

## **Ein integriertes Lehrkonzept zur Business Intelligence Plattform von SAS**

Andreas Hilbert  
Technische Universität Dresden  
Professur für Wirtschaftsinformatik,  
insb. Informationssysteme im  
Dienstleistungsbereich  
Münchner Platz 3  
01187 Dresden  
hilbert@wiid.wiwi.tu-dresden.de

Alexander E. Müller  
Technische Universität Dresden  
Professur für Wirtschaftsinformatik,  
insb. Informationssysteme im  
Dienstleistungsbereich  
Münchner Platz 3  
01187 Dresden  
mueller@wiid.wiwi.tu-dresden.de

Sandro Kürschner  
Technische Universität Dresden  
Professur für Wirtschaftsinformatik,  
insb. Informationssysteme im  
Dienstleistungsbereich  
Münchner Platz 3  
01187 Dresden  
kuerschner@wiid.wiwi.tu-dresden.de

### **Zusammenfassung**

Die Einführung einer Data-Warehousing- bzw. umfassender auch Business-Intelligence-Lösung in Unternehmen ist ein komplexer Prozess, der umfangreiches Basiswissen in verschiedenen Disziplinen voraussetzt. Der Beitrag zeigt, wie die verschiedenen Bausteine in einem integrierten Lehrkonzept zusammengeführt werden. Im Mittelpunkt des Beitrages stehen die für eine Implementierung in SAS zu erwerbenden Kompetenzen. Neben den technischen Kompetenzen sind dies auch notwendige Fähigkeiten auf strategischer, konzeptioneller und logischer Ebene. Ferner wird gezeigt, wie das umfangreiche analytische Potenzial eines einmal erstellten Data Warehouses mit SAS gelehrt werden kann.

**Schlüsselwörter:** Data Warehouse, Data Mining, Business Intelligence, Integriertes Lehrkonzept, Lehre, SAS Management Console, SAS Enterprise Guide, SAS Enterprise Miner, SAS OLAP Cube Studio

## **1 Integriertes Lehrkonzept**

Das auf den folgenden Seiten vorgestellte integrierte Lehrkonzept zur SAS Business Intelligence Plattform ist seit mehreren Semestern erfolgreich an der Professur für Wirtschaftsinformatik, insb. Informationssysteme im Dienstleistungsbereich (WIID) der TU Dresden im Einsatz. Es setzt sich aus drei Hauptkomponenten zusammen: Zum einen finden theoretische Vorlesungen zur Business Intelligence (BI) mit den Schwerpunkten

Data Warehousing und Data Mining statt. Zum anderen werden Rechnerübungen angeboten, die sich ebenfalls in die Schwerpunkte Data Warehousing und Data Mining untergliedern lassen. Als dritten und letzten Punkt finden sich Projektseminare im Angebot, welche teilweise in Kooperation mit Praxispartnern angeboten werden. Diese drei Komponenten sind eng miteinander verknüpft, weshalb der Besuch einer Veranstaltung oftmals notwendige Voraussetzung für den Besuch einer anderen ist.



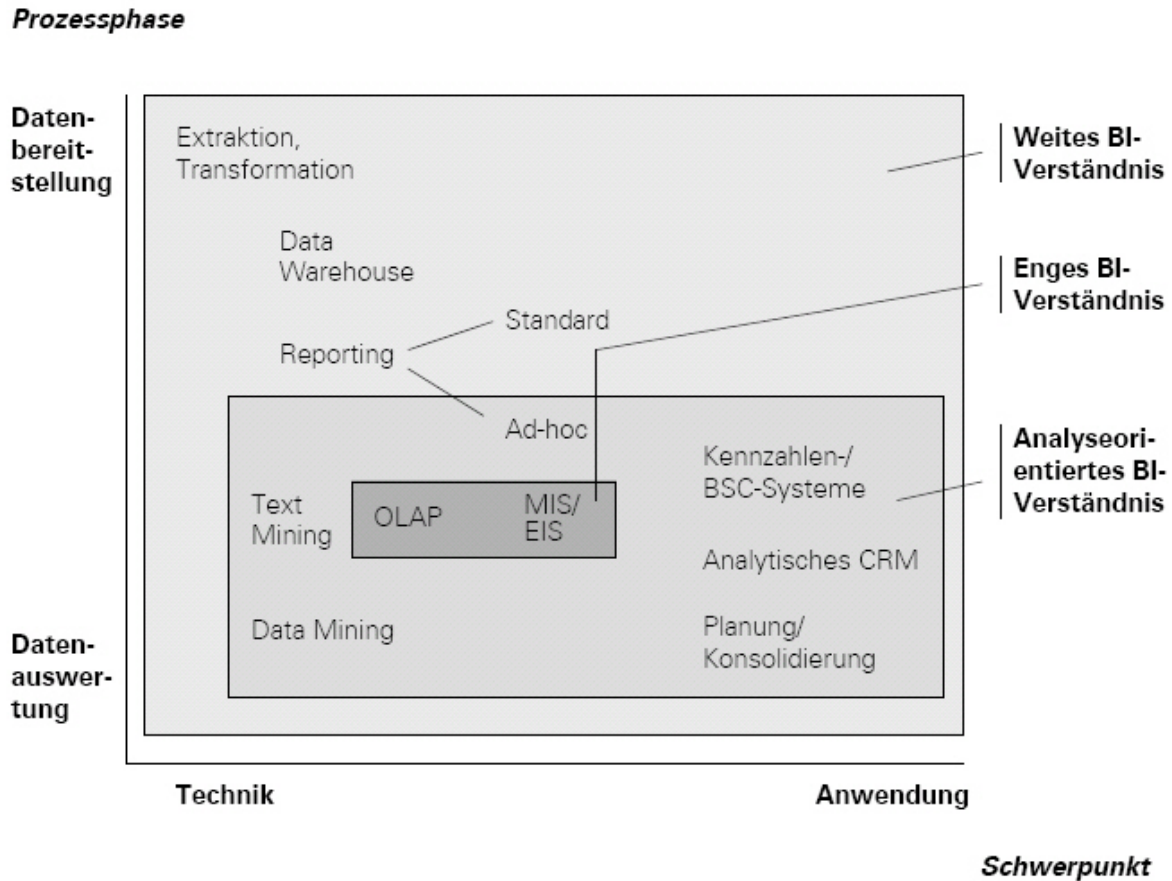
**Abbildung 1:** Bestandteile des Lehrkonzepts (eigene Darstellung)

Wie in Abbildung 1 dargestellt, bilden die Vorlesungen einen theoretischen Rahmen um die Rechnerübungen zum Data Warehousing und Data, Web und Text Mining. Als zusätzliches und fakultatives Angebot für Studierende, die an einer weiterführenden Beschäftigung mit dem Thema interessiert sind, finden sich Projektseminare im Lehrveranstaltungsangebot. Für diejenigen Studierenden, die alle oder zumindest einen Großteil der angebotenen Lehrveranstaltungen durchlaufen haben, besteht die Möglichkeit, von der Professur das Zertifikat „Excellence in Business Intelligence“ zu erhalten, um so die über das Normalmaß hinausgehende Qualifikation im Bereich Business Intelligence nachzuweisen.

