

Data Warehouse in der Marktforschung: Analyse von Benutzeraktivitäten in virtuellen Welten

Bei der Online-Marktforschung läßt sich mit Hilfe virtueller 3D-Welten das Konsumentenverhalten beobachten. Da die Struktur der dabei anfallenden Daten multidimensional ist und die marktforschungsrelevanten Anfragen eine hohe OLAP-Affinität aufweisen, wurde hierzu im Rahmen des Forschungsprojekts Benevit ein Data Warehouse konzipiert und realisiert. Im Rahmen einer Standardisierung der Protokollierung von Konsumentenaktivitäten und deren prototypischen Realisierung erwies sich der Data Warehouse/OLAP-Ansatz zur Speicherung, Analyse und Präsentation der Marktforschungsdaten als adäquates Instrumentenbündel.

1 Online-Marktforschung als Anwendungsgebiet für Data Warehouses und OLAP

Das Internet gehört zu den größten Wachstumsmedien weltweit und verzeichnete in den vergangenen Jahren trotz der letztjährigen Ernüchterung einen enormen Boom. Sowohl das Angebot an Online-Inhalten als auch die Zahl der Nutzer steigen exponentiell. Mittlerweile gehen Schätzungen von über 600 Millionen Usern weltweit aus (vgl. Computer Industry Almanac Inc 2001).

Der Marktforschung sind die daraus erwachsenen Chancen des Mediums Internet nicht verborgen geblieben. Das Methodenspektrum der Marktforschung wurde um internetbasierte Marktforschung erweitert, wobei die Online-Befragung das verbreitetste Online-Instrument darstellt. Dabei ergeben sich die Potentiale der internetbasierten Marktforschung vor allem aus Kostenvorteilen bei großen Stichproben, der einfachen Möglichkeit des globalen Einsatzes und der Ubiquität des Internet. Grundsätzlich lassen sich zwei Beziehungsfacetten zwischen Online-Marktforschung und Internet identifizieren: (a) das Internet selbst als Untersuchungsgegenstand, (b) das Internet als Instrument der Primär- oder Sekundärforschung (vgl. Zerr 2001). Während im Bereich der Online-Marktforschung die Befragung mittels Online-Fragebögen derweil breiten Einzug gefunden hat, sind in virtuellen Online-Welten durchgeführte Testmarktverfahren und Kundenlaufstudien ein offener Untersuchungsgegenstand (vgl. Abbildung

1.1). Durch die Möglichkeit der systematischen Beobachtung virtueller Testumgebungen ist es z.B. möglich, den Einfluss des Preises, der Werbung, der Verpackung, der Regalplatzierung und der Regalbestückung auf das Kaufverhalten der Konsumenten zu testen. Darüber hinaus können Tests zur Neuprodukteinführung und zum Produktwechsel-Verhalten durchgeführt werden (vgl. Simulation Research 2002).

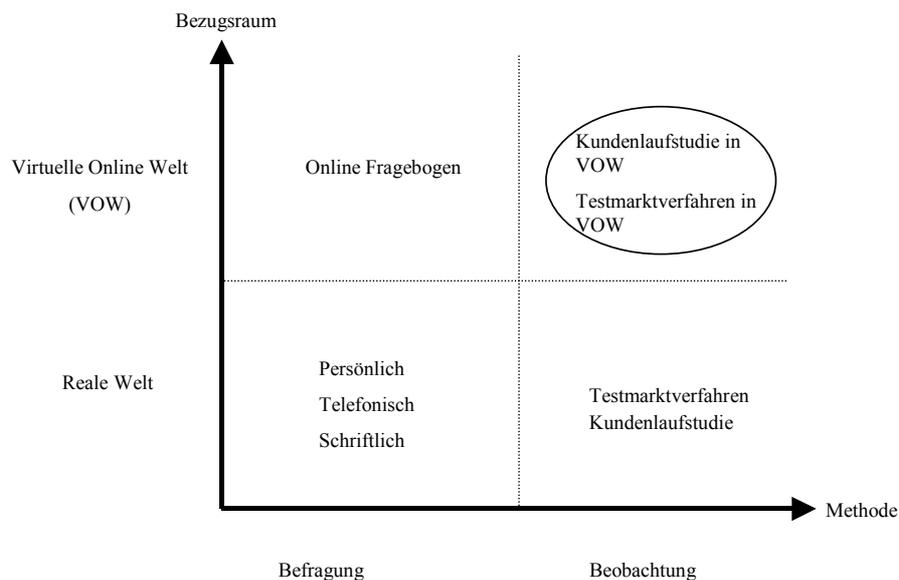


Abbildung 1.1: Offene Untersuchungsgegenstände (Schryen, Herstell 2002)

Die bei der Marktforschung erhobenen Daten sind ausgerichtet auf die zu beantwortenden Fragestellungen (s.o.) adäquat zu speichern. Bauer, Günzel 2000, S. 16, führen in diesem Kontext ein Data Warehouse an: „Einsatzmöglichkeiten eines Data Warehouse bieten sich unter anderem im *Online-Marketing*, dessen wesentliche Aufgabe die Umsetzung des One-to-one-Marketing und die Überführung des Gefühls beim Betreten eines ‚Tante Emma-Ladens‘ in das Internet ist. Aufgabe eines Data Warehouse ist dabei die Sammlung aller möglichen Daten über das Verhalten der Kunden während der Nutzung eines Informationsangebots.“ Praktischen Einsatz findet ein Data Warehouse bereits bei der panelorientierten Marktforschung der GfK¹ Marketing Services im Non-Food-Bereich (vgl. Bauer, Günzel 2000, S. 488-495).

¹ GfK ist die Gesellschaft für Konsumforschung.

Inwiefern der Data Warehouse/OLAP-Ansatz ein geeignetes Instrument zur Speicherung, Analyse und Visualisierung von Marktforschungsdaten aus virtuellen Welten darstellt, wird im folgenden näher untersucht.

Beim marktforscherischen Einsatz virtueller Welten werden neben demographischen Daten vor allem Verhaltensdaten von Probanden erhoben, die sich charakterisieren lassen durch (nicht notwendigerweise vollständige) Angaben zu Proband, Produkt, Zeit, Ort, Aktivität und Fixation², beispielsweise legt der Proband *Hugo Sanchez* am 23.08.2002 um 14:34:45 Uhr das Produkt *Belroyal Erdbeer der Firma Zentis* in den Warenkorb. Die inhärente (semantische) Multidimensionalität der Daten erhebt die Fragen nach Aggregationsebenen bzw. Attributen und Hierarchien; diese werden im Abschnitt 3 thematisiert.

