Internet - Ökonomie

Portale, Push- Technologie, Personalisierung

Michael Augustin

1. September 2003

Inhaltsverzeichnis

Kapitel 1

Einleitung

Die Entwicklung das World Wide Web als Informationsmedium für Konsumenten von Tageszeitungen, Fernsehn und anderen Medien anzubieten, brachte seit Mitte der neunziger Jahre der Wirtschaft eine völlig neue Möglichkeit, sich zu entfalten. Im Vordergrund steht dabei natürlich auch immer noch der Aspekt Gewinne zu maximieren.
So ist es auch nicht verwunderlich, dass Firmen in diesem Metier möglichst jedes nur erdenkliche Mittel einsetzen um Kunden zu gewinnen und an sich zu binden. Gerade ein im wirtschaftlichen Sinne noch weitesgehend unerschlossenes Gebiet, bietet nicht nur großen erfahrenen Firmen Entfaltungsmöglichkeiten, sondern auch Neulingen. Die lukrativen Bedingungen über das Netz etwas anzubieten und möglicher Weise sogar zu verkaufen, ließen Firmen aus dem Boden schießen, die ohne große Konzepte Marktlücken füllten.
Betrachtet man die vielen angebotenen Dienstleistungen im Internet, so wird man zwangsläufig feststellen, dass es teilweise derart verrückte Ideen gibt, für die wohl kaum eine Bank der Welt jemals einen Kredit gewähren würde, wenn es sich hierbei um eine Existenzgründung handeln würde.
Um so kurioser ist jedoch, dass immer wieder, und zwar gerade in der Vergangenheit, die Möglichkeit bestand, unheimlich viel Geld auf diesem Weg zu verdienen. Fakt ist allerdings, dass auch auf diesem Markt nur jenes Unternehmen überlebt, welches mit der Zeit geht und offen für Neues ist. Dieser Punkt ist von überaus großer Relevanz, denn ein Markt der so schnellebig und vielfältig ist wie das World Wide Web, kann selbst für Experten nicht langfristig überschaubar sein. Denn wäre dies der Fall, so hätten sich Pleiten von millionenschweren Firmen vermeiden lassen.
Also müssen ständig neue Strategien und Technologien entwickelt werden, die auch weiterhin das Überleben von Firmen sichern und dazu dienen auf die Kunden innerhalb der virtuellen Welt besser eingehen zu können. Es

müssen neue Überlegungen getroffen werden, die beispielsweise beinhalten, wie Kunden zu erreichen sind, mit denen man kein Verkaufsgespräch mehr von Angesicht zu Angesicht hat, ohne sie zu verschrecken.
Im Folgenden geht es um Technologien, welche für das Internet entwickelt wurden, um Benutzer gezielt mit Informationen zu versorgen und zu beeinflussen. Es werden aber auch Verfahren vorgestellt, die ziemlich genaue Kundenprofile und spezielle Einblicke in das Kundenverhalten bringen. Zum Teil werden Zusammenhänge in den Entwicklungen betrachtet und Vor- und Nachteile ausgewertet.

Kapitel 2

Webportale

2.1 Das Tor Zum WWW

Das World Wide Web stellt den Nutzern mitlerweile ein imenses Wissensrepertoire zur Verfügung. Doch fehlt es innerhalb dieses Informationsmediums an Strukturierung und Organisation. Es gibt keine zentrale Stelle im Netz, welche Veröffentlichungen kontrolliert. Niemand kann also wirklich nachvollziehen, wann im Netz welche Daten beispielsweise gelöscht oder aktualisiert werden. Weitere Probleme liegen in der Datengruppierung. Sämtliche Arten von Daten werden gleichberechtigt im Netz angeordnet. Dies erschwert natürlich das Auffinden spezifischer Information sehr. Abhilfe sollten bereits in den frühen neunziger Jahren sogenannte Suchmaschinen bringen, welche durch einen immer größer werdenden Nutzerkreis, dessen Ansprüche ständig stiegen, zu Plattformen heranwuchsen, die verschiedenste Dienste auf einer Seite vereinigten. Denn selbst für versierte Webbenutzer ist auch heute noch eine Art Anlauf- und Rückkerhstelle im Netz sehr dienlich um einen gewissen Überblick über die Vielfalt des Informationsangebotes und der Webservices im Internet zu bewahren. Seit 1998 hat sich dann für derartige Seiten der Begriff Web- Portal herausgebildet. Der grundlegende Aufbau dieser Portale, besteht aus dem Namen des jeweiligen Portalanbieters, der Werbung, einem Suchelement, einem Navigationsteil für das breit gefächerte Informationsangebot von Nachrichten, Börsen- und Wetterberichten oder Informationen aus anderen Bereichen, Webservices wie E-Mail, SMS-Diensten oder Chaträumen und der Möglichkeit das Portal auf persönliche Belange zuschneiden zu können.
Auf Grund der unterschiedlichen Klientelen die solche Seiten ansprechen sollen, ist es im Allgemeinen etwas schwierig eine Kategoriesierung der Web-