## 2 Theoretische Veranstaltungen

Zwei grundlegende Vorlesungen zur Business Intelligence vermitteln das theoretische Rüstzeug für alle weiterführenden Veranstaltungen zur SAS Business Intelligence Platt-

form. Aufgrund einer sehr weit gefassten Begriffsabgrenzung von BI wird mit diesen Veranstaltungen alles abgedeckt, was sich, beispielsweise nach Kemper, in Abbildung 2 in das weite Verständnis von Business Intelligence einordnen lässt. Stellenweise ist das gewählte BI-Verständnis sogar noch weiter gefasst.



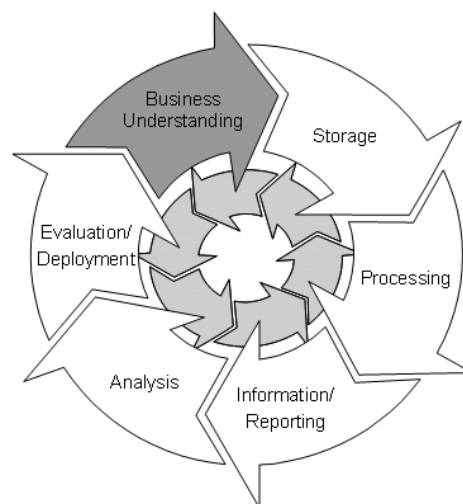
**Abbildung 2:** Facetten der Business Intelligence (vgl. [1], [2])

Jede Vorlesung besitzt ihre individuellen Schwerpunkte, welche zusammengefasst die Hauptforschungsinteressen der Professur widerspiegeln. Dies sind einerseits das Data Warehousing und andererseits das Data, Web und Text Mining, im praktischen Einsatz jeweils mit der SAS Business Intelligence Plattform. Ergänzend zu den Vorlesungen wurde im Zuge der Umstellung auf die Bachelor- und Masterstudiengänge eine semesterbegleitende Projektleistung, bei der sich Studenten in Kleingruppen individuell mit einem Thema ihrer Wahl – aus dem Themenkomplex der Vorlesung – beschäftigen, eingeführt. Die Note hieraus fließt in etwa zu einem Viertel in die Gesamtnote der jeweiligen Prüfung ein.

Die beiden folgenden Kapitel widmen sich dem Aufbau und den Inhalten der theoretischen Veranstaltungen sowie der Relevanz der vermittelten Kompetenzen für die Nutzung der eingesetzten SAS Komponenten.

## 2.1 Business Intelligence, insbesondere Data Warehousing

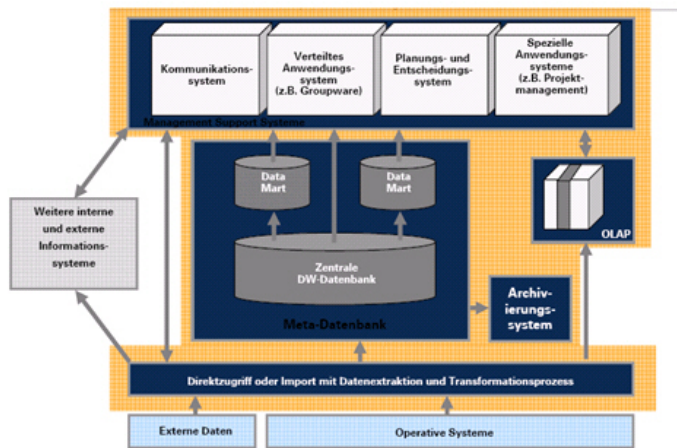
Die Vorlesung Business Intelligence, insb. Data Warehousing führt die Studenten zu Beginn an die Begrifflichkeiten und das Umfeld der Business Intelligence heran. So wird erläutert, wie sich durch die gesteigerte Kundenorientierung im Kundenbeziehungsmanagement die Anforderungen an die Datenbasis verändert haben und dadurch der Einsatz von Business Intelligence bedingt wird. Weiterhin wird geklärt, was unter Business Intelligence im Allgemeinen zu verstehen ist, wie es in der Literatur von den verschiedenen Autoren im Speziellen verstanden wird, aber auch, wie sich die unterschiedlichen Techniken – ohne an dieser Stelle bereits auf die Details einzugehen – konzeptuell einordnen lassen. Ebenso wird das Verständnis von Business Intelligence als iterativer Prozess mit Rückkopplungen (vgl. Abbildung 3), welches an der Professur Anwendung findet, vorgestellt.



**Abbildung 3:** BI als iterativer Prozess mit Rückkopplungen (vgl. [3])

Im Anschluss an diese allgemeinen Überlegungen wechselt der Fokus direkt zum Schwerpunkt der Veranstaltung, dem Data Warehousing (DWH). Unmittelbar nach einer Einführung in die Ziele und Merkmale des DWH, werden die verschiedenen Data-Warehouse-Architekturen auf konzeptioneller Ebene betrachtet (vgl. Abbildung 4). Daran schließt sich die Einführung des OLAP-Begriffs und die theoretische Betrachtung des Würfelkonzepts inklusive der verschiedenen Navigationsarten innerhalb eines solchen OLAP-Würfels an.

Architekturüberblick...