Betrachtet man neben der Speicherung die angestrebte Verwendung dieser Daten und den Nutzerkreis, so finden sich zahlreiche Eigenschaften, die sich auch bei verbreiteten definitorischen Beschreibungen von OLAP-Systemen finden. Kurz 1999, S. 316-325, stellt bedeutende Ansätze – u.a. die von Codd sowie Pendese/Greeth – vor, die vor allem die in Tabelle 1.1 dargestellten Charakteristika aufweisen; ihre Entsprechungen in der (Online-)Marktforschung finden sich dort zugeordnet.

Die dargelegte Data Warehouse/OLAP-Affinität einer (datengestützten) Marktforschung in virtuellen Welten rechtfertigt und motiviert aus Sicht der Autoren eine weitergehende Konzeption und prototypische Realisierung (s. Abschnitte 3 und 4).

² Die Fixation bezeichnet entweder die Blickrichtung des Probanden oder die Art der Drehung eines Produkts, z.B. eine Rotation um x Grad um eine bestimmte Drehachse.

Eigenschaft	OLAP-Ausprägung	MaFo³-Entsprechung
Datensicht	multidimensional	Dimensionen: Proband, Zeit, Produkt etc.
Ziel	Entscheidungsunterstützung	Entscheidungsunterstützung bzgl. des Marketing-Mixes
Methodisches Instrument	Analyse von Daten, z.B. mittels Drill-Optionen	Analyse von Aktivitäten, dynamisch aufgeschlüsselt nach Attributelementen der Dimensionen Produkt, Proband, Zeit etc.
Benutzerbezug	funktionale Transparenz insbesondere für Nicht-IT-Anwender ⁴	Anwender sind primär Marktforscher und Manager

Tabelle 1.1: OLAP-Inhärenz von Marktforschung in virtuellen Welten

2 Projekt Benevit

Bislang dominiert in der Online-Marktforschung die Befragungsmethode bspw. per WWW-Fragebogen, jedoch gestatten die Internettechnologien darüber hinaus auch die insbesondere für kleine und mittelständische Unternehmen relevante (zeit- und kostensparende) Beobachtung von Probanden in virtuellen Verkaufs- und/oder Testräumen. Diese können mit Hilfe von 3D-Darstellungssprachen realisiert und im Web-Browser angezeigt werden. Der Einsatz virtueller Verkaufswelten ist technisch weitgehend gelöst, jedoch fehlen Standards zur Protokollierung von Probandenaktivitäten und eine Verbindung zu marktforscherischen Analysemethoden.

Hier setzt das vom Lehrstuhl für Wirtschaftsinformatik und Operations der RWTH Aachen durchgeführte und von der Stiftung Industrieforschung geförderte Forschungsprojekt Benevit (vgl. Winfor 2002) (Beschreibung von

³ Marktforschung in virtuellen Welten

⁴Die Trennung zwischen Benutzerschnittstelle und zugrundeliegender Basistechnologie soll dazu dienen, die Produktivität von Nicht-SQL-Kennern nicht einzuschränken

Nutzeraktivitäten in virtuellen Test-Umgebungen) an. Es versucht, diese Lücken durch die

- Entwicklung eines offenen Standards zur Protokollierung von Konsumentenaktivitäten in virtuellen Umgebungen,
- prototypische Realisierung und Implementierung dieses Standards in Form eines in der Darstellungssprache *Virtual Reality Modeling Language* (VRML97) erstellten Supermarktes (s. Abschnitt 4) sowie die
- Ankopplung dieses Supermarktes an ein OLAP (Online Analytical Processing)-Marktforschungssystem zur Analyse und Präsentation der erhobenen Daten (s. Abschnitt 4)

zu schließen (s. Abbildung 2.1).

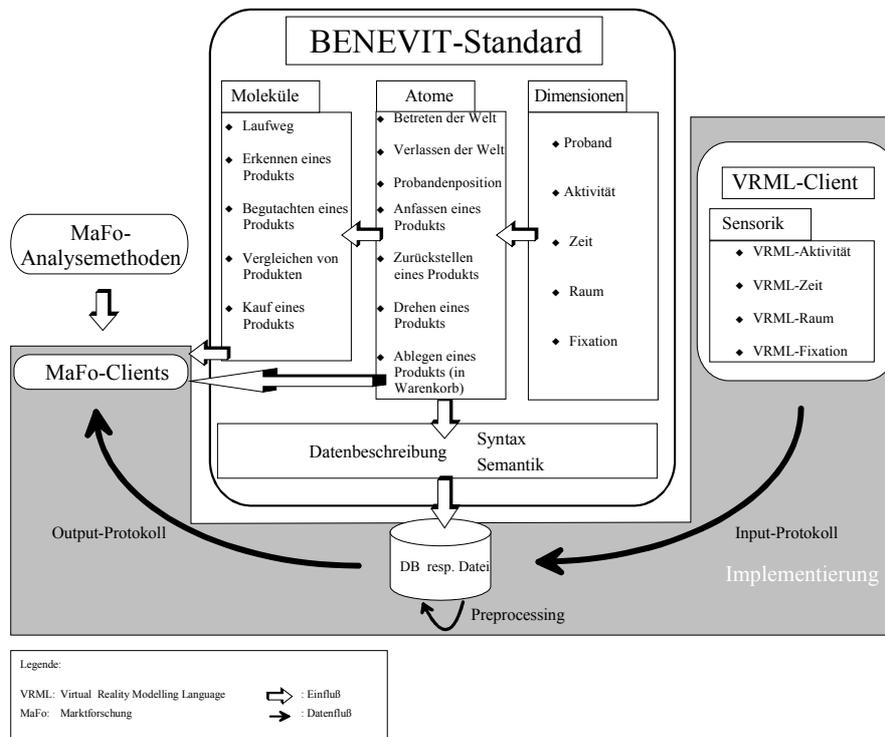


Abbildung 2.1: Standard zur Protokollierung von Benutzeraktivitäten in virtuellen Welten

Der entwickelte Benevit-Standard (vgl. Schryen, Herstell 2002; Schryen 2002) normiert die Beobachtung und Analyse von Benutzeraktivitäten in virtuellen Verkaufswelten und ist realisierungs- und implementierungsunabhängig. Er besteht aus den Aspekten *Dimensionen*, *Atome*, *Moleküle* und *Datenbeschreibung*.