Portale vorzunehmen. So kann man die Portale einerseits nach Informationsinhalt einteilen, wie beispielsweise medizinische Webportale oder Unternehmensportale, allerdings unterscheiden sich die einzelnen Portale der jeweiligen Bereiche auch wieder mindestens im Aufbau oder ihrem Einsatzgebiet.
Um ein Beispiel für eine mögliche Unterteilung zu liefern, soll folgende kurze Differenzierung der Busines to Consumer Portale ausreichen. Im Großen und Ganzen haben diese Web- Portale ihren inhaltlichen Kern auf das Themengebiet E-Commerce ausgerichtet. Allerdings ist beim Aufbau einer solchen Seite zu berücksichtigen, welcher Nutzerkreis angesprochen werden soll um eine möglichst intuitiv bedienbare Oberfläche zu schaffen, die wenig unwichtige Information für den Benutzer enthält. Die Consumer- Portale haben das Ziel möglichst jeden Benutzer des Internets anzusprechen, um ihm Informationen und Funktionen des World Wide Webs an einem zentralen Punkt anzubieten. Diese Portale sollen eine kostenlose Hilfe für den Einstieg, die Weiterführung und die Orientierung im Netz darstellen. Die Enterprise- Information- Portale, welche auch auch als Vertical oder Corporate Portale bekannt sind, sind ebenso wie die Consumer- Portale auf einen offenen Nutzerkreis zugeschnitten. Sie dienen als Einstiegsseite für die Web- Seiten eines Unternehmens, wobei die unternehmensspezifische Ausrichtung mit sich zieht, dass fast ausschließlich Informationen über eine bestimmte Firma angeboten werden. Beim Aufbau eines Extranet- Portals stehen Busines to Busines Aktivitäten im Vordergrund, welche potentiell kooperierenden Geschäftskunden eines Unternehmens beispielsweise Bestell- und Liefervorgänge vereinfachen sollen. Intranet- Portale beziehen sich ebenso wie die Extranet- Portale auf einenen geschlossenen Nutzerkreis. Sie sollen den Mitarbeitern eines Unternehmens einen konsistenten Blick auf das Unternehmen bieten, indem diese einen direkten und personalisierten Zugriff auf Unternehmensapplikationen und Intranetinhalte haben. Bei derartigen Portalen wurde das Portalkonzept, dass sich für den Zugang zur Informationsviefalt des World Wide Web bewährt hat, auf das Unternehmen übertragen.

2.2 Ein Wirtschaftseinblick

Es ist aber auch für jeden Portalanbieter, egal wen er ansprechen möchte, die Frage zu klären, wie er an die jeweiligen Zielgruppen herankommt. Zielgruppen sind hierbei Partner und Nutzer eines Portals, denn diese stehen für den Anbieter aus rein wirtschaftlicher Natur in unmittelbarem Zusammenhang. Betrachtet man die Consumer- Portale, so agieren Partner in diesem Geschäft

