

# **Johannes Kepler Universität Linz**

Institut für Informationsverarbeitung und Mikroprozessortechnik (FIM)

Dipl.-Ing. Michael Sonntag

## **Untersuchungen zur Personalisierung**

## **Kontaktinformationen:**

Dipl.-Ing. Michael Sonntag

E-Mail:

sonntag@fim.uni-linz.ac.at

Adresse:

Universität Linz

Institut für Informationsverarbeitung und Mikroprozessortechnik (FIM)

Altenbergerstr. 69

A - 4040 Linz

Fax:

++43-732-2468-599

Telefon:

++43-732-2468-437

# Inhaltsverzeichnis

<i>Inhaltsverzeichnis</i>	3
<i>Link-Verzeichnis</i>	5
<b>1 Auswahl von personalisierten Websites</b>	<b>5</b>
<b>2 Produkte zur Personalisierung von Websites</b>	<b>6</b>
<b>3 Verschiedenes</b>	<b>7</b>
<i>Untersuchungen zur Personalisierung</i>	<b>9</b>
<b>1 Über die Personalisierung</b>	<b>9</b>
1.1 Allgemeines	9
1.1.1 Definition	9
1.1.2 Probleme für Anbieter einer personalisierten Website	10
1.2 Sinn und Zweck der Personalisierung	13
1.2.1 Zielgerichtete Werbung	13
1.2.2 Erhöhte Kundenbindung bei Electronic Commerce	13
1.2.3 Speziell abgestimmte Informationen	14
1.2.4 Reduktion der Datenflut	14
1.2.5 Verteilung von Informationen in Betrieben	14
1.2.6 Genaue Benutzerüberwachung möglich	15
1.3 Wiedererkennen des Benutzers	15
1.3.1 Benutzername und Paßwort	16
1.3.2 Cookies	17
1.3.3 Spezielle URLs mit codierter Benutzerkennung	18
1.3.4 Position des Benutzers	18
1.3.5 Add-On Programme	19
1.3.6 Login über Benutzername und Paßwort (HTTP-Protokoll)	20
1.3.7 S-HTTP/SSL	21
1.4 Strategien zur Anpassung der Daten an den Benutzer	21
1.4.1 Explizite Abfrage	22
1.4.2 Suchweg-Verkürzung	23
1.4.3 Kollaboratives Filtern	23
1.4.4 Verhaltensbeobachtung	25
1.4.5 Fertige statistische Profile	25
<b>2 Marktübersicht: Programme zur Personalisierung</b>	<b>26</b>
2.1 Beurteilungskriterien	27
2.2 LearnSesame	28

2.3	GuestTrack	29
2.4	IntelliWeb	30
2.5	Firefly	31
2.6	SiteServer	32
2.7	ObjectStore	33
2.8	One-To-One	34
2.9	GroupLens	35
2.10	Dynamo Profile Station	36
2.11	Story Server	37
2.12	WebSell	38
2.13	WiseWire for WebSites	39
2.14	AgentWare	40
2.15	Zusammenfassung	41

# Link-Verzeichnis

## 1 Auswahl von personalisierten Websites

- MyYahoo: <http://my.yahoo.com> (15.7.98): Zusatz zur Suchmaschine Yahoo. Bietet hauptsächlich Nachrichten, Finanzinformationen, Unterhaltung, Sport, Reisen, Gesundheit und Shopping. Darstellung der Seite kann selbst eingestellt werden. Personalisierung nur nach Einstellungen des Benutzers, kein Lernen.
- Microsoft Network: <http://www.msn.com> (22.7.98): Die Informationsseite kann persönlich zusammengestellt werden. Keine Personalisierung nach persönlichen Interessen.
- Infoseek: <http://guide.infoseek.com/news?pg=personalize.html> (22.7.98): Keine echte Personalisierung (anonym, nicht nach pers. Interessen). Die Eingabe von Suchbegriffen sucht nach entsprechenden Nachrichten. Speicherung über ein Cookie.
- My Netscape: <http://my.netscape.com> (30.7.98): Preview (ganz neu, Preview). Bietet Bookmarks und Adressen, Taschenrechner sowie die üblichen Wettervorhersagen, Börseninformationen und Nachrichten. Graphisch gute Oberfläche.
- AOL.com: <http://www.aol.com/mynews> (27.7.98): Bietet die übliche Personalisierung: Nachrichten, Wetter, Börse, Sport, ...
- Excite: <http://my.excite.com> (22.7.98): Sehr gute Personalisierung. Die Seite kann nach eigenen Vorstellungen arrangiert werden und die interessanten Gebiete ausgesucht werden. Kein Lernen nach der Benutzung sondern ausschließlich nach Benutzerkonfiguration.
- Lycos (DE): <http://personal.lycos.de> (23.7.98) (Nur für Deutschland): Bietet Nachrichten, Kalender, Wetter, usw. Keine Registrierung erforderlich (**alles** in Cookies gespeichert).
- Lycos (USA): <http://personal.lycos.com> (23.7.98): Sehr ähnlich zu Excite. Bietet auch persönliche Kontakte, Erinnerungen, ...
- Lycos Webguides: <http://www.lycos.com/webguides/webguides.html> (27.7.98): Bietet eine Kategorisierung von Links, wobei die Personalisierung durch Bewertung der Ergebnisse erfolgt. Verwendet WiseWire.

Time Warner Money: <http://www.money.com> (23.7.98): Bietet persönliche Portfolios, sowie aktive Information bei bestimmten Änderungen. Kostenpflichtig (30 Tage freie Testmöglichkeit).

Hotmail: <http://www.hotmail.com> (23.7.98): Kostenloser E-Mail Account, der über das WWW zugänglich ist. Große Mächtigkeit. Wird durch Werbung finanziert.

Barnes&Noble: <http://www.barnesandnoble.com> (22.7.98): Online-Buchgeschäft, das auch Personalisierung anbietet. Man bekommt Buch-Empfehlungen, nachdem man selbst einige Bücher bewertet hat (Verwendet firefly Technologie).

Amazon: <http://www.amazon.com> (28.7.98): Online-Buchgeschäft, das auch Personalisierung anbietet. Buch-Empfehlungen erfolgen nach: Gekauften Büchern, Autoren, Buchbewertungen (mind. 20). Bietet auch eine Führung durch die Kategorisierung: ähnliche Autoren, Themengebiete, etc.

## 2 Produkte zur Personalisierung von Websites

LearnSesame: <http://www.opensesame.com> (20.7.98): Siehe 2.2

GuestTrack: <http://www.guesttrack.com/guesttrack.html> (20.7.98): Siehe 2.3

Intelliweb: <http://www.micromass.com/intelweb.html> (20.7.98): Siehe 2.4

Firefly: <http://www.agentsinc.com> (20.7.98): Siehe 2.5

SiteServer: <http://www.microsoft.com/siteserver> (20.7.98): Siehe 2.6

ObjectStore: <http://www.odi.com> (20.7.98); Personalization Object Manager:  
[http://www.odi.com/products/managers\\_section.html](http://www.odi.com/products/managers_section.html) (20.7.98): Siehe 2.7

One-To-One: <http://www.broadvision.com> (20.7.98): Siehe 2.8

GroupLens: <http://www.netperceptions.com/product/index.html> (21.7.98): Siehe 2.9

Dynamo Profile Station: <http://www.atg.com/develop/products/profile> (21.7.98): Siehe 2.10

StoryServer: <http://www.vignette.com> (23.7.98): Siehe 2.11

WebSell: <http://www.likeminds.com/solutions/websell> (21.7.98): Siehe 2.12

WiseWire: <http://www.wisewire.com> (27.7.98): Siehe 2.13

Agentware: <http://www.agentware.com> (27.7.98): Siehe 2.14

### 3 Verschiedenes

TrustE: <http://www.truste.org> (16.7.98): Non-profit Organisation, die sich um Vertraulichkeit von persönlichen Daten im WWW bemüht. Gibt allgemeine Richtlinien vor und empfiehlt Regelungen ("Korrekturen persönlicher Informationen sollten möglich sein"), schreibt aber keinen Inhalt vor.

Cookie Central: <http://www.cookiecentral.com> (16.7.98): Extensive Informationen über Cookies, ihre Anwendungen, die Probleme, die sie verursachen können und was man dagegen tun kann.

Richard Dean: Personalizing your Web site:

<http://www.builder.com/Business/Personal/ss01.html> (21.7.98): Allgemeiner Artikel über Personalisierung, der nirgends in die Tiefe geht.

Mark Palmer: A Personalization design pattern for dynamic websites:

[http://www.odi.com/content/white\\_papers/personalization/webapps.html](http://www.odi.com/content/white_papers/personalization/webapps.html) (22.7.98): Handelt hauptsächlich über den Einsatz von ODBMS in Personalisierung.

Richard V. Dragan: Advice from the Web:

[http://www.zdnet.com/pcmag/features/advice/\\_intro.htm](http://www.zdnet.com/pcmag/features/advice/_intro.htm) (27.7.98): Artikel über Empfehlungssysteme im Web.

Will Rodger: Industry Makes Online Privacy Enforcement Recommendations:

<http://www.zdnet.com/intweek/daily/980722a.html> (29.7.98): Artikel über den Umgang mit persönlichen Daten im Zusammenhang mit Websites



# Untersuchungen zur Personalisierung

## 1 Über die Personalisierung

In diesem Kapitel soll Allgemeines über die Personalisierung erklärt werden. Zuerst wird versucht, eine Definition zu geben, um eine gemeinsame Basis für das Verständnis des weiteren zu finden. Anschließend wird der Sinn und Zweck der Personalisierung untersucht. Weiters findet sich ein Abschnitt darüber, wie es möglich ist, einen Besucher auf einer Webseite wiederzuerkennen, was eine Grundvoraussetzung für die Anpassung des Inhalts ist. Zum Abschluß wird detailliert untersucht, welche Möglichkeiten es gibt, den Inhalt an den Benutzer anzupassen: Befragung nach Interessensgebieten, Lernen aus dem Verhalten, etc.

### 1.1 Allgemeines

Sinnvollerweise ist mit einer Definition zu beginnen, da unter Personalisierung mitunter sehr verschiedene Dinge verstanden werden. Anschließend werden die wichtigsten Probleme im Zusammenhang mit Personalisierung näher erläutert: Unrichtige Angaben, Änderungen der Interessen mit der Zeit, Wahrung der Privatsphäre und der Widerspruch zwischen einfachem Zugang und möglichst detaillierten Informationen.

#### 1.1.1 Definition

Es gibt keine allgemein anerkannte Definition, was unter Personalisierung zu verstehen ist, da dieser Begriff auch in sehr unterschiedlichem Kontext verwendet wird: Anpassung von Produkten an Kundenwünsche (auch Individualisierung), Electronic Commerce ("persönlicher Einkauf", auch wenn keinerlei Anpassung an den Benutzer erfolgt), Software-Konfiguration (Benutzereinstellungen), feste News-channel, Anpassung von Webseiten an Kunden oder Besucher und vieles mehr. Daher soll im weiteren unter Personalisierung folgendes verstanden werden:

***Anpassung von auf Webseiten angebotenen Informationen an die Interessen des jeweiligen Betrachters durch Auswahl und Darstellung interessanter und Ausscheiden und Weglassen uninteressanter Daten.***

### 1.1.2 Probleme für Anbieter einer personalisierten Website

Der Versuch, eine Website zu personalisieren, bringt mehrere Probleme mit sich. Diese sind im Einzelnen:

#### 1.1.2.1 Unrichtige Interessensangaben

Die Erfahrung zeigt, daß Personen oft unrichtige Angaben darüber machen, woran sie besonders interessiert sind. Dies hat seinen Grund nur in Ausnahmefällen in absichtlich falschen Angaben, sondern beruht meistens auf einem oder mehreren der folgenden Punkte:

- Die Person hat diesen Punkt vergessen.
- Der Benutzer weiß nicht, daß es dieses Gebiet auch online bzw. hier gibt.
- Es wird angegeben, woran man interessiert sein sollte (z. B. beruflich), aber nicht unbedingt ist.