Die ersten drei standardisieren die erfassbaren marktforschungsrelevanten Aktivitäten, die Datenbeschreibung regelt die syntaktische und semantische Behandlung von elementaren Probandenaktivitäten, im Benevit-Standard Atome genannt, so dass eine einheitliche Interpretation der im Datenpool abgelegten Daten möglich ist.

Die Spezifizierung von Atomen erfolgt anhand der fünf Dimensionen *Proband*, *Zeit*, *Raum*, *Produkt* und *Fixation*, d.h. jede Aktivität lässt sich mit einer Kombination entsprechender Ausprägungen parametrisieren bzw. beschreiben, wobei nicht bei jeder Aktivität alle Dimensionen verwendet werden müssen. Ein Proband hält sich beispielsweise zu einem bestimmten Zeitpunkt an einem bestimmten Ort auf und schaut dabei in eine bestimmte Richtung.

Moleküle bezeichnen solche marktforschungsrelevanten Informationen, die sich aus Atomen zusammensetzen und in der Regel erst dann gebildet werden können, wenn ein Proband die Welt wieder verlassen hat. Beispielsweise gilt dies für seinen Laufweg. Die Erzeugung von Molekülen geschieht mittels einer Aggregation von Atomen, die mit Hilfe einer Analyse des Datenpools (Speicherort von Atomen) vorgenommen wird.

Die prototypische Realisierung eines Benevit-konformen Marktforschungssystems findet sich im Abschnitt 4 beschrieben. Zuvor sollen das semantische und logische Datenmodell des dabei zum Einsatz kommenden Data Warehouses vorgestellt werden.

3 Semantisches und logisches Datenmodell

3.1 Semantische Datenmodellierung

Für die semantische Modellierung des eingesetzten Data Warehouse ist zunächst die Bestimmung der relevanten Dimensionen und Fakten notwendig. Die Dimensionen ergeben sich aus der Spezifikation der Atome und Moleküle (vgl. Abschnitt 2) im Benevit-Standard. Für die Aufgabenstellung an das Data Warehouse sind die Moleküle teilweise nicht relevant, da bspw. die Laufweganalyse der Probanden über ein anderes integriertes Softwarewerkzeug (vgl. Abschnitt 4) realisiert wird. Da Moleküle aber grundsätzlich aus der Aggregation mehrerer Atome gebildet werden, können sie ebenfalls für Auswertungen herangezogen werden.

Im Data Warehouse sollen die Aktivitäten der Probanden in Bezug auf Produkte im Zeitablauf näher analysiert und je nach marktforscherischer Anforderung auch aggregiert werden. Weiterhin ergeben sich durch Konstellationen verschiedener Aktivitäten sowie durch deren zeitliche Abfolge weitere Auswertungsobjekte (Moleküle). Daraus ergeben sich die Dimensionen *Probanden*, *Zeit*, *Sortiment (Produkt)* und *Aktivitäten*. Aktivitäten spielen in diesem Kontext eine Sonderrolle, da sie gleichzeitig Gegenstand der Auswertung sind und eine eigene Auswertungsdimension darstellen. Damit wird klar, dass ein betrachtenswertes Fakt im Data Warehouse durch eben diese Aktivitäten bestimmt wird. Das eigentliche Fakt wird dabei durch das Ereignis (Event) des Auftretens einer Aktivität verkörpert. Hierbei ist zu einem späteren Zeitpunkt zu untersuchen, ob alle Ereignisse in gleichem Maße durch Attributausprägungen aller Dimensionen bestimmt werden. Weitere Fakten entstehen durch die Betrachtung der Dauer bestimmter Moleküle, wie z.B. die Verweildauer oder die Betrachtungsdauer eines Objektes.

Zur strukturierten semantischen Modellierung wird die von Determann 2001 entwickelte KOSMO-Methodik herangezogen, die es ermöglicht, die Modellierungssichten Struktur, (Daten-)Quantität und Aggregation(-spfade) in den Betrachtungsebenen Selektion, Definition/Konkretisierung und Konnektion zu betrachten. Sie bietet für die unterschiedlichen Entwicklungsschritte verschiedene grafische Notationsformen an, die im folgenden auszugsweise für die Darstellung des semantischen Modells des Benevit Data Warehouse verwendet werden. Die erwähnten Betrachtungsebenen beziehen sich hierbei auf einen wiederkehrenden Prozess innerhalb der Modellierungssichten, dessen aktuelle Ergebnisse hier lediglich als Momentaufnahme präsentiert werden. Die folgende Darstellung stellt die Vorgehensweise und die Verwendung der unterschiedlichen Notationsformen grafisch dar.

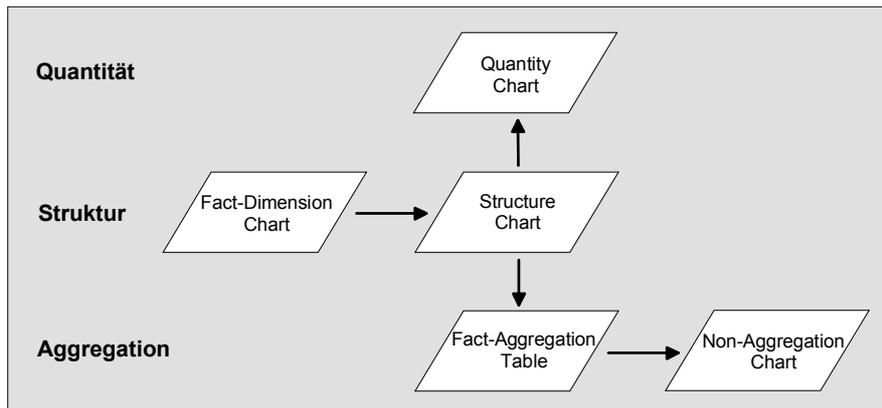


Abbildung 3.1: Notationsformen und ihre Reihenfolge

Zur Erstellung des Fact-Dimension-Chart muss zusätzlich zur Definition und Verifikation der Fakten und Dimensionen geklärt werden, ob und auf welchem Aggregationsniveau die Fakten von den Dimensionen bestimmt werden. Hierzu ist die folgende tabellarische Darstellung hilfreich. Ein „X“ oder Text in einer Dimensionszelle bedeuten, dass das entsprechende Fakt durch Attributausprägungen dieser Dimensionen beschrieben wird. Im Rahmen des Benvit-Standards existieren unterschiedliche Fakttypen, wobei zwischen Typ „A“ für Atom und Typ „M“ für Molekül unterschieden wird. Im Feld Bemerkungen wird die Bedeutung der Fakten näher erläutert.