in der Art und Weise, dass sie die Anbieter von Dienstleistungen und Produkten sind und Entgelder zur Finanzierung der Portale bereitstellen. Da die Portale jedoch auch nur ein Werbeträger von vielen auf dem Markt darstellen, müssen die Portalanbieter sich um eine hohe Frequentierung ihrer Seiten kümmern. Denn meistens bestimmt die Anzahl der Besucher die jeweiligen Marktchancen eines Web- Portals. Wer hohe Nutzerzahlen vorweisen kann, hat wie es im Business beinahe generell der Fall ist, auch die Möglichkeit höhere Preise im Bezug auf die Bereitstellung von Platz für Werbung und Dienste für sich zu beanspruchen. Um dauerhaft Konkurenz für andere, oder eventuell auch ähnliche Portalbetreiber zu sein, ist ein sehr gutes Knowhow im Bereich Marketing unverzichtbar. Dazu gehört, dass jeder Web- Portalanbieter in den Punkten Product, Place, Price und Promotion immer up-to-date sein muss, da in dieser Branche viel kopiert und unterboten wird. Er muss sich auch zwangsfäufig mit der Auswertung Nutzerspezifischer Daten auseinandersetzen wobei ihm Techniken wie das Data Minig von großer Hilfe sein können. Denn nur so hat der Anbieter die Chance, seinen Platz im Internet zu einem attraktiven Werbemedium auszubauen, dessen Name positive Assoziationen hervorruft.
Doch auf Grund der Dynamik und der starken und hohen Konkurenz im Netz, gelang es nicht vielen Anbietern ständig attraktiv für die Kunden im Internet zu bleiben und so entwickelte sich die einst so wunderbare Idee zu einer von vielen im World Wide Web.

Kapitel 3

Push - Technologie

3.1 Funktionsweise

Eine weitere Erfindung, die das Auffinden von Wissen im Internet vereinfachen sollte, war die Push- Technologie, die als Pendant zu Pull 1997 einen großen Erfolg versprach. Vorreiter in der Entwicklung auf diesem Gebiet waren Unternehmen wie Individual Inc. und Crayon. Auch bei der Schaffung dieser Neuheit war man sich dessen bewusst, dass es für die Nutzer fortlaufend schwieriger werden würde, an relevante Informationen im Netz heranzukommen, da die Menge an Informationen, die im World Wide Web abgelegt wurden enorm wuchsen. Der Ansatz der Push- Technologie lag allerdings darin, dem Benutzer ähnlich dem Abonementmodell der Post die gewünschten Informationen automatisiert zukommen zu lassen, so dass der Nutzer eine eher passive Rolle bei der Suche nach den gewünschten Inhalten einnehmen konnte. Der Vorteil dieser Technologie gegenüber dem herkömmlichen Verfahren besteht in der Möglichkeit, die Zustellung der Informationen weitaus dynamischer gestalten zu können. Im Gegensatz zur Post kann der Verbraucher mehrmals am Tag aktuelle Informationen bekommen, wohingegen die Post, im Speziellen bei der Zeitungszustellung nur einmal pro Tag die gewünschten Informationen liefert, denn dem Nutzer können so lange Daten zugesendet werden, wie er online ist.
Das Funktionsprinzip der Push- Technologie besteht aus wenigen einfachen Punkten, die nachfolgend kurz beschrieben werden. Der Benutzer meldet sich bei einem Anbieter oder Speziellen Informationskanal an, indem er sein Informationsprofil an diesen schick. In dem Profil ist festgelegt, wann der Anbieter senden kann und welche Inhalte er dem Benutzer zukommen lassen soll. Anhand des vorgegebenen Zeitplans verbindet sich der User-PC dann zum

Server im Internet und holt sich von dort seine Informationen. Diese Daten werden direkt auf dem PC gespeichert und wenn es vom Nutzer erwünscht ist wird dieser auch umgehend über die neu eingegangenen Informationen benachrichtigt.