... und Architekturvarianten

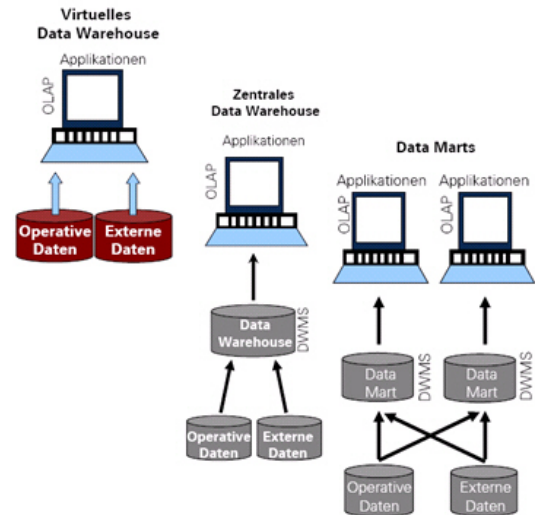


Abbildung 4: Data Warehouse Architekturen (vgl. [4])

Die Studenten kommen an diesem Punkt durch grafische Darstellungen erstmals mit der Software in Berührung, mit der diese Würfel gebaut und betrachtet werden können – dem SAS OLAP Cube Studio und dem SAS Enterprise Guide. Um die Entwicklung eines Data Warehouse aufzuzeigen, werden die drei Ebenen Fachkonzept, DV-Konzept und technische Implementierung auf dem Weg von der betriebswirtschaftlichen Problemstellung zur Informationstechnik näher betrachtet. Anhand der Ausführungen zum DV-Konzept werden den Studenten die Vor- und Nachteile der einzelnen Datenbankschemata vermittelt. Ein ausführliches Beispiel aus dem Lebensmittelhandel lehrt anschließend die praktische Erstellung eines Dimensional Fact Models sowie eines Star- und Snow-Flake-Datenbankschemas. Nach der Theorie zur Entwicklung eines Data Warehouse folgt die Theorie zur Modellierung von Data-Warehouse-Strukturen. In diesem sehr umfangreichen Teil der Veranstaltung werden auf Ebene des konzeptuellen Entwurfs die Modellierungsmodelle Multidimensionales E/R, Multidimensionale UML, ADAPT, Dimensional Fact Model und das Semantische Data-Warehouse-Modell vorgestellt. Der Data-Warehousing-Schwerpunkt der Vorlesung endet mit einer Einführung in die Abfragesprache MDX und der Vorstellung des Common Warehouse Metamodel. Es folgt eine Einführung in die Ziele und Merkmale des Data Mining sowie eine Abgrenzung des Data Mining gegenüber OLAP.

Den Abschluss der Vorlesung bildet eine intensive Auseinandersetzung mit dem KDD-Prozess sowie den Vorgehensmodellen SEMMA und CRISP-DM. Eine kurze Vorstellung der Data-Mining-Problemtypen und -Analysemethoden dient als Ausblick auf die Folgeveranstaltung Business Intelligence, insb. Data Mining.

## 2.2 Business Intelligence, insbesondere Data Mining

Der Aufbau der Veranstaltung Business Intelligence, insb. Data Mining orientiert sich an den einzelnen Schritten des in Abbildung 5 dargestellten KDD-Prozesses und betrachtet im Speziellen die wichtigsten Verfahren des Data Mining.

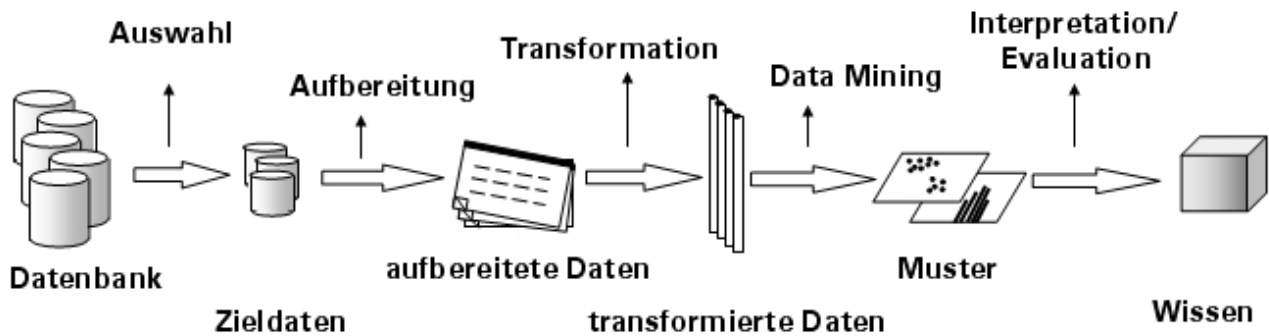


Abbildung 5: Die Phasen des KDD-Prozesses nach [5], S. 9

Den Studenten wird so das Ziel des Data Mining, die Suche nach Zusammenhängen in riesigen Datenmengen, näher gebracht. Zu diesem Zweck werden die grundlegenden Analysemethoden (vgl. Abbildung 6) wie Clusterverfahren, Regressionsansätze, Neuronale Netze, Entscheidungsbäume und Assoziations- bzw. Sequenzanalysen theoretisch erläutert und die zugrunde liegenden Rechenmodelle vorgestellt. Ein wichtiger Aspekt neben der Methodenlehre ist dabei die Interpretation der Ergebnisse.