Aktivitäten	Proband	Sortiment	Zeit	Fakt	Typ	Bemerkungen
ID, Bezeichnung						
1, Eintritt	X	--	X	Ereignis	A	Weltevent
2, Austritt	X	--	X	Ereignis	A	Weltevent
4, Anfassen	X	X	X	Ereignis	A	Produktevent
51, Kippen	X	X	X	Ereignis	A	Produktevent
52, Zurückkippen	X	X	X	Ereignis	A	Produktevent
6, Ablegen	X	X	X	Ereignis	A	Produktevent
7, Zurückstellen	X	X	X	Ereignis	M	Produktevent
10, Kauf	X	X	X	Ereignis	M	Produktevent
--	X	--	--	Dauer	M	Aufenthalts- dauer
--	X	X	--	Dauer	M	Betrachtung Deckel
--	X	X	--	Dauer	M	Betrachtung Rand
--	X	X	--	Dauer	M	Beschäftigung mit Produkt

Tabelle 3.1: Dimensionen und Fakten

Aus dieser Darstellung lässt sich das Fact-Dimension Chart formen, wobei hier insbesondere die unterschiedlichen dimensional Abhängigkeiten deutlich werden. Jede Fakt(-gruppe) ist ausschließlich mit denjenigen Dimensionen verknüpft, die das einzelnen Faktwert beeinflussen. Für Produkt- und Weltevents gilt, dass für jedes Auftreten ein einzelner Faktwert (mit dem Wert „1“) erzeugt wird. Dies ermöglicht Zählungen der Ereignisse und sämtliche darauf aufbauende Berechnungen im zu modellierenden Data Warehouse. Alle anderen Fakten

enthalten als Wert ihre Dauer. Beispielhaft bedeutet dies für die Aufenthaltsdauer, dass für jeden Probanden die Differenz aus Start- und Endzeit seines Aufenthalts im virtuellen Supermarkt als Faktwert abgelegt wird.

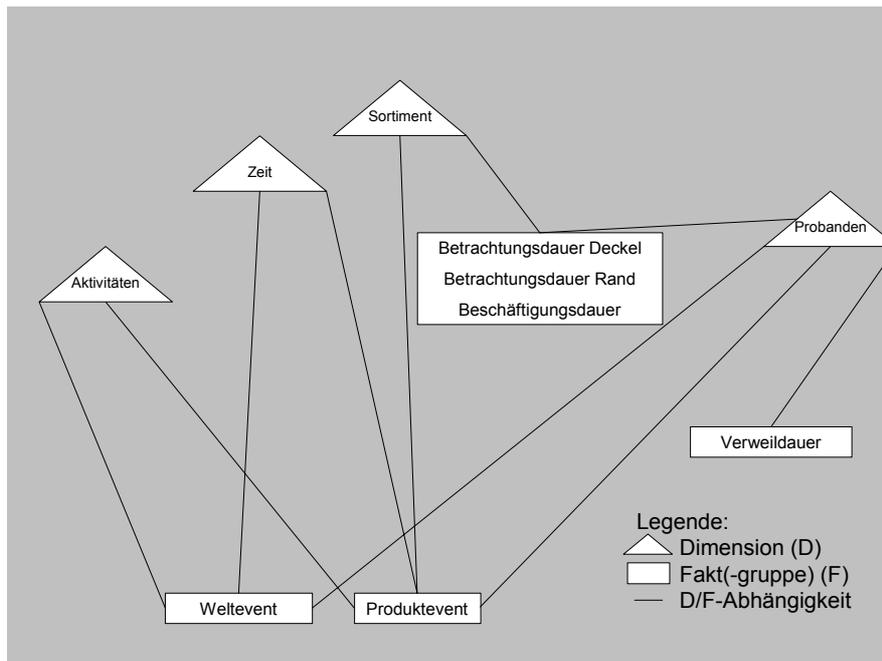


Abbildung 3.2: Fact-Dimension Chart

Im aktuellen Entwicklungsstadium des Data Warehouse ist eine Beschränkung ausschließlich auf die Produktevents vorgenommen worden, so dass aus den in Abbildung 3.2 dargestellten Faktgruppen lediglich diese für die weitere Darstellung bedeutsam ist.

Den nächsten Entwicklungsschritt hin zum semantischen Modell stellt das Structure Chart dar. Hierfür sind alle im Fact-Dimension Chart gezeigten Dimensionen mit ihren Attributen und deren Beziehungen zu ergänzen. Bei den Beziehungen ist zusätzlich die Art der Beziehung zu vermerken; im hier betrachteten Teilmodell sind lediglich 1:N-Beziehungen vorhanden. Die Probanden- und Aktivitätendimension bestehen lediglich aus einer Hierarchiestufe, hier sind für die Zukunft weitere Hierarchisierung möglich und geplant. Insbesondere ist eine weitere Strukturierung der Probandendimension sinnvoll, um anhand von Gruppierungen (z.B. männlich/weiblich, Altersklassen etc.) weitergehende Fragestellungen der Marktforschung beantworten zu können. Die Sortimentsdimension hat eine lineare Struktur, deren Stufung sich mit dem

betrachteten Ausschnitt der Sortimentsstruktur des am Projekt teilnehmenden Unternehmens Zentis deckt. Die Elemente des Dimensionsattributes „Produkt“ stellen in Abgrenzung zu „Artikel“ konkrete Produkte an einem bestimmten Regalstandort im virtuellen Supermarkt dar. „Artikel“ verkörpern allgemein Produkte eines bestimmten Typs, ohne einen solchen Standort zu haben. Die Zeitdimension spannt vom Jahr auf höchster Ebene bis zur stundengenauen Betrachtung eines Tages auf niedrigster Ebene. Die Aggregationspfade laufen sowohl über Monate als auch über Kalenderwochen zum Jahr. Wochentage als Generalisierung von Tagen und Stundentyp als Generalisierung von Tagesstunden ermöglichen Auswertungen der Fakten auf diesen Ebenen, wobei Stundentyp eine stundengenaue Uhrzeit zwischen 00:00 und 23:00 unabhängig von konkreten Tagen darstellt. Für Produktevents ergibt sich aus dieser Modellierungsweise das in Abbildung 3.3 gezeigte Structure Chart.