3.2 Unicast

Zu Beginn lag der Hauptanwendungsbereich dieser Techologie darin, per E-Mail Newsletter zu verteilen. Doch nach und nach wurden die zu verschickenden Daten weitaus größer. So begann man neben Textdateien auch Multimediadateien oder ganze Softwarepakete über das Netz zu senden. Diese Informationsübersendung wurde zunächst per Unicastverfahren realisiert, welches jedoch TCP - basierend arbeitet und viel Bandbreite benötigt. Hierbei ist es nötig, dass zu jedem Nutzer eine eigene Verbindung aufgebaut ist, selbst wenn identische Anforderungen bestehen. Die Notwendigkeit dieser einzelnen Verbindungen liegt bei diesem Verfahren darin, dass dem Anbieter mitgeteilt werden muss, dass der User- PC bereit ist Daten zu empfangen. Zur Realisierung dieses Verfahrens, wurden von Firmen wie Pointcast, Netscape oder Microsoft sogenannte Push- Applikationen entwickelt, die dem Benutzer eine übersichtliche Oberfläche bieten sollten, mit der sie alle nötigen Einstellungen vornehmen oder die empfangenen Daten lesen konnten. Die Arbeitsweise dieser Applikationen unterscheidet sich in der Struktur der Informationszustellung der Anbieter, welche sich in drei Modelle gliedern lassen. Dazu gehören das Push- Server- Modell, das Webserver- Extention- Modell, und das Client- Agent- Modell.
Beim Push- Server- Modell steht ein eigener Push- Server im Netz zur Verfügung, auf den durch Clientprogramme oder spezielle Protokolle vom User- PC aus zugegriffen werden kann. Den Clientprogrammen, beispielsweise Satelite bei Windows 98, wird duch den Benutzer eingegeben, welche Informationen angefordert werden. Dieses Profil wird an den Push- Server weitergeleitet, dieser sucht in bestimmten Informationskanälen nach den gewünschten Inhalten und schickt diese an den Client. Das Webserver- Extention- Modell benutzt keinen speziellen Push- Server und Feedback und Profildaten werden auf einem externen Server dauerhaft abgelegt, von wo aus sie vom Anwender eingesehen und verwaltet werden können. Für derartige Anwendungen reicht meist ein ganz normaler Webbrowser als Clientprogramm aus, um die geforderten Informationen zu bekommen. Beispiel für dieses Modell sind jene Webportale, die dem Benutzer die Möglichkeit zur Personalisierung bieten.

Bei diesem Beispiel wird ein Profil an den Portalanbieter gesendet um dem Benutzer bei jedem Aufruf der Seite die im Profil eingestellte Information zu zeigen. Das Client- Agent- Modell nutzt einen Server zum verwalten von Profilddaten und Updates. Der Benutzer gibt mit der Übersendung eines Profils dem Server vor, welche Art von Information gesuch wird. Der Server hat lediglich die Information darüber, wo die gesuchten Inhalte im Netz zu finden sind und sendet dem Push- Client auf dem User- PC alle zu der Anfrage passenden Verweise zu. Content- Agents holen sich dann Informationen von den zugeschickten Adressen im Netz, wobei jeder Agent unterschiedliche Ergebnisse liefert. Dieses Modell ist ähnlich der Arbeitsweise einer Suchmaschine Im Internet und schafft ein anonymes Verhältnis zwischen Benutzer und Anbieter. Denn der Benutzer kontrolliert den Einsatz und bestimmt den Umfang einer solchen Applikation.

3.3 Multicast

Einen Fortschritt auf dem Gebiet der Push- Technologie brachte die Firma Starbust, mit der Entwicklung des Multicast File Transfer Protokoll (MFTP), welches auf UDP- Ebene arbeiten kann, somit weniger Bandbreite benötigt und einem weitaus empfängerbasierterem Konzept nachgeht. Beim Multicastverfahren werden die Empfänger für eine spezielle Multicastsitzung auf einem Server eingetragen und erhalten daraufhin ständig die Daten dieser Gruppe durch eine Netzwerkinfrastruktur. Dabei ist es nicht nötig, dass der Sender der Daten eine Liste der Empfänger besitzt, denn dieser schickt nur eine Kopie über das Netzwerk, welche beim Routing dann so oft wie nötig verfielfacht wird. Die Anzahl der neuen Kopien wird bestimmt, indem der weiterleitende Router einen Spannbaum zu allen multicastfähigen Routern erzeugt, die einen oder mehrere Hosts beliefern. Ob die Hosts noch zu einer bestimmten Gruppe gehören wird regelmaäßig geprüft um inaktive Zweige des Spannbaums streichen zu können. Dies vermeidet redundanten Datenverkehr und schont die Bandbreite. Bekommt der Server eine negative Empfangsbestätigung, so sendet er so lange die Daten an den Empfänger, bis eine positive Rückmeldung erfolgt. Ein weiterer Vorteil des Multicastverfahrens liegt darin, dass jeder Intranet- Administrator selbst für eine derartige Fähigkeit seines Netzes sorgen kann. Dadurch können beispielsweise inerhalb eines Firmenintranets eigene Informationskanäle angeboten werden, die den Mitarbeitern das auffinden von Informationen über die Konkurenz abnehmen, Details zu neuen Produkten liefern oder aber auch Geschäftsberichte

