

# DaMiT: Data Mining lernen und lehren

**Klaus P. Jantke**  
DFKI GmbH  
Stuhlsatzenhausweg 3  
66123 Saarbrücken  
jantke@dfki.de

**Gunter Grieser**  
TU Darmstadt  
Alexanderstr. 10  
64283 Darmstadt  
grieser@informatik.tu-darmstadt.de

**Steffen Lange**  
FH Darmstadt  
Haardtring 100  
64295 Darmstadt  
s.lange@fbi.fh-darmstadt.de

**Martin Memmel**  
TU Kaiserslautern  
PF 3049  
67653 Kaiserslautern  
mommel@dfki.uni-kl.de

## Abstract

Das Tutorsystem DaMiT stellt Wissen über das Gebiet des Data Mining auf neuartige Weise zur Verfügung. Es repräsentiert sogen. Content – nämlich Wissen über Grundlagen, Prinzipien und Verfahren des Data Mining inkl. ausgearbeiteter Fallstudien und Bewertungen kommerzieller Data Mining-Tools –, der gleichermaßen in der akademischen Ausbildung, in der beruflichen Weiterbildung, im lebenslangen Lernen und für Entscheidungsprozesse in Politik und Wirtschaft relevant ist.

DaMiT ist ein integriertes System, welches, getrieben vom Paradigma des „Learning by Doing“, das Studium des Data Mining organisch mit seinen Anwendungen verbindet. DaMiT ist adaptierbar in dem Sinne, daß die Benutzer ihre jeweils favorisierte Sicht auf die Inhalte leben können. DaMiT ist adaptiv, da es sich den Bedürfnissen des Benutzers hinsichtlich des Präsentationsstils anpaßt. Für die komplexen Anwendungsübungen der Lernenden hat DaMiT ein eigenes Konzept hervorgebracht, entwickelt und implementiert, die sogenannten „Competitive Exercises“. Mit DaMiT kann man nicht nur Data Mining studieren, man kann es praktizieren. DaMiT geht über die Funktionalität konventioneller Systeme des E-Learning hinaus und betritt direkt den Marktplatz Internet, ausgerüstet mit einer integrierten E-Payment-Funktionalität und einer Anbindung an moderne IT-Sicherheitsinfrastrukturen.

## 1 Data Mining

Mit dem Tutorsystem DaMiT kann man Data Mining lernen. Was ist Data Mining und wie kann man es am besten lehren? Die Antworten auf diese Fragen machen deutlich, daß Data Mining ein Themengebiet ist, das für den Einsatz von E-Learning-Angeboten – sowohl in der akademischen Ausbildung als auch in der beruflichen Weiterbildung und dem lebenslangen Lernen – geradezu prädestiniert ist.

### 1.1 Was ist Data Mining?

Beim Data Mining geht es darum, Modelle zu bilden, die Regularitäten und Zusammenhänge in großen Datenmengen erklären (vgl. [WF99]).

Heutzutage kann erfolgreiches Data Mining – nicht zuletzt angesichts der Flut der zur Verfügung stehenden Daten – nur noch in Interaktion von Mensch und Maschi-

ne stattfinden. Im Prozeß der interaktiven Modellbildung leisten beide Partner das, was sie am besten können: Die Maschine schlägt zum Teil hochkomplexe Modelle vor und erstellt quantitative Bewertungen dieser Modelle. Der Mensch analysiert die Qualität und Nützlichkeit der vorgeschlagenen Modelle und akzeptiert bzw. verwirft sie. Durch eine geeignete Auswahl und Vorverarbeitung der zu berücksichtigenden Daten und die richtige Wahl und Parametrisierung der Methoden und Verfahren, die zur automatischen Modellbildung eingesetzt werden, greift der Mensch steuernd ein und legt fest, welche Modelle überhaupt in die engere Wahl kommen.

Abbildung 1 stellt das von der CRISP-DM Special Interest Group entwickelte Prozeßmodell vor. Sie zeigt die einzelnen Phasen des Data Mining-Prozesses und illustriert die Beziehungen und Abhängigkeiten zwischen ihnen.