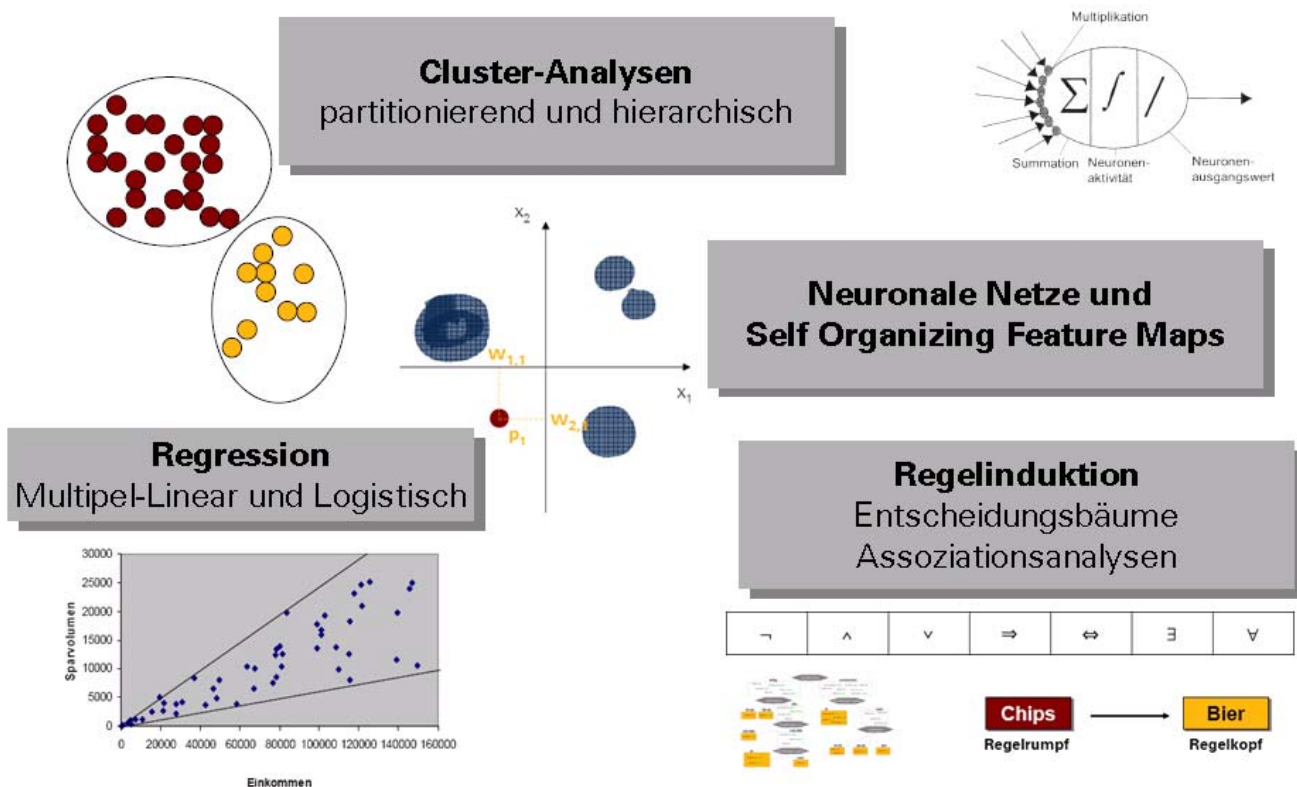


Abbildung 6: Mathematische Methoden (vgl. [6])

Themenbezogene Beispiele aus der Praxis wie die Unterstützung des Cross-Selling bei Finanzdienstleistern, die Tariffindung bei Versicherungen, die Bindung wechselwilliger Kunden (Churn Management) bei Telekommunikationsanbietern oder das Planen und Unterstützen des Kampagnen-Managements im Customer Relationship Management runden die Veranstaltung ab.

Ein wöchentlich stattfindendes Tutorium unterstützt bei den mathematischen Aspekten der Veranstaltung. Unter Anleitung eines Tutors wiederholen die Studierenden mathematische Theorie und spielen weiterführende Modelle durch. Dies soll das Verständnis für die später im Hintergrund ablaufenden Berechnungen der in der Rechnerübung angewendeten Methoden und Modelle weiter ausbauen.

### **3 Praktische Veranstaltungen**

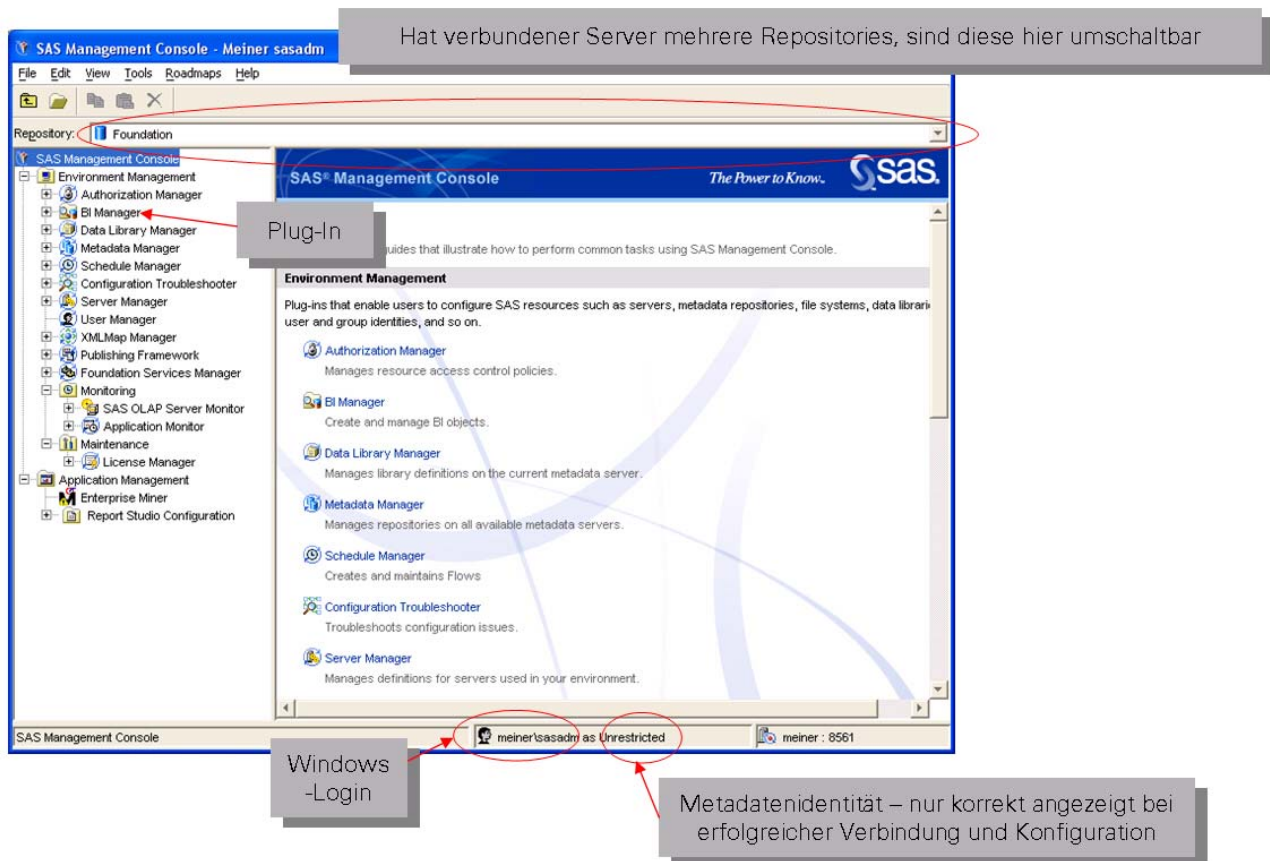
Neben den beiden bereits aufgeführten theoretischen Veranstaltungen wird den Studierenden die Möglichkeit geboten, das in den Vorlesungen erworbene Wissen in rechnergestützten Übungen erstmalig selbst anzuwenden und ihre Kompetenzen im Bereich Business Intelligence weiter auszubauen.