Es ist daher notwendig, die Angaben des Benutzers nur als vorläufig und unvollständig zu betrachten. Nur für sehr einfache Kategorien der Personalisierung mit eindeutiger Datenklassifizierung (geographische Lage: z. B. Bundesland, Wohnort, ...; einige wenige grobe Interessengebiete) ist es möglich, sich auf die Angaben des Kunden zu beschränken. In allen anderen Fällen ist es notwendig, das Profil zumindest durch Beobachtung anzupassen (Siehe dazu im Detail unter Punkt 1.4).

#### 1.1.2.2 Änderungen des Interessensprofils

Im Zusammenhang mit den anfänglich unrichtigen Interessensangaben steht die nachträgliche Veränderung des Profils. Mit der Zeit ändern sich die Interessensgebiete, was bei der Personalisierung berücksichtigt werden muß. Dies kann wiederum durch verschiedene Möglichkeiten festgestellt werden.

Hierbei ist aber zu beachten, daß eine explizite Änderung des Profils durch den Benutzer sehr einfach durchzuführen sein muß, da die Kunden dies ansonsten nicht tun werden. Die Änderung der Interessen macht sich dann durch eine immer geringere Nutzung bemerkbar, da die Auswahl von immer geringerer Treffsicherheit und damit von sinkendem Nutzen ist.

Um größere Änderungen durch Beobachtung des Verhaltens erkennen zu können (z. B. der Benutzer interessiert sich nun auch noch für ein ganz anderes Gebiet und gibt dafür ein bisheriges auf), ist es notwendig, auch verwandte und andere Themen anzubieten, da er ansonsten das neue Thema auf anderem Wege besucht und keine Informationen darüber erlangt werden können. Beispielsweise steigt Kunde der Nutzer über eine Suchmaschine ein

und nicht über Links auf der Seite des Anbieters der Personalisierung, sodaß eine Beobachtung nur äußerst schwierig, wenn überhaupt, möglich ist. Dies ist eine besondere Gratwanderung, da der Benutzer ja nicht mit einer großen Menge für ihn nur potentiell interessanter Links überschwemmt werden möchte. Daher ist die Einbindung eines Suchmaschinen-Interfaces eine gute Idee für einen Anbieter solcher Seiten (Falls die Personalisierung nicht ohnedies vom Anbieter einer Suchmaschine ausgeht, was sehr oft der Fall ist, siehe Links).

#### 1.1.2.3 Rechtliche Probleme

Durch die (zumeist erfolgende) Beobachtung des Benutzers fallen große Mengen persönlicher Daten an, die vor unbefugtem Zugriff gesichert werden müssen. Darüber hinaus wären diese Informationen zwar von großem Wert für andere Firmen (z. B. als Adressen für zielgerichtete Werbung, ...), dürfen jedoch nur unter bestimmten Umständen weitergegeben werden (z. B. Anerkennung von Benützungsbedingungen durch den Kunden).

#### 1.1.2.4 Notwendigkeit eines großen Benutzerkreises

Für viele Techniken zur Personalisierung ist es notwendig, daß der regelmäßige Benutzerkreis groß ist. Dies ist notwendig, um aus dem Verhalten gültige statistische Daten ableiten zu können (insbesondere wenn die Personalisierung vom Vergleich mit anderen Benutzern abhängt). Auch die zusätzlichen Kosten für die Bereitstellung von personalisierbaren Informationen sind hier in Betracht zu ziehen, die sich nur bei vielen Benutzern rechnen.

#### 1.1.2.5 Einfacher Einstieg ↔ Detaillierte Informationen über den Benutzer

Da Benutzer nur sehr unwillig längere Fragebögen ausfüllen (zwei Seiten können schon zuviel sein), sind nur sehr spärliche Informationen über die Interessensgebiete vorhanden. Dies kann u. U. dadurch beseitigt werden, daß zusätzliche Dienste angeboten werden, die einzeln zu konfigurieren sind (Angabe der geograph. Position ermöglicht lokalen Wetterbericht). Dies ist besonders in Hinblick auf den vorigen Punkt wichtig, da detaillierte Abfrage von Informationen nicht nur höchst ungern beantwortet werden ("Was geht das eigentlich die an?"), sondern auch den Benutzerkreis verringern ("Das ist mir zu mühsam!"). Hier ist ein tradeoff besonders schwierig, da nur allgemeine Informationen zwar die Anzahl der Benutzer erhöht, jedoch durch die ungenaue Anpassung nur einen geringen Nutzen bringt und der Ausstieg sehr leicht ist und naheliegt. Dies ist auch mit ein Grund für die Verwendung von statistischen Benutzerprofilen, um neue Kunden mit wenigen Fragen in ein detailliertes Profil einzupassen, welches mit der Zeit dann für die einzelne Person maßgeschneidert wird.

#### 1.1.2.6 Höherer Softwareaufwand

Da auf Grund der Personalisierung die Seiten jeweils für jeden Benutzer gesondert konstruiert werden müssen, besteht ein höherer Softwareaufwand, da ein einfacher Webserver dafür nicht mehr ausreicht. Die meisten kommerziellen Server sind allerdings heute schon dazu in der Lage und es ist lediglich eine Frage des Betreuungsaufwandes, der natürlich ebenfalls steigt.

#### 1.1.2.7 Klassifizierung der Daten

Im Gegensatz zum Softwareaufwand, der nur unwesentlich steigt, ist ein erheblich höherer Betreuungsaufwand gegeben. Er resultiert daraus, daß die Daten, die personalisiert dargestellt werden sollen, in den meisten Fällen klassifiziert werden müssen, um dem Computer eine Auswahl zu ermöglichen. Dies ist auch heute unmöglich zu umgehen, da eine automatische Erkennung des Inhalts zur Zeit unmöglich ist. Lediglich sehr grobe Klassifizierungen nach sehr allgemeinen Kriterien (z. B. Texte in Sport, Wirtschaft, Politik, ... einteilen) können automatisch erfolgen, sind aber auch nicht immer zuverlässig. Es müssen also für jedes Datum Metadaten vorhanden sein, die entweder extern (z. B. in einer Datenbank) oder im Dokument selbst gespeichert sind (z. B. Schlüsselwörter im HTML-Header, Dateiinformati-onen wie Titel, Autor, ...). Durch den Vergleich dieser Zusatzinformationen mit den Präfe-renzen des Benutzers kann dann die Software auswählen, was dem Benutzer präsentiert wird.

Ein für die Zukunft erfolgversprechendes Konzept zur Speicherung von Meta-Informationen im Dokument selbst ist XML<sup>1</sup>. Hiermit können sowohl Struktur als auch zusätzliche Informationen in Dateien integriert werden. Dennoch ist auch dies keine absolute Lösung, da es sich um einen sehr offenen Standard handelt. So können beispielsweise alle möglichen Informationen über die Struktur und Hinweise zur Darstellung in einem Dokument enthalten sein, jedoch jegliche Klassifizierung fehlen. In einem anderen Fall könnte diese Klassifizierung auch nach völlig anderen Regeln und Mustern erfolgen, sodaß sie im Endeffekt vom verwendeten Programm nicht verstanden wird. Es ergibt sich also der Schluß, daß auch bei Verwendung von XML nur in gleicher Weise (nach den gleichen Regeln) annotierte Dokumente eine zusätzliche Klassifizierung erübrigen. In allen anderen Fällen und bis zur Verwendung dieses oder eines ähnlichen Standards wird sich eine explizite und händische Klassifizierung nur in sehr einfachen Fällen vermeiden lassen (Ausnahme: einige Anpassungsstrategien; Siehe 1.4).

---

<sup>1</sup> XML: Extensible Markup Language; Siehe auch <http://www.w3.org/XML/> (15.7.98), [http://javadomain.com/java\\_xml.html](http://javadomain.com/java_xml.html) (15.7.98), <http://www.arbortext.com/xml.html> (15.7.98),

## 1.2 Sinn und Zweck der Personalisierung

Personalisierung bringt sowohl für den Anbieter als auch für den Kunden Vorteile, sodaß es im beiderseitigen Interesse liegt, die (natürlich auch vorhandenen) Nachteile zu überwinden bzw. in Kauf zu nehmen.

### 1.2.1 Zielgerichtete Werbung

Personalisierung hat für den Anbieter von Webseiten den Vorteil, daß er mehr Informationen über seine Benutzer erhält. Dies ermöglicht es ihm, bei ihm gebuchte Werbeeinschaltungen viel zielgerichteter zu plazieren und damit höhere click-through-Raten zu erreichen. In einer ähnlichen Weise wird dies bereits bei Suchmaschinen eingesetzt, wo die Werbung nach den eingegebenen Suchworten ausgewählt wird. Durch die viel extensiveren Informationen, die bei personalisierten Diensten zur Verfügung stehen, kann die Treffsicherheit noch weiter erhöht werden. Unter anderem auch dadurch, indem die Präsentation desselben Produkts dem jeweiligen (potentiellen) Kunden angepaßt wird. Weiters sind Anzeigen umso erfolgreicher, je besser Sie in die Umgebung passen, in die sie eingebettet sind: Eine Anzeige für schnelle Sportwagen wird auf der Seite mit Informationen über die Formel 1 erfolgreicher sein, als auf den Wirtschaftsnachrichten, auch wenn sie von derselben Person gelesen wird, die sich grundsätzlich dafür interessiert. Klickt der Benutzer nun auf eine Anzeige, so ermöglicht es Personalisierung dem Anbieter, auch gleich ein passendes Angebot zu stellen oder die Verkaufsabwicklung zu vereinfachen.

### 1.2.2 Erhöhte Kundenbindung bei Electronic Commerce

Der Einsatz von personalisierten Webseiten in Verbindung mit Electronic Commerce kann zu höherer Bindung der Kunden führen. Dies beruht auf mehreren Gründen:

- Der Kunde wird persönlich angesprochen und ist kein anonymer Besucher mehr. Dies erzeugt ein gewisses Gefühl der Verbundenheit und kann ihn zur Wiederkehr animieren. Es wird ihm sozusagen mitgeteilt, daß er als Person willkommen ist und nicht einfach irgend jemand ist, dem etwas verkauft werden soll.
- Dem Kunden können speziell abgestimmte Angebote gemacht werden: Kunden, die diejenigen Dinge angeboten bekommen, die sie brauchen/interessieren, haben eine viel stärkere Bindung als solche, die viele nicht relevante Produkte angeboten erhalten.
- Da die Kundeninteressen und seine bisher gekauften Waren bekannt sind, ist eine weitere Betreuung möglich, um zusätzliche Produkte gezielt zu bewerben (Siehe auch Punkt 1.2.1).