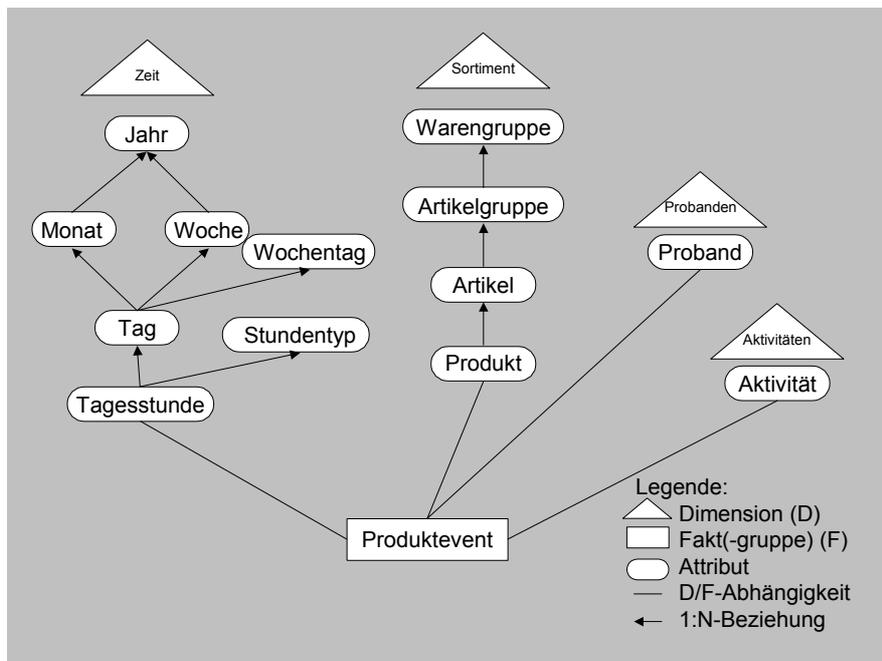


Abbildung 3.3: Structure Chart

In ähnlicher Form lassen sich für die anderen Faktgruppen ebenfalls Structure Charts erstellen. Je nach Komplexität der Darstellung können diese schließlich zu einem einzigen Structure Chart zusammengefasst werden oder in ihrer separaten Darstellung verbleiben, wobei sich dann die Darstellung einiger Dimensionen wiederholt.

Eine Identifizierung und Kategorisierung von *marktforschungsrelevanten* Kennzahlen (Fakten) kann anhand der Dimensionen vorgenommen werden. Für die Kombination Produktdimension/Aktivitätsdimension zeigt Abbildung 3.4 ein Kennzahlensystem; weitere Kennzahlensysteme unter Einbeziehung der anderen Dimensionen werden derzeit erarbeitet. Das dargestellte Structure Chart deckt mit dem Fakt Produktevent und durch die Einbeziehung aller betrachtbaren Dimensionen einen weiten Teil dieser marktforschungsrelevanten Kennzahlen ab.

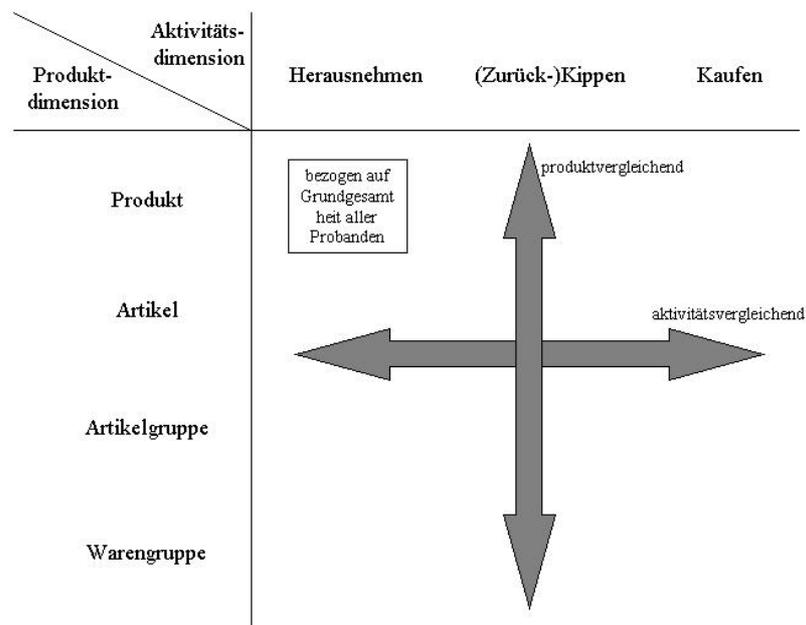


Abbildung 3.4: System von aktivitäts- und produktvergleichenden Kennzahlen

Aktivitätsvergleichende Kennzahlen fixieren beispielsweise einen bestimmten Artikel und ermitteln das Verhältnis der Anzahl herausgenommener und gekaufter Produkte dieses Artikels zu der Anzahl der herausgenommenen. Analog lassen sich Verhältnisse auf anderen Attribut(eben)en bilden. Produktvergleichende Kennzahlen hingegen fixieren z.B. die Aktivität *Kaufen* und bilden das Verhältnis der Anzahl gekaufter Produkte einer Artikelgruppe zur Anzahl der gekauften Produkte der übergeordneten Warengruppe. Im Sinne der Produkt- und Aktivitätsdimension nicht-vergleichende Kennzahlen ermitteln das Verhältnis der Anzahl derjenigen Probanden, die die fixierte Aktivität bezogen auf das fixierte Attributelement der Produktdimension durchgeführt haben, zu der Anzahl aller Probanden.

Die KOSMO-Methodik schlägt bereits an dieser Stelle des semantischen Modellprozesses vor, eine Abschätzung der zu erwarteten Datenmengen (Fakten und Dimensionen) anhand eines sogenannten Quantity Charts vorzunehmen. Dies soll verhindern, dass im weiteren Erstellungsprozess des Data Warehouse Fakten und Granularitätsniveaus von Dimensionen abgelegt werden, deren Wert für Fragestellungen im Vergleich zur dafür benötigten Speicherkapazität nur sehr gering ausfällt. Da sich das Data Warehouse im Benevit-Projekt noch in einem recht frühen Entwicklungsstadium befindet, wird auf eine Entwicklung des Quantity Charts verzichtet, da sich aktuell die insgesamt betrachtete Grundmenge der Daten noch auf einem prototypisch niedrigen Niveau befindet.