Hierzu werden unmittelbar die verschiedenen Teile der SAS Business Intelligence Plattform eingesetzt. Die Rechnerübungen orientieren sich an den Schwerpunktinhalten der Vorlesungen und sind diesen thematisch zugeordnet. Aus diesem Grund gibt es getrennte Rechnerübungen zum Data Warehousing und zum Data, Web und Text Mining. Diese finden momentan semesterweise alternierend zu den Vorlesungen statt.

Die rechnergestützten Übungen wurden im Zuge der Bachelor-/Masterumstellung ebenfalls um eine semesterbegleitende Projektleistung erweitert. Dadurch wird der Relevanz von eigenverantwortlich bearbeiteten Aufgaben bei der praktischen Einübung von theoretisch vermitteltem Wissen Rechnung getragen. Die Wichtigkeit dieser begleitenden Projektleistungen zeigt sich dadurch, dass durch sie der Erwerb eines Viertels der Gesamtpunkte (Note) der Rechnerübungen möglich ist.

#### **3.1 Rechnerübung Data Warehousing**

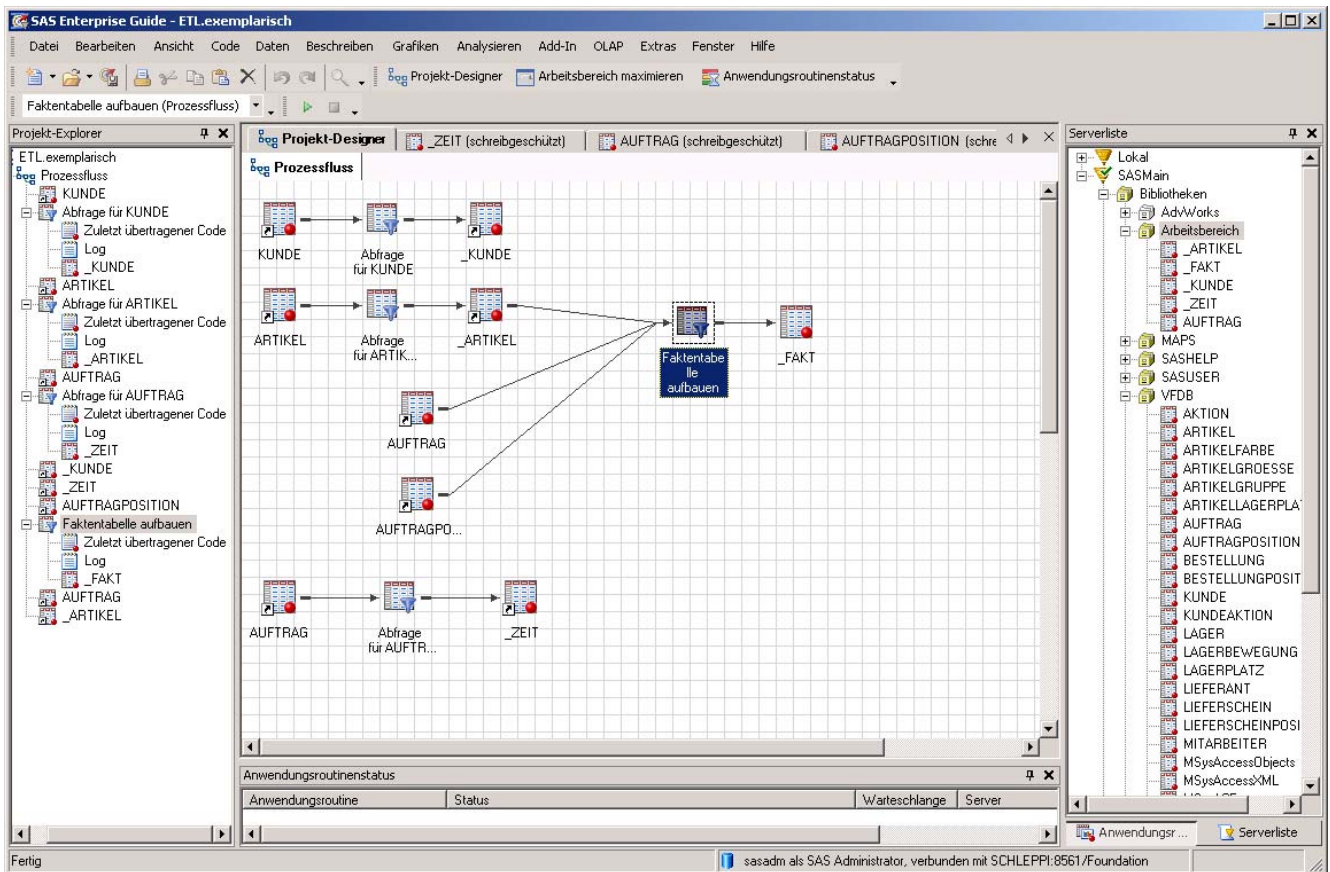
Angelehnt an die Inhalte der Vorlesung Business Intelligence, insb. Data Warehousing vermittelt die rechnergestützte Übung Data Warehousing den Studenten die Kompetenzen, um das vorhandene Theoriewissen unter Zuhilfenahme mehrerer SAS Komponenten umzusetzen. Sie lernen zu Beginn die Administrationsmöglichkeiten kennen, die durch die SAS Management Console (vgl. Abbildung 7), sowohl für das lokale System als auch für entfernte Rechner, bereitgestellt werden.