### 1.2.3 Speziell abgestimmte Informationen

Auch für den Benutzer bietet Personalisierung einen großen Vorteil: Er erhält gezielt die Informationen, die ihn interessieren, zugestellt (Push-Technologie) oder zur Verfügung gestellt, ohne daß er sich selbst darum kümmern muß. Dies bedeutet eine (unter Umständen erhebliche) Arbeitersparnis, was allgemein als positiv bewertet wird. Für den Benutzer sind wenige, dafür ausgewählte und relevante, Informationen von größerem Wert, als wenn er zwar enorme Mengen an Daten erhält, aber nicht weiß, ob und wenn ja wo, sich die wichtigen Teile befinden. Auch die Zusammenstellung von Informationen aus verschiedenen Quellen bringt zusätzlichen Wert, da dies eine eindeutige Zeit- und Arbeitersparnis bedeutet.

### 1.2.4 Reduktion der Datenflut

Sowohl der einzelne Benutzer wie auch die Anbieter auf dem WWW profitieren von Personalisierung in der Weise, daß nur mehr relevante Informationen übertragen werden und daher das Bandbreitenproblem verringert wird. Im Gegensatz dazu werden sonst große Datenmengen geladen, weil wichtige Informationen in ihnen vermutet werden (z. B. einen Link in einer Suchmaschine gefunden). Doch in sehr vielen Fällen stellt sich dann heraus, daß die Information doch nicht die gewünschte ist und die Daten werden einfach vernichtet. Bis jedoch diese Entscheidung getroffen wird, wurde bereits eine erhebliche Bandbreite belegt und stand für wirklich wichtige Informationen nicht zur Verfügung.

Dies kann zwar das Bandbreitenproblem im Großen nicht lösen, aber zumindest verringern. Für den Betreiber einer einzelnen größeren Website kann dies jedoch einen erheblichen Unterschied machen, da er unter Umständen einen viel schmäleren Zugang zum Internet benötigt. Er hat dann zwar absolut weniger Besucher auf seiner Seite, doch sind diese für ihn wertvoller, da sie wirklich am Inhalt interessiert sind.

### 1.2.5 Verteilung von Informationen in Betrieben

Personalisierung kann auch dazu verwendet werden, Informationen in Betrieben schnell und zielsicher den Personen zur Kenntnis zu bringen, die sie benötigen. Dies ist ein entscheidender Fortschritt gegenüber firmeninternen Wissens-Datenbanken, die bei Bedarf durchsucht werden. Aufgrund des dazu notwendigen expliziten Aufwandes wird dies gerne unterlassen und daher trotzdem vieles mehrfach durchgeführt. Ist hingegen das Personalisierungssystem über die Arbeit unterrichtet, kann es automatisch verwandte Informationen suchen und neue Daten direkt an die Betroffenen weiterleiten, ohne daß der Ersteller der Information wissen muß, wen sie aller betrifft.

### 1.2.6 Genaue Benutzerüberwachung möglich

Natürlich gibt es auch (zumindest potentielle) Nachteile. Wenn man ein persönlich maßgeschneidertes Service erhält, ist es natürlich ohne weiteres möglich, die dafür unabdingbar notwendigen Informationen auch auf eine Weise zu verwenden, die nicht den Wünschen des Benutzers entspricht. Dies kann von verstärkter Werbung (nur lästig) über Vertreterbesuche (unangenehm) bis hin zur Verwertung durch staatliche Behörden (u.U. sehr unangenehm) gehen. Leider gibt es auch keine erfolgversprechenden Ansätze, um dieses Problem zu verhindern.

Die meisten der Suchmaschinen, die auch personalisierte Nachrichten anbieten (Siehe Link-Verzeichnis), bieten Informationen über die Verwendung der persönlichen Daten eher verschämt an (winzige Links auf nur wenigen Seiten, etwa auch nicht auf der Startseite). Allgemein ist der Standard in Bezug auf die Vertraulichkeit eher gering (z. B. Weitergabe von persönlichen Informationen nur dann, wenn die Firma **annimmt**, daß sie gesetzlich dazu verpflichtet sei). Dennoch sind wenigstens einige wichtige Grundmerkmale erfüllt: Informationen werden an Werbungs-Inserenten nur in aggregierter Form weitergegeben (z. B. 65 % der Besucher sind männlich) und die E-Mail Adressen werden nicht weitergegeben. Ebenso kann der Empfang von weiteren Informationen oder Werbungen unterbunden werden. Dennoch ist meiner Meinung nach eine allgemeine Regelung, welche Verwendung der persönlichen Daten erlaubt ist, notwendig, wobei auch Überwachung und Sanktionen miteinbezogen werden müssen. Nur dann haben Websites mit Personalisierung wirklich die Chance, das Vertrauen der Besucher zu erhalten.

### 1.3 Wiedererkennen des Benutzers

Um die angebotenen Informationen auf den einzelnen Besucher anzupassen, ist es notwendig, diesen Besucher zu identifizieren, da es undenkbar ist, jeweils alle Informationen bei jedem Besuch neu eingeben zu lassen (was auch jegliches Lernen aus Beobachtung des Benutzers verhindern würde). Dies ist mit erheblichen technischen Schwierigkeiten verbunden, unter anderem weil alle diese Mechanismen auch auf eine Art verwendet werden können, die nicht den Erwartungen des Benutzers entspricht. Ebenso gibt es sehr viele verschiedene technische Möglichkeiten, die jedoch alle verschiedene Nachteile haben, sodaß eine Kombination verwendet werden muß.

Es ist bei den im folgenden vorgestellten Möglichkeiten zu beachten, daß meistens nur der Rechner identifiziert wird, aber nicht unbedingt, welche Person diesen nun wirklich benützt. Dies bereitet in zweierlei Hinsicht Probleme:

- 1) Es gibt Benutzer, die mehrere Rechner verwenden und daher eventuell als verschiedene Personen identifiziert werden, wenn keine Vorsichtsmaßnahmen getroffen werden.
- 2) Es gibt Rechner, die von mehreren Personen benutzt werden. Diese müssen trotzdem auseinandergehalten werden, was meist schwierig ist.

### 1.3.1 Benutzername und Paßwort

Die einfachste und üblichste Form ist die, daß jeder Benutzer eine Kennung und ein Paßwort erhält, welche er beim Besuch der Seite auf einem Formular eingeben muß (Eine andere Art der Paßworteingabe siehe unter Punkt 1.3.6). Da HTML jedoch ein zustandsloses Protokoll ist (jede Seite wird einzeln und unabhängig von früheren Anforderungen gesondert geladen, wobei keine Informationen für spätere Anforderungen aufgehoben werden, wie z. B. der Browsertyp oder die Adresse des Anforderers), reicht dies nur aus, um eine einzige Seite zu personalisieren, die direkt auf die Eingabe folgende. Da dies jedoch nicht ausreicht, muß dieses Verfahren mit einem der unten angeführten kombiniert werden. Diese Methode hat den großen Vorteil, daß es die einzige ist, die wirklich den Benutzer identifiziert und nicht den Rechner oder etwas anderes (solange das Paßwort geheim bleibt).

Da das Paßwort jedoch leicht zu vergessen ist (insbesondere wenn man sich bei mehreren Diensten anmeldet, wo jeweils ein eigener Benutzername und ein eigenes Paßwort notwendig ist), werden oft zusätzliche Hinweise gegeben:

- Bei Angabe der E-Mail Adresse (muß schon vorher mit dem Benutzer assoziiert worden sein), wird das Paßwort per E-Mail an diese Adresse geschickt: Nur der wirkliche Benutzer hat Zugang zu dieser Mailbox (hoffentlich).
- Beim Anlegen des Accounts können zusätzliche Hinweise zum Paßwort angegeben werden, die das System später anzeigt. Hierbei ist es natürlich besonders wichtig, daß diese Hinweise nur für den tatsächlich Berechtigten etwas bedeuten (Eselsbrücke).

Im Hinblick darauf, welche Menge an persönlichen Informationen über dieses Paßwort zugänglich sind, ist die Sicherheit erschreckend gering: Unverschlüsselte Übertragung über das Internet durch Eingabe in Formulare, ungesicherter Versand über E-Mail, keinerlei zusätzliche Prüfungen, etc. Einzig und allein bei Websites, die auch kommerziell tätig sind (Electronic Commerce), sind die Sicherheitsanstrengungen größer, da zumindest bei der

Übertragung von Kreditkartennummern auch das Problembewußtsein der Anwender höher ist und diese daher eine höhere Sicherheit verlangen.

### 1.3.2 Cookies

Die in Verbindung mit der vorigen Variante üblichste Methode sind Cookies<sup>2</sup>. Es handelt sich hierbei um kurze Textdateien, die der Webserver auf dem Rechner des Benutzers ablegt und auch wieder auslesen kann. Sie wurden unter anderem für genau diesen Zweck erfunden, aber von einigen Firmen auch für andere Zwecke mißbraucht und sind so in Verruf gekommen. Manche Benutzer lassen daher Cookies auf ihrem Computer grundsätzlich nicht zu.

Cookies ermöglichen es einer Website, ohne Befassung des Benutzers völlig unbemerkt die Identifizierung vorzunehmen und so die Seiten entsprechend zusammenzustellen. Dies ist für den Benutzer sicher die angenehmste Art, da er keinerlei Aktionen dafür setzen muß. Dennoch gibt es auch einige Nachteile:

- Wie bereits erwähnt, verbieten manche Benutzer die Speicherung von Cookies und verhindern daher effektiv diese Form der Identifizierung.
- Obwohl im Standard<sup>3</sup> keine Beschränkungen vorgesehen sind, haben die derzeitigen Implementierungen meist die Mindestanforderungen auch als Höchstlimits festgesetzt: Maximal 20 Cookies von einem Server und alle Cookies können höchstens 4 kB groß sein. Dies erlaubt es normalerweise nicht, die Daten zur Personalisierung direkt in den Cookies zu speichern, sondern ermöglicht nur die Identifizierung des Benutzers, um die Daten dann lokal von einem Server des Anbieters zu holen.
- Cookies haben ein Ablaufdatum (wird allerdings vom Server gesetzt) und können jederzeit gelöscht werden (z. B. bei Platzmangel). Der Server muß daher damit rechnen, daß Cookies plötzlich verschwinden (u. U. auch während des Besuchs der Seite!) und daher auch zumindest eine andere Form der Identifizierung beherrschen.
- Cookies identifizieren nicht den Benutzer sondern nur den Rechner. Es sollte daher eine Möglichkeit angeboten werden, daß/die Cookies zu löschen, wenn man den Computer verläßt, damit der folgende Benutzer nicht falsch erkannt wird.

---

<sup>2</sup> Siehe dazu Link-Verzeichnis, aber auch Netscape: [http://home.netscape.com/newsref/std/cookie\\_spec.html](http://home.netscape.com/newsref/std/cookie_spec.html) (16.7.98), wo ein kurzer Überblick zu finden ist.

<sup>3</sup> Siehe RFC 2109: <http://www.pasteur.fr/cgi-bin/mfs/01/21xx/2109> (16.7.98)

All diese Nachteile sollen jedoch nicht darüber hinwegtäuschen, daß Cookies zur Zeit die komfortabelste (keine Paßworteingabe, ...), einfachste (keine Aktion des Benutzer nötig; einfach zu verwenden) und verlässlichste (keine Probleme mit Schreibfehlern) Methode sind. Die einzigen Probleme, die damit nicht gelöst werden können, sind mehrere Benutzer an einem Computer und ein Benutzer mit mehreren Rechnern.