An die semantische Abschätzung der notwendigen Speicherkapazitäten schließt sich die Betrachtung der Aggregationsperspektive des Data Warehouse an. Im Fact-Aggregation Table wird definiert, welche Fakten sich in welcher Weise entlang der Dimensionenspfade aggregieren lassen (Navigationspfade). Weiterhin stehen die Berechnungsvorschriften zur Datenverdichtung im Kern dieser Darstellung. Aufgrund der Tatsache, dass in der semantischen Modellierung von Data Warehouse Systemen üblicherweise eine Vielzahl verschiedener Fakten Verwendung findet, wird für das Fact-Aggregation Table einer tabellarischen Darstellung der Vorzug vor einer graphischen Darstellung gegeben, um die Übersichtlichkeit zu erhöhen. Dargestellt werden die Einflussdimensionen des jeweiligen Fakts, die Art der Aggregierbarkeit in ihren möglichen Ausprägungen „voll“, „semi“ und „nicht“ sowie die Berechnungsformel zur Aggregation. Außerdem wird vermerkt, ob es sich bei dem jeweiligen Fakt um eine Ableitung aus anderen Fakten handelt; so läßt sich beispielsweise ein einfacher Gewinnbegriff aus der Subtraktion des Fakts „Kosten“ vom Fakt „Umsatz“ ableiten. Bezüglich der Art der Aggregierbarkeit bedeutet „semi“, dass ein Fakt nicht in allen Einflussdimensionen aggregierbar ist; z.B. wäre ein Lagerbestandswert sinnvoll geographisch durch Summation zu aggregieren, aber nicht zeitlich.

Für den Prototypen des Benevit Data Warehouse ergibt sich die folgende Fact-Aggregation Darstellung:

	Dimension				...
Bezeichnung Fakt	Aktivitäten	Zeit	Sortiment	Probanden	...
Produktevent	X	X	X	X	...

...	Aggregierbarkeit			Berechnung	
	...	voll	semi	nicht	abgeleitet
...	X			nein	COUNT distinct (ProdEvent)

Tabelle 3.2: Fact-Aggregation Table

Das Fakt Produktevent wird hier mit einer Berechnungsvorschrift versehen, die es ermöglicht, innerhalb der Dimensionen zu aggregieren. Eine solche Kombination aus Fakt und Berechnungsvorschrift wird auch als Metrik bezeichnet.

Als letzter Schritt der semantischen Modellierung schließt die KOSMO-Methodik mit dem Non-Aggregation Chart ab, in dem die Faktengruppen in Bezug auf die Art ihrer Aggregierbarkeit zusammen mit den Dimensionen dargestellt werden. Nur semi- und nichtaggregierbare Fakten werden hier aufgenommen. Nicht aggregierbare Fakten werden in der Grafik gesondert ohne Bezug zu den Dimensionen dargestellt, was auf die Unmöglichkeit der Navigation innerhalb der Dimensionen für diese Fakten hinweist. Da es sich bei dem aktuell betrachteten Fakt im Rahmen des Benevit Data Warehouse um ein voll aggregierbares Fakt handelt, entfällt diese Darstellung.

3.2 Logische Datenmodellierung

Innerhalb der logischen Datenmodellierung wurde die Übertragung der Semantik auf ein relationales Datenmodell gewählt, da die spätere Auswertung mit einem ROLAP-Werkzeug (MicroStrategy 7) erfolgen soll.

Die Repräsentation des semantischen Modells in einer relationalen Struktur erfolgt über ein vollständig denormalisiertes Snowflake Schema (vgl. Kurz 1999, S. 164ff; Hahne 1999), welches im aktuellen Ausbaustand des Data Warehouse lediglich eine Faktentabelle enthält. Für jede Dimension wird auf jeder Stufe eine sogenannte Lookup-(Dimensions-)Tabelle gebildet, welche alle Attributelemente der jeweils betrachteten und aller in der Hierarchie oberhalb befindlichen Stufen enthält. Somit ist es möglich, Faktentabellen anderer Dimensionalität oder einfach Aggregatfaktentabellen zur Performance-Steigerung ohne großen Aufwand im Data Warehouse zu modellieren und abzulegen. Aufgrund der relativ geringen Dimensionstiefe ist der durch die Verwendung dieses Snowflake-Schemas steigende Speicherplatzbedarf zu vernachlässigen und das Schema somit als adäquat zur Problemlösung im Projektumfeld anzusehen. Die folgende Darstellung zeigt exemplarisch einen Ausschnitt aus dem entwickelten Schema.

Gezeigt werden lediglich die Dimensionen Zeit und Sortiment mit einigen Dimensionsstufen bzw. Attributen.