zusenden.
Da allerdings der Einsatz der Push- Technologie auch zu einer gewissen Informationsüberlutung beim Nutzer entartete, indem Unarten aufkamen wie das unaufgeforderte Zusenden von E- Mails (Spam) oder Benachrichtigungen über Änderungen von Webseiten mittels Popups, ging die große Zeit des Aufstiegs dieser Technologie auch wieder vorüber. Von vielen abgestempelt als Bandbreitenfresser nutzt man allerdings noch heute selbst in großen Unternehmen ein vom Grudaufbau her gleiches Verfahren zur Übersendung von Updates, Teten, Video- und Audiodaten.

Kapitel 4

Personalisierung

4.1 Der gläserne Kunde

Wenn ein Online- Anbieter seinen Kunden binden, oder gar seinen Umsatz am Kunden steigern möchte, muss er immer einen gewissen Point of Interest für diesen darstellen. Ein solches Ziel ist für den Anbieter nur dann erreichbar, wenn er dem Kunden gefällt und auf dessen Bedürfnisse eingeht, ohne ihn durch zu viele Verpflichtungen und Fragen zu bedrängen. Doch um zu erfahren, welche Vorlieben jeder individuelle Nutzer hat, ist es nötig Marktforschungen vorzunehmen und Kundenanalyse zu betreiben. Dies beinhaltet die Erstellung von Datensätzen über bestimmte Kundenkreise, bis hin zu deren Auswertung.
Im Bereich der Web- Portale werden Profile von Kunden auf sehr unterschiedliche Art und Weise erstellt. Es kommt lediglich darauf an, welche Information der Portalbetreiber vom Kunden haben möchte. Der Einsatz von Cookies beispielsweise bietet den Portalanbietern vor allem Aufschluss über das Verhalten ihrer Kunden und ist für den Benutzer sehr diskret und unkompliziert. Dieses Verfahren gibt einem Anbieter zunächst einmal Aufschluss über die häufigsten Messparameter für die Erfolgskontrolle eines Portals. Dazu gehören Messdaten die Eindruck über die Dauer eines Besuches und die Angebotsnutzung geben. Diese werden durch die Summe der Sichtkontakte beliebiger Nutzer als Page Views angegeben. Weitere Parameter sind die Visits einer Seite, sie drücken die Anzahl der Zugriffe von Web- Seiten ausserhalb des eigenen Angebotes aus. Die Zahl der AdClicks liefert die Häufigkeit der Clicks auf ein werbetragendes Objekt und gilt seit 1998 als Standardonlinewerbewährung. AdImpressions zeigen die Anzahl der Sichtkontakte von Nutzern auf Werbemittel zu bestimmten Tageszeiten im Zu-