### 1.3.3 Spezielle URLs mit codierter Benutzererkennung

Eine Benutzererkennung kann auch direkt in einen URL codiert werden. Dies führt jedoch dazu, daß der URL praktisch unlesbar wird (Klartext wird aus Sicherheitsgründen nicht verwendet, sondern eine verschlüsselte Form) und man ihn sich nicht mehr merken kann, wodurch ein Transport zu einem anderen Rechner sehr schwierig wird. Er muß daher jedesmal aus einer E-Mail (Zusendung nach Anmeldung) oder einer Liste (Bookmarks) herausgepickt werden, was einigermaßen unkomfortabel ist. Hier wird weder der Benutzer noch ein bestimmter Rechner identifiziert, sondern allein die Kenntnis des URLs, was für eine eindeutige Identifizierung eine unsichere Sache ist (sie werden üblicherweise unverschlüsselt übertragen; dafür ist ein über die Schulter schauen und merken auszuschließen). Mit abnehmender Bedeutung des CGI-Interfaces gerät diese Form langsam in Vergessenheit, wird aber manchmal eingesetzt, wenn ein Benutzer keine Cookies auf seinem Rechner zuläßt.

### 1.3.4 Position des Benutzers

Eine im Internet sehr einfache Möglichkeit ist, den Rechner, den ein Benutzer benützt, an seiner eindeutigen IP-Adresse zu erkennen. Da jedoch immer mehr Firmen dazu übergehen, nicht einfach global gültige Internet-Adressen zu verwenden (auch weil diese sehr knapp sind), sondern ein lokales Intranet mit privaten Adressen zu verwenden, ist diese Methode nur mehr sehr eingeschränkt verwendbar. Ein Intranet wird üblicherweise über einen Proxy oder einen Firewall an das Internet angeschlossen, wobei eine Adressumsetzung erfolgt (so auch bei praktisch allen Accounts bei einem Internet-Provider). In diesem Fall werden alle (dahinterliegenden) privaten Adressen auf eine einzige externe Adresse umgewandelt, sodaß keinerlei Benutzererkennung mehr möglich ist. Zusätzlich kommt hinzu, daß es nicht möglich ist festzustellen, ob dieser Fall vorliegt oder eine global eindeutige Adresse verwendet wird (und auch nur eine geringe Anzahl der Internet-Teilnehmer darüber informiert ist, welcher Fall für sie zutrifft). Auch hier stellt sich das Problem, daß nur der Rechner und nicht der Benutzer identifiziert wird.

### 1.3.5 Add-On Programme

Viele Services verwenden spezielle Programme (eigenständige Browser, die entweder nur Ihre Daten darstellen können, oder auch einen Webbrowser integriert haben oder in einen Webbrowser eingebaut werden), um Ihre Daten darzustellen. Dies ist insbesondere bei Push-Technik üblich, da kein allgemeiner Standard existiert. Diese Software kann natürlich feststellen, auf welchem Computer sie läuft, welcher Benutzer eingeloggt ist und diese Daten dann über das (spezielle) Protokoll an den Server schicken, welcher dann Seiten persönlich maßschneidern kann. Bisher wurden solche Programme aber nur dazu verwendet, nach der eingestellten Konfiguration bestimmte Kategorien von Daten vom Server zu holen/zuz empfangen und anschließend darzustellen, die Personalisierung erfolgte also auf dem Rechner des Benutzers und nicht des Anbieters. Dies erlaubt eine größere Freiheit in den angebotenen Diensten und der Darstellung, hat jedoch auch mehrere Nachteile:

- Ein zusätzliches Programm ist zu installieren, wobei für jede Plattform eine eigene Version angeboten werden muß. Bei einer Änderung oder Erweiterung des Angebots zur Personalisierung, das sich nicht auf neue Kategorien beschränkt, ist eine neue Version der Software nötig.
- Es gibt keinen allgemeinen Standard, man ist daher je nach Software auf bestimmte Anwender/Zielgruppen fest ausgerichtet.
- Der Benutzer kann nicht mehr den Rechner wechseln, da die Einstellungen und das Programm lokal gespeichert sind. Es wäre also nötig, das Programm mehrfach zu installieren und auch die Konfiguration zu übernehmen.
- Bei einer Verwendung in lokalen Netzwerken (Verbindung über einen Proxy) kann es zu Schwierigkeiten kommen, da die übertragenen Daten nur selten dem HTTP-Protokoll entsprechen und daher eventuell eine Spezialkonfiguration notwendig ist.

Bei einer Minimalversion eines zusätzlichen Programmes übernimmt dieses nur die Identifizierung des Benutzers und überläßt die restliche Arbeit dem Server. Dies beseitigt zwar einige Probleme, die grundlegenden Nachteile bleiben aber bestehen.

### 1.3.6 Login über Benutzername und Paßwort (HTTP-Protokoll)

In den Spezifikationen des HTTP-Protokolls<sup>4</sup>, das für die Übertragung von WWW-Seiten verwendet wird, ist die Möglichkeit der Authentifizierung des Benutzers vorgesehen. In der derzeit fast überall verwendeten Version 1.0 erfolgt die Übertragung jedoch im Klartext (Bestimmte Server/Browser erlauben auch proprietäre (sichere), aber nicht allgemein verwendbare, Übertragungarten). Bei Version 1.1 existiert ein besseres System (Name und Paßwort werden verschlüsselt übertragen) erst als Vorschlag. Das größte Problem hierbei ist, daß die meisten Webserver mit diesem System nur den Zugriff zu festen Seiten regeln können, es aber keine Schnittstelle gibt, um aufgrund der empfangenen Identifikation spezielle Seiten zusammenszustellen.

Bei dieser Methode wird vor dem Senden jeder Seite vom Server eine Identifizierung des Benutzers verlangt. Der Benutzer muß Name und Paßwort aber nur einmal eingeben, da beides vom Browser gespeichert und bei der nächsten Anfrage wieder verwendet wird. Die Sicherheit dieses Ansatzes hängt davon ab, wie sicher der Browser die Daten speichert. Üblicherweise bleiben sie nur solange existent, wie das Programm läuft und gehen dann verloren. Das bedeutet allerdings, daß sich der Benutzer bei jedem neuen Start erneut identifizieren muß, was z. B. gegenüber Cookies ein Nachteil ist. Ein weiterer Nachteil ist noch, daß die Anfrage nach Name und Paßwort in einem Fenster erfolgt, auf das der Ersteller der Seite keinen Einfluß hat. Dies bedeutet, daß dem Benutzer ein ganz einfacher Dialog präsentiert wird, auf dem er Name und Paßwort eingeben soll, ohne daß irgendwelche Erläuterungen möglich sind. Bei direktem Zugriff auf eine solche Seite über einen URL weiß der Anwender nicht einmal, wohin diese Daten geschickt werden, da er erst danach die erste Seite präsentiert bekommt. Ein Einstieg wäre daher nur über den Umweg einer Informationsseite mit einem Link zur tatsächlichen (dann personalisierten) Seite möglich. Im Gegensatz dazu ist bei der Verwendung von Formularen eine völlig freie Gestaltung und beliebige Erläuterung möglich. Die Notwendigkeit einer Einstiegsseite läßt sich hier jedoch auch nicht verhindern, doch ist sie komfortabler, da dort die Benutzeridentifikation direkt eingegeben werden kann. Ein weiterer Nachteil ist, daß bei falschem Benutzernamen oder Paßwort es nur die Möglichkeiten gibt, andere Werte einzugeben, oder abzubrechen. Die Lieferung einer speziellen Seite (z. B. zur Neuanmeldung) ist nicht möglich. Siehe dazu auch das unter Punkt 1.3.1 angeführte in entsprechender Weise.

---

<sup>4</sup> Siehe dazu RFC 1945 für HTTP/1.0 und RFC 2068 für HTTP/1.1 sowie RFC 2069 für eine bessere Authentifizierung: <http://www.w3.org/Protocols/Specs.html> (18.7.98)

### 1.3.7 S-HTTP/SSL

Bei diesen beiden Protokollen<sup>5</sup> handelt es sich um Möglichkeiten, HTTP-Verbindungen (S-HTTP) bzw. allgemein Netzwerkverbindungen (SSL) zu sichern, indem der Nachrichteninhalte verschlüsselt und die Teilnehmer authentifiziert werden. Heute ist hauptsächlich SSL in Verwendung, dient aber nur dazu, den Server verlässlich zu identifizieren und die Übertragung zu verschlüsseln, wobei jedoch der Benutzer anonym bleibt. Auch diese Methode der Identifizierung ist für einen gesicherten Zugriff auf feste Seiten vorgesehen, sodaß eine Adaptierung auf eigens zusammengestellte Seiten nach der bei der Authentifizierung festgestellten Identität schwierig und wahrscheinlich nur über Zusatzprogramme möglich wäre. Da sich ein Benutzer mit einem von einer Zertifizierungsinstanz bestätigtem Zertifikat anmelden muß, ist keinerlei Anonymität mehr gegeben: der Besucher muß sich praktisch ausweisen, was eine psychologische Schranke ist. Auch hier stellt sich die Frage, wie dieses Zertifikat verwaltet wird: Wird der Benutzer oder der Web-Browser identifiziert, wie gut ist der Zugriff auf fremde Zertifikate abgesichert und wie einfach kann ein Zertifikat transportiert werden? Dennoch hat dieser Ansatz auch Vorteile: die Authentifizierung erfolgt vollkommen automatisch und somit unbemerkt vom Benutzer. Im Gegensatz zu Cookies muß hier auch nicht damit gerechnet werden, daß die Identifikations-Information jederzeit verschwinden kann.

## 1.4 Strategien zur Anpassung der Daten an den Benutzer

Nachdem der Benutzer identifiziert ist, müssen nun für ihn relevante Daten zusammengestellt werden. Dazu ist einerseits in den meisten Fällen eine (relativ einfache aber arbeitsaufwendige) Klassifizierung der angebotenen Daten und andererseits die Feststellung der Präferenzen dieses besonderen Benutzers notwendig. Das Abstimmen der Daten mit dem Interessensprofil sowie die Präsentation auf einer einzigen Seite in einem bestimmten Layout ist dann eine (technisch) meist relativ einfache Aufgabe. Mit welchen Möglichkeiten die Präferenzen und Wünsche der Benutzer herausgefunden werden können, wird im folgenden erläutert. Es werden hier die reinen Methoden mit den jeweiligen Vor- und Nachteilen erläutert. In vielen Fällen ist jedoch eine Mischung zweier oder mehrerer Methoden sinnvoll, da bei jeder Art zumindest ein besonders schwerwiegendes Problem besteht, das für einen guten Erfolg umgangen werden sollte.

---

<sup>5</sup> Siehe zu S-HTTP: <http://search.ietf.org/internet-drafts/draft-ietf-wts-shttp-06.txt> (18.7.98) und **Vorläufiger Entwurf**. SSL: <http://home.netscape.com/eng/ssl3/draft302.txt> (18.7.98) sowie <http://www.pcwebopedia.com/SSL.htm> (20.7.98)

### 1.4.1 Explizite Abfrage

Die einfachste Möglichkeit, um die Interessen des Benutzers herauszufinden, ist, ihn einfach zu fragen. Dies erlaubt es dann, anhand von fertigen Regeln den Inhalt von Webseiten anzupassen: Wenn er als Sportart Surfen betreibt, wird ihm ein Meeresurlaub angeboten, betreibt er jedoch Bergsteigen, so werden andere Reiseziele vorgeschlagen. Zumindest als Einstieg ist diese Methode sehr geeignet, da initiale Daten eine relativ gute Personalisierung erlauben.