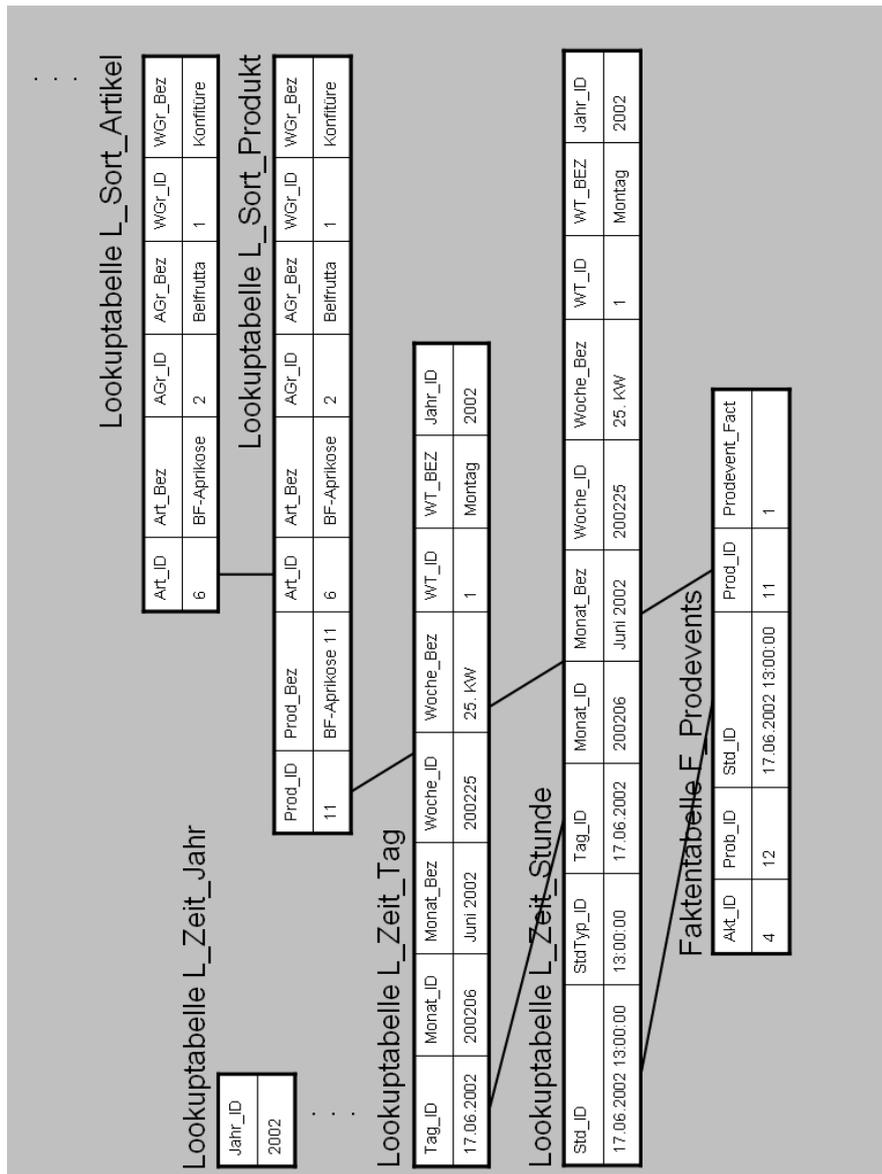


Abbildung 3.5: Snowflake Schema Benevit (Ausschnitt)

Die Strukturierungskriterien der Dimensionen werden durch die ID-Spalten definiert. Die Bezeichnungsspalten dienen der verbesserten Darstellung der Werte im Auswertungswerkzeug. Die graphische Darstellung zeigt die bewusst in Kauf genommene Redundanz innerhalb der Dimensionstabellen nochmals deutlich. Dies führt innerhalb der physischen Modellierung und der Abbildung in der verwendeten Data Warehouse-Datenbank zu größerer Flexibilität (Aggregat-speicherung) und zur Steigerung der Berichtsperformance durch die Verringerung von JOIN-Operationen bei der Navigation innerhalb der Dimensionen.

4 Prototypische Realisierung im Rahmen einer Fallstudie

Die prototypische Realisierung einer auf dem Benevit-Standard basierenden Marktforschungsumgebung erfolgt client-server-basiert: mit Hilfe der international standardisierten 3D-Darstellungssprache VRML (vgl. VRML 1997) wurden zwei Supermärkte⁵ erstellt. Diese können mit einem Web-Browser geladen und einem Plug-In zur VRML-Darstellung angezeigt werden. Der Proband kann sich nun frei im Supermarkt bewegen und Produkte greifen, drehen, in den Warenkorb legen usw. (s. Abbildung 4.1).

⁵ Der eine Supermarkt dient primär der Ermittlung von Laufwegen, der andere wird in Kooperation mit dem Nahrungsmittelhersteller Zentis GmbH & Co. primär für die Betrachtung von produktbezogenen Aktivitäten verwendet.



Abbildung 4.1: Screenshots der Virtuellen Supermärkte

Der in einer Kooperation mit dem Nahrungsmittelhersteller Zentis entstandene Supermarkt kann online unter <http://www.winfor.rwth-aachen.de/inhalte/forschung/vrmafo/supermarkt/eingabe.html> besucht werden.

Die elementaren Probandenaktivitäten (s. Atome im Benevit-Standard und Abbildung 2.1) werden mittels der im Supermarkt integrierten Sensorik als http-Requests an einen Webserver übermittelt, dort von einem Anwendungsserver aufbereitet (u.a. werden Moleküle erzeugt, s. Benevit-Standard und Abbildung

2.1) und über die ODBC-Schnittstelle in die Faktentabellen des Microstrategy-Data Warehouses eingespeist. Dieses dient als Marktforschungs-Frontend zur Analyse der Konsumentenaktivitäten; zur Analyse und Visualisierung von Laufwegen wurde eine proprietäre Grafikanwendung erstellt. Abbildung 4.2 illustriert das Zusammenwirken am Szenario der produktbezogenen Data Warehouse-Analyse.

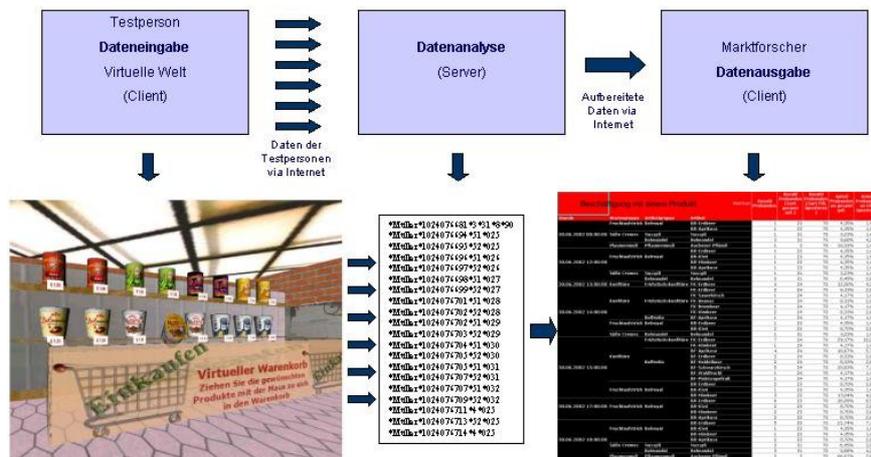


Abbildung 4.2: Benevit-Marktforschungssystem

Exemplarisch für die vielseitig gestaltbaren Marktforschungsberichte (s. Abschnitt 6) ist vergrößert in Abbildung 4.3 ein Bericht zu sehen, der die absolute und relative Beschäftigung von Probanden mit bestimmten Produkten wiedergibt.