sammenhang mit dem Nutzerverhalten bei einem Sichtkontakt. AdRequests messen, ob konkrete Abfragen eines Werbemittels auf einem Werbeserver existieren. Sie spiegeln nicht nur den Click zu der gesamten großen Seite wieder. Die ViewTime bezeichnet jene Zeit, die ein Potentiell werbeführender Teil eines Internetangebotes während eines Nutzervorganges sichtbar ist. Die Cost per AdClick stellt die Werbekosten im Bezug zur Clickzahl dar. Weitere Parameter wie Cost per Order und Cost per Customer lassen sich ähnlich berechen.
Andere Verfahren wie das Bereitstellen von Fragebögen, die durch den Benutzer freiwillig ausgefüllt werden können, geben eher Aufschluss über persönliche Daten, die allerdings nicht immer einhundertprozentig korrekt und vollständig sind. So geben durchschnittlich etwa 90 Prozent ihre Hobbys, 89 Prozent ihr Alter, 67 Prozent ihren Namen, 29 Prozent ihr Gehalt aber nur 4 Prozent ihre Kreditkartennummer an. Der Erfolg der Informationssuche auf diese Weise hängt also von der Art der Daten ab, die der Anbieter von seinem Kunden wünscht.
Mit Hilfe der Auswertung solcher Zahlen, kann ein Portalbetreiber nachweislich durch intelligentes Schalten von Werbung seinen Umsatz steigern. Die Werbebranche bedient sich schon seit langer Zeit einiger Weisheiten der Psychologie und so ist es auch nicht verwunderlich, dass die Onlinewerbung mit sehr geschickten Mitteln eine individuelle, dynamische und gewinnbringende Beeinflussung auf das Kaufverhalten des Kunden ausübt. Diese Art von Beeinflussung geschieht beispielsweise durch Werbebanner. Dabei ist zwischen einfachen und stichwortsensitiven Bannern zu unterscheiden. Die einfachen Banner bedürfen meist einer guten Platzierung auf der werbeschaltenden Seite und einem passenden Layout, wohingegen stichwortsensitive Werbebanner eher durch unauffällige Botschaften bestechen, sobald ein Schlagwort von Benutzer eingegeben oder ein bestimmter link aufgerufen wird.

4.2 Der Personalisierungsprozess

Die Personalisierung wird in der Praxis in zwei Prozesse eingeteilt. Dies ist zum Einen das Profiling, welches die Aquisition und Benutzermodellierung beinhaltet und zum Anderen das Match Making, was sich mit der Auswertung der gewonnenen Daten und der Reaktion auf die daraus folgende Informationsgewinnung beschäftigt.
Beim Profiling werden explizite Daten, wie Postanschrift, Name und Alter und implizite Daten wie Click- Streams und Verweilzeiten mit den bereits

beschriebenen Verfahren gesammelt und danach durch beispielsweise dynamisches Filtering verwendet um Affinitäts- und Präferenzengemeinschaften aufzubauen, die dann dazu führen sollen, automatisiert Beratung und Empfehlung anbieten zu können. Statische Verfahren haben den Nachteil, dass nicht sofort auf den Benutzer eingegangen werden kann. Sie führen aber im Allgemeinen zu einem weitaus größeren Informationsgewinn über den Kunden. Beispiel für das statische Match Making ist das Data Mining, welches im folgenden Kapitel genauer betrachtet werden soll.

Kapitel 5

Data Minig

5.1 Einführung

Data Mining ist das derzeit geeignetste Mittel, um aus großen Datenbeständen implizit vorhandenes, allerdings bislang unbekanntes und potentiell nützliches Wissen herauszuziehen, damit direkt Maßnahmen daraus abgeleitet werden können.
Diese Verfahren findet sowohl in der Wirtschaft, als auch in der Medizin, im Sport und in anderen Bereichen Anwendung, um Strukturen und Zusammenhänge in komplexen Datenmengen herauszufinden. Denn die Summe der Informationen einer Datenbank ist größer als die Summe der Einzeldaten, wenn man Zusammenhänge zwischen den einzelnen Daten findet. In der Wirtschaft wird das Data Mining beispielsweise zur Kategorisierung von Kunden, zur Webseiten- Optimierung oder zur frühzeitigen Ermittlung von Trendwechseln genutz. Dieses Verfahren ermöglicht einem Unternehmen, zielgerichtet auf seine Kunden eingehen zu können und up-to-date zu sein.

5.2 Der KDD- Prozess

Der Knowledge Database Discovery Prozess ist der wesentliche Bestandteil des Data Mining. Er beschreibt in welchem Zusammenhang das Sammeln von Daten, die Datenvorbereitung, die Anwendung der Data Mining Verfahren und die Datenauswertung stehen, um neue Erkenntnisse aus einem Datensatz zu ziehen.
Zu Beginn des Prozesses steht das Sammeln von Daten im Vordergrund, da-