#### 1.4.1.1 Vorteile

- Dies ist die am einfachsten zu implementierende Möglichkeit, die auch den geringsten Aufwand an Rechenzeit und Speicherplatz benötigt.
- Daten können bei der Registrierung einfach mit einem Formular abgefragt werden.
- Nach den Fragen ist einfach festzustellen, wie die Daten in Klassen einzuteilen sind.
- Sind die Daten den Fragen entsprechend klassifiziert, sind die Regeln meist sehr einfach aufzustellen.
- Es ist eine eingeschränkte Vorhersage der Benutzerinteressen nach Regeln möglich: Wenn der Kunde A und B mag, dann wird er auch C mögen. Das Finden dieser Regeln ist schwierig (Sie können im Einzelfall auch völlig falsche Ergebnisse liefern).

#### 1.4.1.2 Nachteile

- Die Personalisierung ist relativ ungenau, da nur eine begrenzte Anzahl von Informationen abgefragt werden kann. Kein Benutzer wird 10 Seiten voll mit Fragen beantworten! Dies hat einerseits den Grund in der kurzen Aufmerksamkeitsspanne und andererseits im Schutz der Privatsphäre: niemand läßt sich gerne so lange ausfragen.
- Wenn der Besucher seine Interessen nicht selbst wieder explizit ändert, bleibt die Personalisierung gleich. Da sich jedoch die Benutzervorlieben mit der Zeit sicherlich verändern, wird die Personalisierung so mit der Zeit für ihn immer weniger wert.
- Um die Personalisierung sinnvoll zu betreiben, ist eine komplexe Klassifizierung notwendig. Dies bedeutet einen Arbeitsaufwand und klare Richtlinien, damit verschiedene Personen gleichartige Daten gleich klassifizieren.
- Da Benutzer oft auf manche Gebiete vergessen oder weniger wichtige auslassen und eben wegen der nicht unbedingt immer vollen Treffsicherheit des Systems, ist es sinnvoll, einen Graubereich rund um die festgestellten Interessen anzubieten. In vielen Fällen erfolgt dies einfach dadurch, daß nicht die gesamte Seite personalisiert wird, sondern es einen allgemeinen Teil gibt, der für alle Personen gleich ist.

### 1.4.2 Suchweg-Verkürzung

Diese Art der Benutzeranpassung besteht daraus, daß Links, denen öfter gefolgt wurde, immer höher in der Hierarchie hinaufwandern. Auf diese Art wird die Dauer der Navigation zu öfter benötigten Informationen stark verkürzt. In manchen Fällen erfolgt auch noch die Abfrage einer Bewertung der Qualität des Ergebnisses (z. B. bei Suchmaschinen), wodurch die Anpassung schneller vonstatten geht.

#### 1.4.2.1 Vorteile

- Diese Strategie lernt dauernd hinzu und kann daher auf sich ändernde Interessen dynamisch reagieren.
- Die Anpassung erfolgt genau für diese eine Person und basiert nicht auf Erfahrungswerten oder statistischen Kennzahlen.
- Es ist keine Klassifizierung der Daten notwendig. Jede beliebige Art von Daten (Bilder, Text, Applikationen, etc.) kann so personalisiert werden.
- Der Benutzer muß keine expliziten Angaben machen und es ist keine Registrierung notwendig (lediglich eine Identifizierung als dieser Besucher).

#### 1.4.2.2 Nachteile

- Die Personalisierung benötigt einige Zeit, um Resultate zu zeigen und schreitet nur langsam fort. Ebenso reagiert sie nur langsam auf Änderungen.
- Nur regelmäßige Suchen können so optimiert werden. Ändert sich das Informationsangebot oder die Interessen, wird die bisherige Personalisierung komplett wertlos.
- Es sind keine Vorhersagen möglich, ob der Benutzer sich für ein bestimmtes, bisher nicht angebotenes, Thema interessieren wird.

### 1.4.3 Kollaboratives Filtern

Der Gedanke hinter dieser Anpassungsstrategie ist, daß viele Personen eine Bewertung über die Daten abgeben und dann versucht wird, Gruppen von Benutzern mit ähnlichen Bewertungen zu finden. Nun wird versucht, jeden Benutzer in eine solche Gruppe einzuordnen, wodurch ihm alle Bewertungen dieser Gruppe als zumindest einigermaßen zutreffend zugeschrieben werden. Beispiel: Wenn ein Benutzer A und B mag und seine Gruppe sich für A und B und C interessiert, wird auch er C wahrscheinlich gut finden. Der Prototyp dieser Art von Personalisierung sind Kinofilme, die von Besuchern bewertet werden. Anhand der Einordnung in eine bestimmte Gruppe werden dann andere Filme empfohlen.

#### 1.4.3.1 Vorteile

- Eine Klassifizierung der Daten ist nicht erforderlich. Es können daher verschiedenste Datentypen und diese auch untereinander vermischt verwendet werden.
- Je mehr Benutzer teilnehmen und je mehr Elemente diese bewerten, um so besser wird die Personalisierung. Diese Methode lernt ständig hinzu.
- Die Personalisierung ist zwar nicht auf einzelne Personen bezogen, entspricht aber bei fast allen Teilnehmern sehr gut den tatsächlichen Interessen.
- Es ist eine Vorhersage des Interesses für völlig andere Gebiete möglich (inkl. Wahrscheinlichkeit des Zutreffens), sofern diese von den selben Leuten bewertet wurden.
- Schon mit wenigen Bewertungen können bei neuen Benutzern relativ gute Vorhersagen getroffen werden, sofern eine große Datenbasis vorliegt.

#### 1.4.3.2 Nachteile

- Ein kommerzielles Paket ist notwendig, da die notwendigen Algorithmen sehr kompliziert sind. Ein Vergleich der Ergebnisse und damit der Produkte ist ohne eine praktische Anwendung und einen Test unmöglich.
- Die Bildung von Gruppen ist ein Angelpunkt dieser Strategie und kann nur schwer vorhergesagt werden. Die Qualität der Ergebnisse kann daher differieren. Manche Produkte versuchen das zu umgehen, indem ein einzelner, möglichst passender, Vergleichsbenutzer gesucht wird.
- Eine größere Anzahl von regelmäßigen Benutzern ist notwendig, um sinnvolle und gute Ergebnisse zu erhalten.
- Der Benutzer muß immer wieder Bewertungen abgeben, auch wenn die Personalisierung (für ihn) schon relativ gut ist.
- Der Beginn ist schwierig, da noch keine Bewertungen vorliegen. Es gibt zwar Möglichkeiten zur Abhilfe (die Ergebnisse wie Regeln liefern), doch sind diese qualitativ nicht besonders gut.
- Sind die Daten nur von kurzfristigem Interesse (z. B. Nachrichtenartikel), so kann diese Methode nicht angewendet werden, da die Information veraltet ist, bevor genügend Bewertungen zur Personalisierung vorliegen.

#### 1.4.4 Verhaltensbeobachtung

Diese Strategie beruht darauf, daß von einer unpersonalisierten Seite ausgegangen wird und man die Aktionen des Benutzer beobachtet. Aus der Auswahl der Seiten und der Verweilzeit darauf wird versucht, die Interessen des Benutzer zu identifizieren (ebenso über sonstige Aktionen wie Bookmark erstellen, lokal speichern, drucken, ...). Diese werden dann über Regeln dazu verwendet, den Inhalt persönlich zu gestalten.

##### 1.4.4.1 Vorteile

- Es werden die tatsächlichen Interessen herausgefunden und man ist nicht auf Angaben angewiesen, die (wie oben erläutert) oft nicht ganz zutreffen.
- Der Benutzer muß keinen Fragenkatalog beantworten, sondern kann gleich beginnen.
- Eine "echte" Personalisierung, da auf genau diesen einen Benutzer spezifisch eingegangen wird.
- Die Strategie lernt dauernd hinzu und kann so auf sich ändernde Interessen reagieren.

##### 1.4.4.2 Nachteile

- Da große Mengen an Daten anfallen (Klicks, Verweilzeit, ...), ist ein hoher Aufwand für die Speicherung und Verarbeitung zu betreiben.
- Eine lange Verweilzeit bedeutet noch kein besonderes Interesse: u. U. wird gerade eine Tasse Kaffee getrunken. Auch andere Aktionen können so mißdeutet werden.
- Die Informationsgewinnung erfolgt nur sehr langsam, da aus einmaligen Aktionen nicht sehr viel abgeleitet werden kann. Sie ist daher nur für regelmäßige Benutzer mit starker Interaktion geeignet.
- Insbesondere bei schon personalisiertem Inhalt kann es schwierig sein, aus den Aktionen bestimmte Vorlieben abzuleiten.
- Zu Beginn hat der Benutzer noch keinen Vorteil, da erst eine Lernphase notwendig ist.
- Eine Vorhersage über neue Gebiete ist nicht möglich.

#### 1.4.5 Fertige statistische Profile

Bei dieser Methode werden dem Benutzer nur einige wenige Fragen gestellt, die sich kaum oder gar nicht auf den eigentlichen Inhalt der Personalisierung beziehen. Aufgrund der Antworten wird der Benutzer dann in eine statistische Gruppe eingeordnet (z. B. jung, männlich, sportlich, Technikfreak). Durch diese Einordnung werden dann die von dieser Gruppe statistisch festgestellten Interessen als für den Benutzer zutreffend angenommen.

#### 1.4.5.1 Vorteile

- Man muß dem Benutzer nur einige wenige allgemeine Fragen stellen.
- Die Strategie liefert sehr vielfältige und detaillierte Interessen.
- Vorhersagen über komplett andere Sachrichtungen sind möglich.

#### 1.4.5.2 Nachteile

- Die Erstellung des Fragenkatalogs kann sehr schwierig sein, um dann eine sichere Gruppenzuordnung zu ermöglichen.
- Es sind eine enorme Menge an statistischen Daten und große Erfahrung notwendig, um Gruppen bilden zu können.
- Niemand entspricht genau dem Durchschnitt. Das Profil ist daher mit Sicherheit in manchen Punkten falsch, unter Umständen auch gerade in den interessierenden Hauptpunkten.
- Eine einmal getroffene Einordnung kann nicht geändert werden.
- Kein Eingehen auf den Benutzer als Einzelperson, d. h. kein Hinzulernen.

## 2 Marktübersicht: Programme zur Personalisierung

In diesem Kapitel wird versucht, einen Überblick über die wichtigsten derzeit am Markt befindlichen Programme bzw. Programmpakete zu geben, welche die Personalisierung von Webseiten ermöglichen. Eine vollständige Auflistung aller Programme ist wegen des Umfangs nicht möglich, da viele Produkte den einen oder anderen Aspekt von Personalisierung beinhalten. Es wurde daher eine Auswahl getroffen, die explizit auf Personalisierung ausgerichtet sind oder starke Unterstützung dafür bieten.

Die Bewertung wurde nach von den Firmen eingeholten Informationen zusammengestellt, da ein praktischer Test aller Produkte in keinem Verhältnis zum Ergebnis stehen würde. Weiters sind auch nicht von allen Applikationen Demo-Versionen erhältlich. Der Großteil der Informationen wurde über das Internet eingeholt, wobei sowohl Web-Pages wie auch E-Mail zum Einsatz kam. Bei einigen Produkten wurden auch Prospekte herangezogen. Links zu den Webseiten der Produkte finden sich im Link-Verzeichnis unter Punkt 2: Produkte zur Personalisierung von Websites.

Am Beginn steht eine Zusammenfassung der Kriterien, nach denen die verschiedenen Produkte verglichen werden, dann folgt die Beschreibung der einzelnen Programme und am Ende eine Zusammenfassung der Ergebnisse.