Beschäftigung mit einem Produkt				Metriken	Anzahl Probanden	Anzahl Probanden (Sort. gesamt gef.)	Anzahl Probanden (Sort. Filt. ignorieren)	Anteil Probanden an gesamt gef.	Anteil Probanden an Filt. ignorieren
Stunde	Warengruppe	Artikelgruppe	Artikel						
30.06.2002 00:00:00	Fruchtaufstrich	Belroyal	BR-Erdbeer	1	23	70	4,35%	1,43%	
			BR-Aprikose	1	23	70	4,35%	1,43%	
	Süße Cremes	Nusspli	Nusspli	1	31	70	3,23%	1,43%	
			Belmandel	3	31	70	9,68%	4,29%	
30.06.2002 12:00:00	Pflaumenmuß	Pflaumenmuß	Aachener Pflümli	1	23	70	4,35%	1,43%	
			BR-Erdbeer	1	23	70	4,35%	1,43%	
	Fruchtaufstrich	Belroyal	BR-Kiwi	1	23	70	4,35%	1,43%	
			BR-Himbeer	1	23	70	4,35%	1,43%	
30.06.2002 13:00:00	Süße Cremes	Nusspli	Nusspli	1	23	70	4,35%	1,43%	
			Belmandel	2	31	70	6,45%	2,86%	
	Konfitüre	Frühstückskonfitüre	FK-Erdbeer	3	24	70	12,50%	4,29%	
			FK-Sauerkirsch	2	24	70	8,33%	2,86%	
30.06.2002 14:00:00	Konfitüre	Frühstückskonfitüre	FK-Ananas	1	24	70	4,17%	1,43%	
			FK-Brombeer	2	24	70	8,33%	2,86%	
	Belfrutta	Belfrutta	BF-Aprikose	1	24	70	4,17%	1,43%	
			BR-Erdbeer	1	23	70	4,35%	1,43%	
30.06.2002 15:00:00	Fruchtaufstrich	Belroyal	BR-Kiwi	2	23	70	8,70%	2,86%	
			BR-Erdbeer	7	24	70	29,17%	10,00%	
	Süße Cremes	Belmandel	Belmandel	1	31	70	3,23%	1,43%	
			Frühstückskonfitüre	1	24	70	4,17%	1,43%	
30.06.2002 17:00:00	Konfitüre	Belmandel	BF-Aprikose	4	24	70	16,67%	5,71%	
			BF-Erdbeer	2	24	70	8,33%	2,86%	
	Belfrutta	Belfrutta	BF-Heidelbeer	2	24	70	8,33%	2,86%	
			BF-Schwarzkirsch	5	24	70	20,83%	7,14%	
30.06.2002 18:00:00	Fruchtaufstrich	Belroyal	BR-Erdbeer	1	24	70	4,17%	1,43%	
			BR-PinkGrapefruit	1	24	70	4,17%	1,43%	
	Fruchtaufstrich	Belroyal	BR-Erdbeer	2	23	70	8,70%	2,86%	
			BR-Kiwi	1	23	70	4,35%	1,43%	
30.06.2002 17:00:00	Fruchtaufstrich	Belroyal	BR-Himbeer	3	23	70	13,04%	4,29%	
			BR-Erdbeer	6	23	70	26,09%	8,57%	
	Fruchtaufstrich	Belroyal	BR-Kiwi	2	23	70	8,70%	2,86%	
			BR-Himbeer	2	23	70	8,70%	2,86%	
30.06.2002 18:00:00	Fruchtaufstrich	Belroyal	BR-Aprikose	2	23	70	8,70%	2,86%	
			BR-Erdbeer	5	23	70	21,74%	7,14%	
	Fruchtaufstrich	Belroyal	BR-Kiwi	1	23	70	4,35%	1,43%	
			BR-Himbeer	1	23	70	4,35%	1,43%	
30.06.2002 18:00:00	Süße Cremes	Nusspli	Nusspli	2	31	70	6,45%	2,86%	
			Belmandel	3	31	70	9,68%	4,29%	
	Pflaumenmuß	Pflaumenmuß	Aachener Pflümli	2	3	70	66,67%	2,86%	

Abbildung 4.3: Exemplarischer Bericht zur Beschäftigung mit Produkten

5 Resümee und Ausblick

Die Speicherung und Analyse von Benutzeraktivitäten in virtuellen Welten stellt ein adäquates Anwendungsgebiet für Data Warehouses und OLAP dar. Der vorliegende Beitrag schlägt hierzu erste semantische und logische Datenmodelle vor, deren Eignung derzeit im Rahmen einer empirischen Studie überprüft werden. Hinsichtlich der Generierung von Kennzahlensystemen wurde aus aktivitätsproduktbezogener Sicht systematisiert, die Einbeziehung weiterer Dimensionen ist Gegenstand zukünftiger Betrachtungen. Von marktforscherischem Interesse können neben OLAP auch Verfahren des Data Mining sein, die sich beispielsweise der Aufgabe der Probandentypisierung bzw. -clustering widmen.

6 Literatur

- Bauer, A., Günzel, H. (Hrsg.): Data Warehouse Systeme - Architektur, Entwicklung, Anwendung; dpunkt; Heidelberg 2000.
- Computer Industry Almanac Inc.: There will be 625 Million Computers-in-Use Year-End 2001; 15.07.2001; <http://www.c-i-a.com/pr0701.htm>, Stand: 03.05.02.
- Determann, L.: Modellierung Analytischer Informationssysteme: Entwurf einer Methodik zur multidimensionalen Datenstrukturierung, Deutscher Universitäts-Verlag, erscheint 2002.
- Hahne, M.: Logische Datenmodellierung für das Data Warehouse – Bestandteile und Varianten des Star Schemas, in: Analytische Informationssysteme: Data Warehouse, On-Line Analytical Processing, Data Mining (Hrsg.: Chamoni, P., Gluchowski, P.); 2. neubearb. Aufl.; Springer; Berlin, Heidelberg 1999, S. 145-170.
- Kurz, A.: Data Warehousing - Enabling Technology; MITP; Bonn 1999.
- Schryen, G.: Spezifikation des Benevit-Standards zur Beobachtung und Analyse von Benutzeraktivitäten in Virtuellen Verkaufswelten, Arbeitsbericht des Instituts für Wirtschaftswissenschaften der RWTH Aachen, NR. 02/06, Aachen April 2002.
- Schryen, G., Herstell, J.: Online-Marktforschung –Analyse von Konsumentenverhalten in virtuellen Umgebungen; Information Management & Consulting (erscheint 2002).
- Simulation Research: Visionary Shopper has been used for a number of different purposes; <http://www.simulationresearch.com/PRODUCT.htm>, Stand: 03.05.02.
- VRML Consortium: The Virtual Reality Modeling Language, Spezifikation VRML97, International Standard ISO/IEC 14772-1:1997, http://www.vrml.org/fs_specifications.htm, Stand: 03.05.02.
- Winfor 2002: Webseite des Lehrstuhls für Wirtschaftsinformatik und OR, <http://www.winfor.rwth-achen.de/inhalte/forschung/vrmafo/ueberblick/ueberblick.html>, Stand: 04.07.02.
- Zerr, Konrad: Online-Marktforschung – Erscheinungsformen und Nutzenpotentiale; in: Online-Marktforschung - Theoretische Grundlagen und praktische Erfahrungen (Hrsg.: Theobald, A.; Dreyer, M.; Starsetzki, T.); Gabler; Wiesbaden 2001; S. 8-25.