mit eine Grundlage vorhanden ist, mit der man arbeiten kann. Um sich mit der Datensammlung beschäftigen zu können, ist es zunächst einmal nötig, einen tiefen Einblick in das Themengebiet der Daten zu haben. Dazu gehört die Kenntnis über das Vorkommen und die Formen von Daten. Daten sind eine formalisierte Darstellung von Sachverhalten, Begriffen oder Befehlen, die für die Übermittlung, Interpretation oder Verarbeitung durch den Menschen oder automatische Mittel geeignet ist. Diese Definition gibt darüber Aufschluß, dass Daten fast überall in der Natur vorkommen oder auf andere Art und Weise generiert sein können, wie beispielsweise die Metadaten, welche die Daten über Daten sind. Es gibt aber auch weitere Formen, wie die Historischen Daten, die abgespeichert und nicht mehr aktuell sind, in Anwendung befindliche, beispielsweise kundenspezifische Daten oder zeitkonstante Daten.
Nachdem die einzelnen Informationen gesammelt wurden, werden sie in geeigneter Form abgespeichert. Da Datenbanken den Vorteil, des dauerhaften Speicherns und effizienten Suchens in großen Datenmengen bieten, werden diese auch häufig für die Datenhaltung verwendet. Doch auch Flat Files und Tabellenkalkulationen finden ihren Einsatz.
Da die Daten den Data Mining Spezialisten meist nicht in der gewünschten Form geliefert werden können, folgt zunächst einmal die Datenaufbereitung. Dabei geht es darum, die Daten zu ordnen, unbrauchbare Daten zu entfernen, Lücken in der Datensammlung zu füllen oder aber auch Daten zu codieren. Die jeweilige Art der Datenvorbereitung hängt immer vom Einsatz des im nachfolgenden Verlauf verwendeten Data Mining Verfahrens ab. So wird zum Beispiel für den Einsatz neuronaler Netze beim Data Mining das Data Coding verwendet, um eine Eingabe für das neuronale Netz zu erhalten. In der Praxis wird die Vorbereitung meist mit sogenannten Preprozessing- Tools, Programmiersprachen wie awk, C oder Perl, mit Datenbanken oder manuell in Editoren vorgenommen.
Wenn die Daten in der gewünschten Art und Weise vorliegen, werden die unterschiedlichen Data Mining Verfahren auf diese enorm großen Datensätze angewendet. Darunter fallen Methoden wie die Warenkorbanalyse, bei der nach bestimmten Assoziationen (Regeln) innerhalb eines Datensatzes gesucht wird. Ein Beispiel einer solchen Regel ist, wer Produkt 1 kauft, kauft auch Produkt 7. Entscheidend für die Wichtigkeit einer Regel sind allerdings letztendlich Support und Confidence. Der Support gibt an, wie oft alle Elemente der Regel in der Datenbank vorkommen und der Confidence bestimmt den Durchschnitt von Head und Body einer Regel. Head und Body wiederum sind die wenn- dann- Bedingungen. Die Punkte Support und Confidence sind genau aus dem Grund wichtig, da es nicht von großer Relevanz ist, wenn aus einem riesigen Datensatz genau ein Kunde eine Regel erfüllt. Beispiele für

andere Verfahren sind das Erstellen von Klassifikationen oder das Clustering. Beim Clusering werden Gruppen von Datensätzen zusammengesucht, die gemeinsame Merkmale aufweisen. Diese werden abstrahiert und in ihrer Ähnlichkeit auf ein gewisses Distanzmaß untersucht. Dies ist sowohl für eine räumliche, als auch für eine farbliche Unterscheidung möglich und führt zu Erkenntnissen, die nicht trivial aus den anfänglichen Datensätzen hervorgehen. Sollten allerdings die gewonnenen Informationen nicht ausreichend sein, was in der Regel der Fall ist, so wird der KDD- Prozess an einer geeigneten Stelle von Neuem begonnen, da es durchaus möglich ist, dass durch eine andere Datenvorbereitung und einen anderen Betrachtungsmittelpunkt völlig neue Informationen zum Vorschein kommen.

Literaturverzeichnis

[1] VolesungsFolien zum Datamining - Christoph Schommer - www.cs.uni-potsdam.de/~borchi/Folien/

[2] www.iicm.edu/thesis/hforstinger/Kapitel6.html

[3] www.8ung.at/mobileworkshop/artikel_id65.htm

[4] www.networkworld.de

[5] www.ibusines.de/shop/db/shop.0472hr.html