## 2.1 Beurteilungskriterien

Die Software wird nach folgenden Kriterien bewertet:

- 1) **Benutzererkennung:** Welche Möglichkeiten stehen zur Identifizierung der Benutzer offen? Was geschieht bei vergessenem Paßwort/verlorenem Cookie/...? Können auch anonyme Benutzer (ohne Registrierung) Empfehlungen erhalten?
- 2) **Anpassungsstrategie:** Welche der oben angeführten Strategien wird/werden verwendet, um die Präferenzen des Benutzers herauszufinden?
- 3) **Datenhaltung:** Wie werden die Daten gespeichert, die personalisiert werden? Welche zusätzlichen Programme sind dafür notwendig? Wie werden die Benutzerdaten verwaltet (Datenschutz)? Ist eine Extraktion nach verschiedenen Kriterien möglich?
- 4) **Integration in Webseiten:** Wie werden die Daten in Webseiten eingebaut? Werden Teile an bestimmte Stellen in Webseiten eingesetzt oder werden komplette Seiten nach Vorlagen konstruiert? Werden eigene Programmiersprachen verwendet?
- 5) **Kompatibilität:** Ist die Software zu vorhandenen Webservern kompatibel? Ist ein eigener Webserver integriert (kann dieser ausgewechselt werden)? Für welche Plattformen wird das System angeboten?
- 6) **Verbreitung:** Welche Verbreitung hat das Produkt? Auf welchen Websites wird es eingesetzt (URLs)?
- 7) **E-Commerce:** Ist die Software auch E-Commerce fähig bzw. ist eine Integration anderer Produkte möglich?
- 8) **Werbungspersonalisierung:** Ist es möglich, auch personalisierte Werbung zu integrieren? Welche Programme können eingebunden werden?
- 9) **Erweiterungsmöglichkeiten:** Welche Zusatzpakete gibt es? Was kann sonst noch integriert werden?
- 10) **Testmöglichkeit:** Ist ein kostenloser Test der Software möglich?
- 11) **Preis:** Wie ist die Preisgestaltung? Anmerkung: Nur bei sehr wenigen Produkten konnte der Preis festgestellt werden. Bei vielen Lösungen dürfte jedoch der Aufwand zur Integration in die eigenen Webseiten und eventuell nötige Consulting-Leistungen den Preis weit übersteigen.
- 12) **Kommentare:** Sonstige Bemerkungen

## 2.2 LearnSesame

Dies ist ein Produkt der Firma OpenSesame. Einige Informationen konnten nicht direkt erhalten werden und beziehen sich auf die Online-Demo eGenie (Siehe auch unten).

- 1) **Benutzererkennung:** Cookies, Name und Paßwort (Eingabe eines Hinweises möglich).  
Es werden auch spezielle URLs verwendet
- 2) **Anpassungsstrategie:** Die Klicks des Benutzers werden beobachtet und daraus seine Interessen abgeleitet. Es gibt die Möglichkeit des Opt-out (Ansehen und Verändern des erlernten Profils, z. B. Interessensgebiete löschen). Es werden sowohl von Personen als auch von Gruppen Profile erstellt und daraus Interessen abgeleitet.
- 3) **Datenhaltung:** Die Daten werden in einer Datenbank gespeichert (nicht enthalten). Jede ODBC-kompatible Datenbank kann verwendet werden. Auch bereits existierende Datenbanken können integriert werden.
- 4) **Integration in Webseiten:** Keine Informationen
- 5) **Kompatibilität:** Folgende Betriebssysteme werden unterstützt: Windows NT 4.0, Solaris 2.5 und höher sowie SGI Irix 6.2. An Web-Servern werden unterstützt (kein eigener integriert): Microsoft IIS, Netscape Enterprise Server. Für die dynamische Zusammenstellung der Seiten wird eines der folgenden Produkte/Komponenten benötigt: Active Server Pages, C++, C, Java, Cold Fusion, Web Objects, StoryServer, Verity Topics.
- 6) **Verbreitung:** Ericsson Employee Network, Toronto Dominion Bank, eGenie
- 7) **E-Commerce:** Keine Integration mit bestimmten Programmen vorgesehen. Eine Demo (Slideshow) ist Buch-Verkauf: Integration müßte vermutlich selbst erfolgen.
- 8) **Werbungspersonalisierung:** Folgende Programme können eingebunden werden: Net Gravity und Accipiter (extra Interface notwendig).
- 9) **Erweiterungsmöglichkeiten:** Personalisierung von Push-Channels, Interface zur Integration von Programmen für personalisierten Anzeigen.
- 10) **Testmöglichkeit:** Demos von verschiedenen Diensten mit Personalisierung<sup>6</sup>, Online-Beispiel<sup>7</sup> mit tatsächlichem Lernen (Bücher, TV, Kino, Musik, Ereignisse).
- 11) **Preis:** eGenie (fertige Lösung): Ab \$ 5000, abhängig von der Anzahl der Benutzer/Besucher; Produkt selbst: Keine Informationen
- 12) **Kommentare:** Bisher eher geringe Verbreitung.

---

<sup>6</sup> [http://www.opensesame.com/prod\\_05.html](http://www.opensesame.com/prod_05.html) (21.7.98)

<sup>7</sup> <http://egenie.opensesame.com> (21.7.98)

### 2.3 GuestTrack

Ein Produkt der Firma GuestTrack. Es sind keine besonders detaillierten Informationen auf der Website erhältlich. Weitere Informationen wurde per E-Mail angefordert.

1. **Benutzererkennung:** Erfolgt über eine persönlichen Identifikationsnummer in den URLs oder durch Cookies.
2. **Anpassungsstrategie:** Beobachtet die Clicks der Benutzer, die Personalisierung erfolgt aber nur über abgefragte Informationen (statisch). Es können jederzeit zusätzliche Fragen gestellt werden. Empfehlungen müssen über eigene Programme aus der Datenbank zusammengestellt werden.
3. **Datenhaltung:** Datenbank namens mySQL(enthaltene). ODBC Anbindung geplant.
4. **Integration in Webseiten:** Es müssen Scripts und Tags in statische Webseiten integriert werden, die andere Webseiten aufgrund der Informationen über die Person inkludieren oder erzeugen.
5. **Kompatibilität:** Die Personalisierung erfolgt über ein CGI-Interface und ist daher zu praktisch allen Webservern kompatibel. Betriebssysteme: Solaris, Linux (Windows NT geplant).
6. **Verbreitung**<sup>8</sup>: Family Success! (American Blind & Wallpaper Factory, Noel Group: Bei Besuch wurde keine Personalisierung gefunden)
7. **E-Commerce:** Selbst zu machen.
8. **Werbungspersonalisierung:** Nicht vorgesehen.
9. **Erweiterungsmöglichkeiten:** Weiteres Produkt zur Web-Veröffentlichung von Katalogen aus Datenbanken kann integriert werden.
10. **Testmöglichkeit:** Nach persönlicher Absprache
11. **Preis:** \$ 3.000 für einen Server.
12. **Kommentare:** Dies scheint ein Programm für kleinere Lösungen zu sein. Für große Websites dürfte es schwierig sein, die Personalisierung konsistent zu halten (z. B. alle Kategorien auf allen Seiten zu berücksichtigen, ...).

---

<sup>8</sup> Siehe <http://www.guesttrack.com/customers.html> (21.7.98)

## 2.4 IntelliWeb

IntelliWeb ist ein Produkt von MicroMass Communications Inc.

1. **Benutzererkennung:** Erfolgt über Cookies bzw. Einloggen mit einem Benutzernamen und einem Paßwort.
2. **Anpassungsstrategie:** Die Anpassung erfolgt regelbasiert. Diese Regeln werden in einer Datenbank gespeichert.
3. **Datenhaltung:** Die zu personalisierenden Daten sind in einer externen Datenbank gespeichert. Neue Daten werden einfach der Datenbank hinzugefügt und werden dann (den Regeln entsprechend) eingesetzt. Benutzerdaten aus existierenden Datenbanken können übernommen werden.
4. **Integration in Webseiten:** Es sind spezielle eigene HTML-Tags in die Seiten einzufügen, wo später personalisierter Inhalt dargestellt werden soll.
5. **Kompatibilität:** IntelliWeb ist als CGI-Programm konzipiert, das als Interface zu einem externen Programm dient, welches die Seiten dynamisch zusammenstellt. Folgende Betriebssysteme werden unterstützt: Irix 5.x, 6.x, Solaris 2.5, Sun OS 4.1.3u1, Windows NT Server 3.5.1, 4.0. Browser-Voraussetzung ist Netscape Navigator 3.0.1 oder Internet Explorer 3.0
6. **Verbreitung:** Nicorette CommittedQuitters<sup>9</sup>
7. **E-Commerce:** Keine Informationen
8. **Werbungspersonalisierung:** Keine Informationen
9. **Erweiterungsmöglichkeiten:** Derselbe Datensatz, der zur persönlichen Veröffentlichung im Internet verwendet wird, kann auch für Printmedien verwendet werden (IntelliPrint).
10. **Testmöglichkeit:** Online Demos<sup>10</sup> sind vorhanden, aber eher klein.
11. **Preis:** Keine Informationen
12. **Kommentare:** Professionelles Programm, das vollständig auf Regeln basiert. Kann grundsätzlich für fast alle Medien zur Personalisierung eingesetzt werden und ist auch für das WWW erhältlich. Zur Erzeugung der Daten wird MicroMass Studio benötigt.

---

<sup>9</sup> Siehe <http://www.nicorette.com/cgi-bin/committedquitters> (22.7.98, konnte nicht erreicht werden)

<sup>10</sup> Siehe <http://www.micromass.com/demos.html> (22.7.98)

## 2.5 Firefly

Firefly ist ein Produkt/Service der Firma Firefly Network Inc.

1. **Benutzererkennung:** Cookies in Verbindung mit Name und Paßwort, ein Hinweis zum Paßwort kann integriert werden.
2. **Anpassungsstrategie:** Es werden fixe Datensätze angeboten. Zusätzliche Daten können über Formulare abgefragt werden. Diese Informationen stehen zur Personalisierung zur Verfügung. Hauptmethode ist jedoch kollaboratives Filtern. Die persönlichen Interessen werden durch Bewertung herausgefunden und mit denen anderer Benutzer verglichen.
3. **Datenhaltung:** Die Daten werden in externen Datenbanken gespeichert (nicht enthalten). Benutzerdaten werden zwischen verschiedenen Anbietern, bei denen mit demselben Account eingeloggt werden kann, verschlüsselt übertragen. Benutzer können ihre Daten selbst verändern.
4. **Integration in Webseiten:** HTML-Erweiterungen und C++/Java-API
5. **Kompatibilität:** Betriebssysteme: Sun Solaris, Windows NT. Datenbanken: Microsoft SQL Server, Informix Server, Sybase Server. Web-Server: IIS, Enterprise Server.
6. **Verbreitung:** Hat relativ große Verbreitung. Seiten, die den Firefly Passport verwenden: CollegeBeat<sup>11</sup>, My Yahoo<sup>12</sup>, filmfinder: derzeit außer Betrieb, Venues: wurde eingestellt. Firmen, die FireFly Software verwenden: Barnes&Noble, myLaunch, The QuitNet, ZDNet's My Hot Files, Virtual Emporium, The Mountain Zone<sup>13</sup>.
7. **E-Commerce:** "My Wallet" ist vorgesehen für elektronische Zahlungen (Kreditkarteninformationen, ...), aber zur Zeit noch nicht verfügbar.
8. **Werbungspersonalisierung:** NetGravity AdServer 3.0.
9. **Erweiterungsmöglichkeiten:** Kann über ein Java oder C++ API programmiert werden.
10. **Testmöglichkeit:** Keine Information
11. **Preis:** Kein Informationen
12. **Kommentare:** Anfänglich sollten einige Benutzer angelegt und extensiv verwendet werden, um einen Grundstock an Bewertungen zu erreichen. Eine Einbindung in das

---

<sup>11</sup> Siehe <http://www.collegebeat.com> (22.7.98)

<sup>12</sup> Siehe <http://www.firefly.net/ywfLogin.fly> (22.7.98, Timeout)

<sup>13</sup> Links dazu siehe <http://www.firefly.net/places.fly> (22.7.98)

firefly Netzwerk ist möglich, sodaß dort registrierte Benutzer sich auch einloggen können (deren Informationen stehen dann zur Verfügung).