**Abbildung 7:** SAS Management Console (vgl. [7])

Sobald die Arbeit mit verschiedenen Servern und Bibliotheken funktioniert, lernen die Teilnehmer, wie sie sich die nun verfügbaren Ausgangsdaten im SAS Enterprise Guide (EPG) verfügbar machen und dort damit weiter arbeiten können. Für den Einstieg in den EPG wird auf ein von SAS angebotenes Onlinetutorium zurückgegriffen. Im Anschluss daran wird anhand von vorbereiteten Projekten das Design und die Umsetzung von ETL-Prozessen vermittelt. Hierbei wird ein Schwerpunkt auf die Import- und Transformationsmöglichkeiten und die hierbei einzustellenden Parameter gelegt. Ziel dieses Teils der Veranstaltung ist es, dass die Studenten anschließend in der Lage sind, einen exemplarischen ETL-Prozess wie den in Abbildung 8 eigenständig zu realisieren und als Input für weitere Lehrinhalte eine Fakten- und mehrere zugehörige Dimensionstabellen aus einer gegebenen Datenbank sowie mehreren Excel- und Textdateien zu erstellen.





**Abbildung 8:** Exemplarischer ETL-Prozess (vgl. [7])

Anschließend wird das SAS OLAP Cube Studio vorgestellt, um aus den Fakten- und Dimensionstabellen einen Würfel aufzubauen. Die Auswirkungen der Einstellmöglichkeiten in den einzelnen Schritten des Cube Designers werden erläutert und das Wissen vermittelt, welches die Teilnehmer benötigen, um auch ohne Unterstützung einen Würfel zu bauen. An dieser Stelle wird ein Bezug zu dem aus der Vorlesung bekannten Theoriewissen über das Verhältnis von Dimensionen, Ebenen und Hierarchien hergestellt. Ein kurzer Exkurs zu dem im Hintergrund des Cube Designers erstellten PROC OLAP Code schließt diesen Teil der rechnergestützten Übung ab.

Den letzten Abschnitt der Veranstaltung stellt die Betrachtung und Abfrage des erstellten OLAP-Würfels dar. Die Studenten erfahren, wie sie sich mit dem Enterprise Guide zu einem erstellten Würfel verbinden können und wie sie die Oberfläche des EPG nutzen, um innerhalb des Würfels zu navigieren (vgl. Abbildung 9).

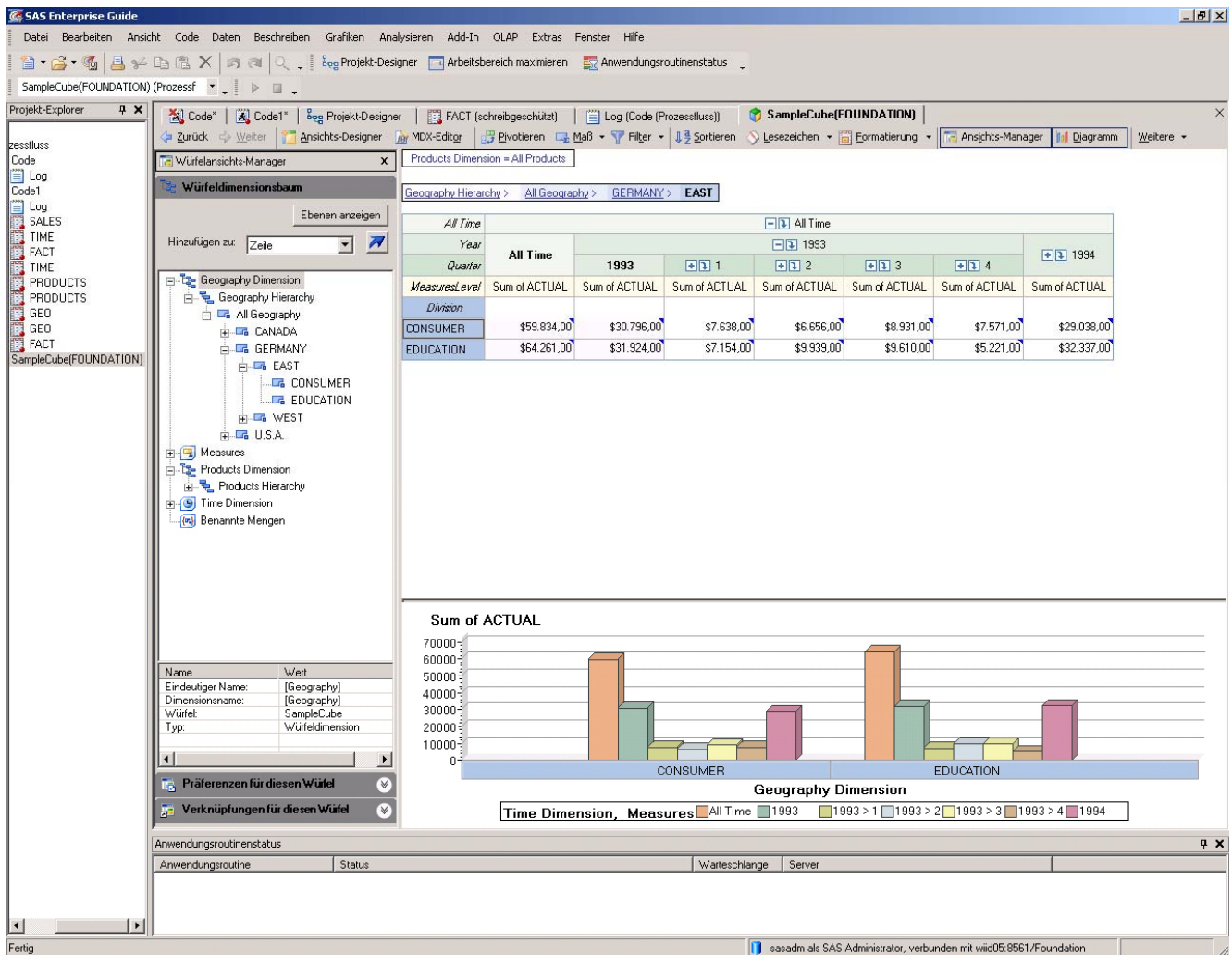


Abbildung 9: OLAP-Cube im Enterprise Guide (vgl. [7])

Die Anwendung der in der Vorlesung vorgestellten Abfragesprache MDX auf den erstellten Würfel und das Erzeugen von zusätzlichen Kennzahlen mittels MDX bilden den Abschluss der rechnergestützten Übung.

