantischen Suche in einem Bestand von Dokumenten (2008)
14. Tanudjaja, F., Mui, L.: Persona: A Contextualized and Personalized Web Search (2002)
15. Jeh, G., Widom, J.: Scaling Personalized Web Search (2003)
16. Challam, V., Gauch, S., Chandramouli, A.: Contextual Search Using Ontology-Based User Profiles (2007)
17. Liu, F., Yu, C., Meng, W.: Personalized Web Search by Mapping user Queries to Categories (2002)
18. Florian König: Adaptive Suche mit thematischen Ontologien (2008)
19. W Kim, D Choi, J Kim, B Jin: Development of a meta product search engine with web services (2005)
20. Ramezani, M., Bergman, L., Thompson, R.: Selecting and Applying Recommendation Technology (2008)
21. Cho, Y. H., Kim, J. K., im, S. H.: A personalized recommendation system based on product taxonomy for one-to-one marketing online (2002)
22. Nelson, P.: Information and Consumer Behavior, In: Journal of Political Economy, University of Chicago Press, vol. 78(2), pp. 311-29, March-Apr. (1970)
23. Forsyth, J., McGuire, T., Lavoie, J.: All Visitors are not created equal (2000)
24. Wirtz, B Electronic Business (2001)