## 2.6 SiteServer

Ein Produkt von Microsoft Inc. aus der BackOffice Produktlinie.

1. **Benutzererkennung:** Es werden Cookies sowie Name und Passwort verwendet. Hinweise jeder Art müssen selbst programmiert werden.
2. **Anpassungsstrategie:** Es wird eine regelbasierte Anpassung verwendet, die jedoch großteils selbst programmiert werden kann. Unterstützung durch fertige Regeln und einen Regel-Manager.
3. **Datenhaltung:** Ein Zugang zu externen Datenbanken ist über ODBC möglich. Benutzerdaten werden in der internen Datenbank gespeichert. Die Zugangskontrolle kann auch über das Betriebssystem (Win NT 4.0) durchgeführt werden, wodurch sich hohe Sicherheit ergibt. Bei dieser Art der Zugangskontrolle müssen Benutzer extra angelegt werden (Für firmeninterne Gruppen gut geeignet).
4. **Integration in Webseiten:** Spezielle Metainformationen in Dokumenten zur Kennzeichnung zu personalisierender Teile. Informationen aus Benutzerprofilen werden über Active Server Pages (ASP) eingebunden.
5. **Kompatibilität:** Betriebssystem: Windows NT Server 4.0 SP3. Frontpage 98 sowie Option Pack 4.0 (enthält IIS 4.0 und ASP) wird benötigt zur Seitenerstellung. Kann nur mit Microsoft IIS eingesetzt werden (nicht enthalten).
6. **Verbreitung:** Weltweit für kleine und mittlere Unternehmen. Einsatz eigens zur Personalisierung unbekannt.
7. **E-Commerce:** Nicht enthalten (Siehe dazu SiteServer Commerce Edition)
8. **Werbungspersonalisierung:** Nicht enthalten (Siehe dazu SiteServer Commerce Edition)
9. **Erweiterungsmöglichkeiten:** Auch personalisierte E-Mails können versendet werden (integriert). Push Technik ist integriert. Microsoft SiteServer Commerce Edition ist eine integrierte E-Commerce Lösung.

10. **Testmöglichkeit:** Download einer Trial-Version (90 Tage) ist möglich<sup>14</sup>. Auch ein Beispiel zur Personalisierung kann gratis heruntergeladen werden<sup>15</sup>.
11. **Preis:** \$ 1240 für System und 5 Benutzerlizenzen, \$ 3000 für unbegrenzte Anzahl gleichzeitiger Benutzer
12. **Kommentare:** Kann nur auf einem bestimmten System eingesetzt werden, bietet dort aber hervorragende Integration. Personalisierung ist nur ein Unterpunkt, auf dem daher auch nur mittleres Gewicht liegt.

## 2.7 ObjectStore

Ein Produkt der Firma ObjectDesign Inc.

Hierbei handelt es sich nicht direkt um ein Produkt zur Personalisierung von Webseiten, sondern um eine objektorientierte Datenbank, für die es u. a. eine spezielle Erweiterung zur Personalisierung von Websites gibt. Dies bedeutet natürlich eine sehr große Flexibilität und Mächtigkeit, hat aber den Nachteil, daß viel selbst programmiert werden muß. Es sind jedoch bereits fertige Objekte vorhanden, um die wichtigsten Teile schnell und einfach zu implementieren: Login, Registrierung neuer Benutzer, Änderungen des Benutzerprofil sowie Abspeicherung der Daten in der Datenbank (Paßwörter auch verschlüsselt). Interfaces zu Java und C++ sind vorhanden. Die verwendete Datenbank besitzt spezielle Mechanismen für Caching, sodaß nach eigenen Angaben eine sehr hohe Performance erreicht wird.

Das Produkt wird von folgenden Firmen verwendet<sup>16</sup>: Time Warner Pathfinder Personal Edition, Time Warner Money, Hotmail, Excite, Global On-line Directory

Da es sich hier nicht um ein fertiges Produkt handelt, werden die verschiedenen Punkte nicht im einzelnen untersucht.

---

<sup>14</sup> Siehe <http://backoffice.microsoft.com/downtrial/default.asp?product=20> (22.7.98)

<sup>15</sup> Siehe <http://www.backoffice.microsoft.com/Dev/details.asp?cHidIndex=23> (22.7.98): Seattle a la carte

<sup>16</sup> Siehe Link-Verzeichnis und [http://www.odi.com/customer\\_stories/stories.html](http://www.odi.com/customer_stories/stories.html) (23.7.98) für weitere Kunden

## 2.8 One-To-One

Ein Produkt der Firma BroadVision.

1. **Benutzererkennung:** Erfolgt über Name und Paßwort, digitale Zertifikate oder Cookies. Identifikation erfolgt über spezielle URLs bzw. Cookies.
2. **Anpassungsstrategie:** Es werden Regeln definiert, die dann automatisch Benutzer und Daten verknüpfen. Laut Angaben kann dazu auch Fuzzy-Logik verwendet werden. Präferenzen werden über Formulare, Beobachten der Handlungen und Import existierender Daten herausgefunden.
3. **Datenhaltung:** Die Daten werden in einer externen Datenbank gespeichert (nicht enthalten). Die Übertragung von Benutzerdaten kann über SSL oder SET gesichert werden. Auswertungen müssen selbst programmiert werden.
4. **Integration in Webseiten:** Die Integration erfolgt über Templates, die wie HTML-Dokumente aufgebaut sind. In ihnen können spezielle Tags verwendet werden. Java und JavaScript können ebenfalls eingesetzt werden.
5. **Kompatibilität:** Betriebssysteme: Solaris, Windows NT, HP UX 10.20. Es kann jeder Webserver verwendet werden, der NSAPI, ISAP oder CGI unterstützt. An Datenbanken sind möglich: Oracle, Microsoft SQL Server, Sybase SQL Server, Informix
6. **Verbreitung:** Große Verbreitung, insbesondere auch bei großen Firmen: American Airlines, Argentaria, Banco Santander, Blackwells, Siemens Nixdorf, Baan Corporation, Union Bank of Switzerland, debis Systemhaus, ...
7. **E-Commerce:** Nicht enthalten. Eine Erweiterung (One-To-One Commerce) ist erhältlich.
8. **Werbungspersonalisierung:** Nicht enthalten. Keine Integration mit anderer Software. Muß separat programmiert werden.
9. **Erweiterungsmöglichkeiten:** Finanzdienstleistungen (insbesondere zwischen Banken), Knowledge Management zur innerbetrieblichen Informationsverteilung. Programmierung ist in C++ oder Java möglich.
10. **Testmöglichkeit:** Keine Informationen.
11. **Preis:** Keine Informationen.
12. **Kommentare:** Ein professionelles System, das sehr viele Möglichkeiten bietet.

## 2.9 GroupLens

Ein Produkt der Firma Net Perceptions.

GroupLens ist kein eigentliches Produkt zur Webseiten-Personalisierung, sondern ein Modul, um Empfehlungen auf Grund von sowohl expliziten als auch impliziten Bewertungen der Benutzer abzugeben. Die eigentlichen Teile für die Personalisierung müssen selbst programmiert oder zugekauft werden.

1. **Benutzererkennung:** Nicht enthalten. Muß von der Applikation erfolgen.
2. **Anpassungsstrategie:** Es wird kollaboratives Filtern verwendet: Die eigene Bewertung wird mit der anderer Personen verglichen. Zusätzlich ist auch eine Beobachtung der Klicks und der Verweildauer integriert (auch Bookmarks, drucken, speichern und scrollen wird beobachtet), um weniger Bewertungen zu benötigen. Die Ableitung von Interessen erfolgt ausschließlich durch Vergleich mit anderen Personen/Gruppen.
3. **Datenhaltung:** Die Daten werden in einer ODBC oder SQL-Datenbank gespeichert (nicht enthalten).
4. **Integration in Webseiten:** Nicht enthalten.
5. **Kompatibilität:** Betriebssysteme: Windows NT 4.0, Sun Solaris 2.5.1. An Datenbanken können Microsoft SQL Server, Oracle und ODBC Datenbanken verwendet werden. Kann direkt in Microsoft Site Server, Broadvision One-To-One und Vignette's Story Server integriert werden.
6. **Verbreitung:** Mittlere Verbreitung: CD-Beat, Takeoff Travel Service, Net By-Line, Amazon.com, E-Online, Sierra On-Line.
7. **E-Commerce:** Nicht enthalten.
8. **Werbungspersonalisierung:** Kann auch dafür eingesetzt werden.
9. **Erweiterungsmöglichkeiten:** Java, COM, CGI (Perl), C/C++ APIs stehen zur Programmierung zur Verfügung.
10. **Testmöglichkeit:** Keine Informationen.
11. **Preis:** Keine Informationen.
12. **Kommentare:** Spezielles Paket, um andere Produkte um kollaboratives Filtern zu erweitern, vom Erfinder dieser Technik.

## 2.10 Dynamo Profile Station

Ein Produkt der Firma Art Technology Group (ATG). Es basiert auf dem Dynamo Application Server und ist komplett in Java geschrieben.

1. **Benutzererkennung:** Erfolgt über Cookies oder URLs. Eine Einbindung von bestehenden Datenbanken ist möglich. Informationen von anonymen Benutzern, die sich registrieren lassen, werden übernommen
2. **Anpassungsstrategie:** Es werden sowohl explizite als auch implizite Interessen zur Personalisierung verwendet, wobei regelbasiert vorgegangen wird. Andere Strategien können selbst programmiert werden. Die Zusammenfassung von Benutzern nach demographischen Profilen ist möglich.
3. **Datenhaltung:** Die Daten werden in einer externen Datenbank gespeichert (Eine Klein-Datenbank zum Testen ist enthalten. Für größere Projekte ist eine kommerzielle DB notwendig). Die Benutzerdaten können auch für mehrere Websites verwendet werden.
4. **Integration in Webseiten:** Erfolgt durch Einfügen von Java-Code oder Scripts in JHTML-Seiten. Die Compilierung des eingebetteten Java-Codes erfolgt erst bei Bedarf.
5. **Kompatibilität:** Betriebssysteme: Solaris 2.5, 2.6 sowie Windows NT 4.0. An Webservern werden unterstützt: Netscape Enterprise, Netscape FastTrack, Microsoft IIS, Apache sowie Servlet-API kompatible Server (ein einfacher Server zum Testen ist enthalten). An Datenbanken werden unterstützt (über JDBC): Oracle SQL Server 7.3, Sybase SQL Server 11.0.2 und Microsoft SQL Server 6.5.
6. **Verbreitung:** Größere Verbreitung<sup>17</sup>: Stream International, Harvard Business School, NTT Data, JavaSoft.
7. **E-Commerce:** Nicht enthalten. Add-on (Retail Station) ermöglicht eine Integration.
8. **Werbungspersonalisierung:** Nicht enthalten (Siehe Add-on Dynamo Ad Station).
9. **Erweiterungsmöglichkeiten:** Ein Java-API zur Programmierung beliebiger Erweiterungen steht zur Verfügung.
10. **Testmöglichkeit:** Kostenloser Test möglich (Anmeldung bei ATG notwendig). Eine Demo einer fiktiven Website<sup>18</sup> kann besichtigt werden.