### 3.2 Rechnerübung Data, Web und Text Mining

Die rechnergestützte Übung zum Data Web und Text Mining war bis dato zweigeteilt in Data Mining sowie Web und Text Mining. Aufgrund der teilweisen Methodenüberschneidungen und der Umstellung auf Bachelor- und Master-Studiengänge werden beide Veranstaltungen ab dem Wintersemester 08/09 zusammengelegt.

Die Veranstaltung startet mit einer allgemeinen Einführung in das Data Mining und motiviert die Thematik mit anschaulichen Beispielen und Anwendungsmöglichkeiten. Anschließend werden die Oberflächen von SAS Base und dem SAS Enterprise Miner vorgestellt.

Nachdem die Grundlagen zur Bedienung des Tools gelegt sind, werden in den weiteren Veranstaltungen die einzelnen Schritte einer kompletten Data-Mining-Analyse mit den

Studenten durchlaufen, wobei die Schwerpunkte jeweils auf der Datenaufbereitung und explorativen Analyse sowie der Anwendung der Modelle liegen.

Im ersten Schritt lernen die Studenten das Einlesen unterschiedlichster Datenformate und das Zusammenführen von Datensätzen, die dann im weiteren Verlauf der Veranstaltung analysiert werden sollen. Darauf aufbauend werden die einzelnen Knoten mit den möglichen Einstellungsparametern zum Aufbereiten und Transformieren der Daten vorgestellt. Die Erläuterungen zu den einzelnen Methoden erfolgen immer beispielbezogen – wobei der Praxisbezug in Form von konkreten Übungen direkt umgesetzt wird. Nachdem die Studenten zum Beispiel Bondaten eines Supermarkts explorativ aufbereitet haben, führen sie eine Assoziations- und eine Linkanalyse durch. Besonders wichtig ist in diesem Zusammenhang die Diskussion und Interpretation der Ergebnisse in der Gruppe. Bezogen auf den genannten Fall entstehen Handlungsempfehlungen zur Layoutplanung für den Supermarktleiter. Die oben angesprochenen Knoten eignen sich besonders gut für den Einstieg in die weiteren Analysemethoden, da die visuelle Darstellung der Ergebnisse für die Studenten sehr anschaulich und zu einem großen Teil intuitiv interpretierbar ist.

The screenshot displays the SAS Text Miner interface. At the top left, there is a button labeled 'KL\_06\_07. ANZEIGEN' and a 'Text Miner' icon. The main window is titled 'Text Miner Settings' and is divided into three tabs: 'Parse', 'Transform', and 'Cluster'. The 'Parse' tab is active, showing options for text location, language (set to 'GERMAN'), and term identification. The 'Identify as Terms' section has several checked options: 'Words occurring in a single document', 'Same word as different part of speech', 'Stemmed words as root form', and 'Noun groups'. Below this, there are 'Initial word lists' and 'Synonyms' sections. On the right side, a table titled '4,792 Terms' is displayed, showing a list of terms with their frequencies and document counts. A 'Clusters' table is also visible, showing six clusters of terms with their respective frequencies and percentages. A red circle highlights the 'Clusters' table.

Term	Freq	# Documents	Keep	Weight	Role	Attribute
+ erfahrung	59	27	Y	0.081	Noun	Alpha
wirtschaft	29	21	Y	0.136	Noun	Alpha
+ team	34	21	Y	0.162	Noun	Alpha
+ fähigkeit	47	20	Y	0.157	Noun	Alpha
	24	19	Y	0.157	Noun	Alpha
	19	19	Y	0.134	Noun	Alpha
	23	15	Y	0.220	Noun	Alpha
	26	14	Y	0.258	Adj	Alpha
	24	13	Y	0.293	Noun	Alpha
	15	13	Y	0.258	Noun	Alpha
	30	12	Y	0.347	Noun	Alpha
	32	11	Y	0.381	Noun	Alpha
	14	11	Y	0.322	Noun	Alpha
	14	10	Y	0.370	Noun	Alpha
	9	9	Y	0.354	Noun	Alpha

#	Descriptive Terms	Freq	Percentage	RMS Std.
1	+ system, + projekt, + leistung, informatik, + tätigkeit	4	13%	0.183259867
2	+ bank, beratung, ausland, reise, + leistung	5	17%	0.1045398047
3	+ marketing, data mining, daten, studium, + leistung	6	20%	0.1397432357
4	leitung, führung, + projekt, + information, informatik	5	17%	0.1277160249
5	finance, denken, + einsatz, + international, studium	3	10%	0.0706217093
6	studium, + system, beratung, wirtschaft, + international	7	23%	0.1699386728

Abbildung 10: Text-Miner-Knoten (vgl. [8])

Den weiteren Schwerpunkt der rechnergestützten Übung stellt die Betrachtung und die Anwendung der Neuronalen Netze, der Regressionsanalyse, der Entscheidungsbäume und des Clusterings dar, wobei mit Daten aus unterschiedlichen Bereichen gearbeitet

wird, wie z. B. Kundendaten eines Telekommunikationsunternehmens, Daten eines Onlineauktionshauses oder auch Logfiles eines Webservers. Den Abschluss der Veranstaltung bilden die Konfiguration (vgl. Abbildung 10) und der Einsatz des Text-Miner-Knotens.

Aufgebaut ist diese rechnergestützte Übung als Blockveranstaltung über vier Tage mit jeweils fünf Zeitstunden. So verfügen alle Studenten am Anfang des Semesters über den gleichen Wissensstand. Während des Semesters existiert ein wöchentliches Tutorium, in dem Übungsaufgaben besprochen und etwaige Fragen zur semesterbegleitenden Projektleistung geklärt werden. Diese Projektleistung ist, wie in Abbildung 11 dargestellt, beispielsweise eine Logfileanalyse, die von den Studenten in kleinen Gruppen während des Semesters bearbeitet wird.

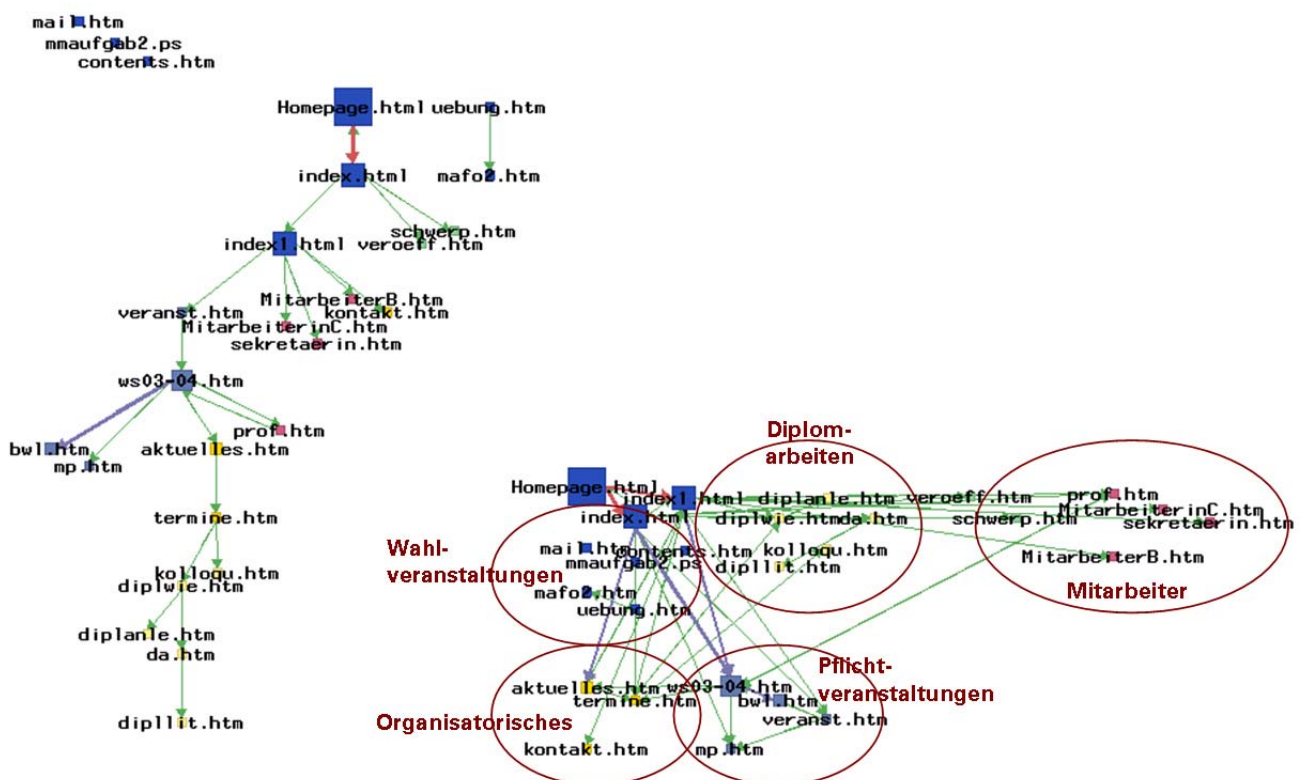


Abbildung 11: Grafiken zur Web Mining Fallstudie (vgl. [8])

Ziel dieser Fallstudie ist es, eine deskriptive Analyse durchzuführen sowie Benutzer, Sessions und Transaktionen zu identifizieren.

## 4 Projektseminare

Den Studenten, die die vorgestellten Veranstaltungen besucht haben, bietet sich an der Professur die Möglichkeit, das erworbene Wissen in Projektseminaren anzuwenden. Hierzu gibt es jedes Semester neue Themen, die teilweise auf Problemstellungen lokaler Wirtschaftspartner zurückgehen und so einen echten Praxisbezug schaffen.

Die Teilnahme am jährlich stattfindenden Data-Mining-Cup, die ebenfalls in Form eines Projektseminars angeboten wird, ist ein weiteres Angebot an die Studenten, das es ihnen

ermöglicht, erlerntes Wissen nicht nur auf reale und praxisnahe Fragestellungen anzuwenden, sondern sich auch mit anderen Studenten und Universitäten zu messen. Beim Data-Mining-Cup 2006 konnte ein Teilnehmer bereits einen 20sten Platz erzielen, was einer Topplatzierung bei diesem internationalen Wettbewerb entspricht.

## **5 Fazit**

Das vorgestellte integrierte Lehrkonzept unterliegt stetigen Anpassungen an neueste Forschungsergebnisse und aktuelle Softwareversionen. Das Grundkonzept bleibt jedoch gleich und hat sich aus Sicht der Professur seit mehreren Semestern bewährt. Durch das Feedback der Studenten im Anschluss an die Veranstaltungen einerseits und die Rückmeldungen der Alumni, die nicht zuletzt wegen ihrer ausgeprägten Business-Intelligence-Kenntnisse hochdotierte Arbeitsstellen gefunden haben, andererseits werden sowohl der Erfolg dieses Aufbaus als auch die Nachhaltigkeit der Ausbildung bestätigt.

## **Literatur**

- [1] Kemper, H.-G.; Mehanna, W.; Unger, C.: Business Intelligence - Grundlagen und praktische Anwendung, 2. Aufl., Vieweg Verlag, Wiesbaden, 2006.
- [2] Gluchowski, P.: Business Intelligence, HMD - Praxis der Wirtschaftsinformatik, Bd. 38, N. 222, dpunkt.verlag, Heidelberg, 2001, S. 5–15.
- [3] Hilbert, A., Schönbrunn, K.: Business Intelligence im strategischen Hochschulmanagement, unveröffentlichtes Arbeitspapier, Dresden, 2007.
- [4] Hilbert, A.: Skript zur Vorlesung Business Intelligence, insb. Data Warehousing, unveröffentlicht, Dresden, 2007.
- [5] Fayyad, U. M.; Piatetsky-Shapiro, G.; Smyth, P.: From Data Mining to Knowledge Discovery: An Overview, in: Fayyad, U. M.; Piatetsky-Shapiro, G.; Smyth, P.; Uthurusamy, R. (Hrsg.): Advances in Knowledge Discovery and Data Mining, AAAI Press, Cambridge, MA, 1996, S. 1–34.
- [6] Hilbert, A.: Skript zur Vorlesung Business Intelligence, insb. Data Mining, unveröffentlicht, Dresden, 2007.
- [7] Müller, A. E.: Skript zur Rechnerübung Data Warehousing, unveröffentlicht, Dresden, 2008.
- [8] Schönbrunn, K.: Skript zur Rechnerübung Data Mining, unveröffentlicht, Dresden, 2008.