---

<sup>17</sup> Siehe auch: <http://www.atg.com/solutions/customers> (27.7.98)

11. **Preis:** Keine Informationen.

12. **Kommentare:** Sehr ausführliche Online-Dokumentation vorhanden. Ein professionelles System, das für alle Arten von Web-Applikationen geeignet ist.

### 2.11 Story Server

Ein Produkt der Firma Vignette.

1. **Benutzererkennung:** Erfolgt über Cookies oder URLs.
2. **Anpassungsstrategie:** Für die Anpassung wird GroupLens verwendet (Siehe 2.9). Auch explizite Angabe von Interessen und Anpassung ist möglich.
3. **Datenhaltung:** Die Daten sind in einer externen Datenbank gespeichert (nicht enthalten).
4. **Integration in Webseiten:** Komponentenbasiert: z. B. Navigationselemente, die für jeden Benutzer speziell zusammengestellt werden.
5. **Kompatibilität:** Betriebssysteme: Solaris 2.5.1, Windows NT Server 4.0 SP3. Folgende Webserver werden unterstützt (keiner enthalten): Netscape Server, Enterprise Server, FastTrack Server, Apache, IIS. An Datenbanken können verwendet werden: Sybase, Oracle, Informix, MS SQL.
6. **Verbreitung:** Große Verbreitung<sup>19</sup>: Bay Networks, Chicago Tribune, Citibank, CNET, Lycos, Macromedia, National Semiconductor, Playboy, Time Warner, MovieFinder.
7. **E-Commerce:** Kann auch programmiert werden (teilweise unterstützt).
8. **Werbungspersonalisierung:** Ist teilweise integriert, bzw. selbst zu programmieren.
9. **Erweiterungsmöglichkeiten:** C und C++ API für die Programmierung eigener Erweiterungen oder als Interface zu anderen Produkten.
10. **Testmöglichkeit:** Keine Informationen.
11. **Preis:** Keine Informationen.
12. **Kommentare:** Bietet starke Unterstützung für Workflow bei der Entwicklung. Automatische Anpassung an Fähigkeiten der benutzten Browser-Software.

---

<sup>18</sup> Siehe <http://www.atg.com/share/walkthrus/ad-n-profile/p1.html> (27.7.98) megatravel.com. Bietet zwar keine echte Website, zeigt aber auch den Hintergrund (Administration).

<sup>19</sup> Für Gesamtliste siehe <http://www.vignette.com/Customers> (23.7.98)

## 2.12 WebSell

Ein Produkt der Firma LikeMinds.

1. **Benutzererkennung:** Erfolgt über Cookies. Auch anonyme Benutzer können Empfehlungen erhalten.
2. **Anpassungsstrategie:** Es werden die Interessen des Benutzers (Bewertung von einzelnen Elementen, Beobachtung oder Import aus existierenden Datenbanken) mit denen anderer verglichen und daraus weitere mögliche Interessen abgeleitet. Es werden keine Gruppen gebildet, sondern möglichst passende Einzelpersonen gesucht.
3. **Datenhaltung:** Die Daten werden in einer ODBC-kompatiblen externen Datenbank gespeichert (nicht enthalten).
4. **Integration in Webseiten:** CGI- und JavaScript Module werden die in die Webseiten eingefügt.
5. **Kompatibilität:** Betriebssysteme: Solaris, IRIX, Windows NT, Linux, ... Das Produkt kann mit jedem Server eingesetzt werden, der ein CGI oder FastCGI-Interface besitzt. Darüber hinaus sind direkte Interfaces für Netscape Enterprise Server und Microsoft Site Server vorhanden.
6. **Verbreitung:** Geringe Verbreitung (für Personalisierung): CineMax (die anderen Angaben waren nicht erreichbar oder es konnte keine Personalisierung gefunden werden)
7. **E-Commerce:** Die Personalisierung ist nur ein Teil eines E-Commerce Paketes.
8. **Werbungspersonalisierung:** Ist enthalten.
9. **Erweiterungsmöglichkeiten:** Ein C++ API ermöglicht eigene Erweiterungen.
10. **Testmöglichkeit:** Online-Demo-Website: MovieCritic<sup>20</sup>
11. **Preis:** Typischerweise zwischen \$50.000 und \$ 100.000.
12. **Kommentare:** Das Produkt ist ausschließlich auf kollaboratives Filtern ausgelegt. Eine der ersten Firmen die dies entwickelten (besitzt auch 2 Patente dazu).

---

<sup>20</sup> Siehe <http://www.moviecritic.com> (27.7.98)

### 2.13 WiseWire for WebSites

Ein Produkt der Firma WiseWire. Dieses Programm sucht eigenständig das Internet nach neuen Informationen ab und kategorisiert diese automatisch (und speichert sie lokal). Diese Kategorien werden als sogenannte "Wires" angeboten, wobei auch private Firmen-Wires möglich sind.

1. **Benutzererkennung:** Erfolgt durch Name und Paßwort oder Cookie.
2. **Anpassungsstrategie:** Es wird kollaboratives Filtern verwendet. Durch die Bewertung werden die Dokumente gereiht. Die Bewertung durch alle Benutzer wird anschließend durch die persönliche Bewertung verfeinert, was bedeutet, daß ein Eingehen auf Interessen einer einzelnen Person eher in geringem Maße erfolgt.
3. **Datenhaltung:** Eine externe Datenbank zur Speicherung der Artikel wird benötigt (nicht enthalten).
4. **Integration in Webseiten:** Keine Informationen.
5. **Kompatibilität:** Keine genauen Angaben: "Hauptsächlich" Windows NT Betriebssystem, Oracle Datenbank, Netscape Server und andere.
6. **Verbreitung:** Geringe Verbreitung<sup>21</sup> als eigenständiger Service: Lycos Webguides. Mittlere Verbreitung als Provider.
7. **E-Commerce:** Keine Informationen.
8. **Werbungspersonalisierung:** Ist integriert (Accipiter AdEngine and AdManager sind integriert).
9. **Erweiterungsmöglichkeiten:** Keine Informationen.
10. **Testmöglichkeit:** Einige Wires können online besichtigt werden<sup>22</sup>, wobei bei einigen die Artikel selbst aber nur mit Paßwort zugänglich sind.
11. **Preis:** Keine Informationen.
12. **Kommentare:** Nicht für alle Gebiete der Personalisierung geeignet, da sehr stark auf Nachrichten ausgerichtet.

---

<sup>21</sup> Siehe <http://www.wisewire-corp.com/indexpart.html> (27.7.98)

<sup>22</sup> Siehe <http://www.wisewire-corp.com/indexsample.html> (27.7.98) für eine Liste

## 2.14 AgentWare

Ein Produkt der Firma Autonomy. AgentWare ist nur als Nebenprodukt zur Personalisierung einsetzbar. Der Hauptaugenmerk liegt auf der selbständigen Klassifizierung von Dokumenten nach von Agenten automatisch identifizierten Konzepten.

1. **Benutzererkennung:** Erfolgt durch Log-In oder Cookies. Auch für anonyme Benutzer sind Empfehlungen möglich.
2. **Anpassungsstrategie:** Der Benutzer wird beobachtet und die besuchten Seiten genau untersucht. Anschließend werden Seiten mit ähnlichen Konzepten angeboten.
3. **Datenhaltung:** Keine Informationen.
4. **Integration in Webseiten:** Keine Informationen.
5. **Kompatibilität:** Betriebssysteme: SUN Solaris, Windows NT, DEC UNIX.
6. **Verbreitung:** Geringere Verbreitung in Bezug auf Personalisierung<sup>23</sup>: LineOne, Macmillan. Große Verbreitung im Bereich Knowledge Management.
7. **E-Commerce:** Nicht enthalten.
8. **Werbungspersonalisierung:** Nicht enthalten.
9. **Erweiterungsmöglichkeiten:** Einige Zusätze erhältlich: Automatische Zensur, Nachrichten-Zusammensteller, Benachrichtigung bei neuesten Informationen bestimmter Art, Groupware Unterstützung.
10. **Testmöglichkeit:** Demos zur Kategorisierung<sup>24</sup> sind Online verfügbar.
11. **Preis:** Keine Informationen.
12. **Kommentare:** Nur bedingt zur Personalisierung zu verwenden. Die Stärke liegt in der automatischen Klassifizierung, daher z. B. gut für die Suche in großen Archiven nach einem bestimmten Thema oder Zusammenstellung von Nachrichten zu einem Thema aus dem Internet geeignet.

---

<sup>23</sup> Siehe <http://www.agentware.com/main/cases/index.html> (27.7.98)

<sup>24</sup> Siehe <http://ultra.agentware.com/KnowledgeSuite> (27.7.98)

## 2.15 Zusammenfassung

Die einzelnen Programme lassen sich grob in vier Kategorien einteilen:

- 1) **Basisprodukte:** Einfachere Programme, die für kleine Firmen geeignet sind und keinen großen Aufwand bedeuten, aber nicht besonders gut erweiterbar sind und reduzierte Funktionalität bzw. Erweiterungsmöglichkeiten bieten.
- 2) **Personalisierungssoftware:** Programme für mittlere bis große Aufgaben, die oft aus dem Bereich des Electronic Commerce kommen. Sie sind gut für Personalisierung geeignet und enthalten viele fertige Bausteine, die einfach anzupassen sind. Sie bieten aber wenig Möglichkeiten für spezielle Wünsche, welche selbst zu programmieren sind.
- 3) **Web-Applikationen:** Programmpakete, die für große Programme im Web geeignet sind. Sie sind nicht direkt für Personalisierung entworfen, bieten aber auch dafür Unterstützung. Sie sind für den High-End Bereich sehr gut geeignet, bedeuten aber einen größeren Programmieraufwand und können meist in alle beliebigen Richtungen erweitert werden, wenn Bedarf besteht.
- 4) **Sonstige:** In dieser Kategorie werden sonstige Produkte eingeordnet, die aus irgendeinem Grund nicht in die obigen Kategorien hineinpassen (z. B. Zusatzprodukte).

Die untersuchten Programme können folgendermaßen eingeteilt werden, wobei die Abgrenzung zwischen Personalisierungssoftware und Web-Applikationen u. U. nicht immer ganz klar ist:

- 1) **Basisprodukte:** GuestTrack
- 2) **Personalisierungssoftware:** IntelliWeb, FireFly, SiteServer, WebSell
- 3) **Web-Applikationen:** LearnSesame , One-To-One, Dynamo Profile Station, StoryServer
- 4) **Sonstige:** ObjectStore (Datenbank+Personalisierungs-Erweiterung), GroupLens (Plug-In für kollaboratives Filtern), WiseWire (Dokumenten-Klassifizierung), AgentWare (Dokumenten-Klassifizierung)

Je nach den verfolgten Zielen sollte daher ein Produkt aus der entsprechenden Kategorie ausgewählt werden. Innerhalb der Kategorie ist dann besonders auf die bestehende Hard- und Software-Umgebung sowie Programmier- bzw. Webseitenerstellungs-Erfahrung Rücksicht zu nehmen.