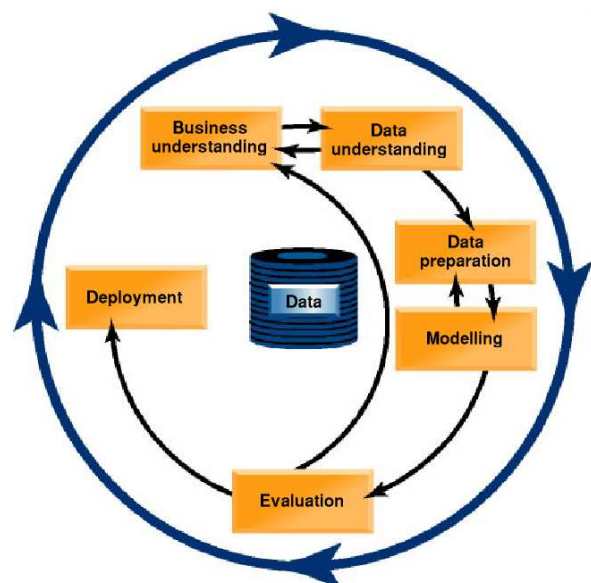


Abb. 1: Die einzelnen Phasen des Data Mining-Prozesses

Data Mining ist Kunst und Wissenschaft zugleich. Es bedarf eines reichhaltigen Reservoirs an theoretisch fundierten Methoden und Methodologien zur Datenvorverarbeitung, automatischen Modellbildung, quantitativen Bewertung von Modellen und Unterstützung des Menschen bei der Analyse der Qualität der vorgeschlagenen Modelle. Darüber hinaus hat der Mensch im Prozeß der interaktiven Modellbildung notwendigerweise eine Vielzahl von ganz essentiellen Entscheidungen zu treffen, bei denen nur Erfahrung, Intuition und – zu einem gewissen Teil – auch Glück helfen können.

## 1.2 Was heißt es, Data Mining zu lehren?

Um erfolgreich Data Mining praktizieren zu können, bedarf es neben einschlägigem Wissen über das Reservoir der zur Verfügung stehenden Methoden und Methodologien ganz entscheidend eines tieferen Verständnisses vom Ursprung, der Bedeutung und der Aussagekraft der zur Verfügung stehenden Daten.

Üblicherweise sind diejenigen, die sich mit den Daten und der intendierten betrieblichen Nutzung der gesuchten Modelle auskennen, keine Spezialisten, die die zur Verfügung stehenden Methoden und Methodologien bis ins letzte verstehen und ausreizen können. Denjenigen, die das intellektuelle Know-How mitbringen, um diese Methoden und Methodologien zu durchdringen, fehlt oft das anwendungsspezifische Hintergrundwissen und die Erfahrung bei der Lösung betrieblicher Aufgaben.

Data Mining „richtig“ zu lehren bedeutet, beide Aspekte zu berücksichtigen. Gelingt dieses, eröffnet sich – quasi ganz nebenbei – die Möglichkeit, mit ein und demselben Inhalt unterschiedliche Gruppen von Lernenden zu adressieren: solche, die im Rahmen ihrer akademischen Ausbildung an Hochschulen und Universitäten theoretisch fundiertes, anwendungsnahe Wissen erwerben wollen, und solche, die im Rahmen der beruflichen Weiterbildung und des lebenslangen Lernens mit der Thematik Data Mining konfrontiert sind. Der Zugang zu den Inhalten und deren Darstellung ist entsprechend an die Bedürfnisse dieser recht heterogenen Gruppen anzupassen.

Will man Data Mining „richtig“ lehren, ist u.a. Wissen zu vermitteln über

- die relevanten Grundlagen aus den Gebieten Mathematik, Statistik und theoretische Informatik,
- die prinzipielle Funktionsweise, die Parametrisierung und die Vor- und Nachteile von Standardalgorithmen des maschinellen Lernens und der Statistik,
- Ansätze zur Kombination solcher Verfahren,
- mögliche Ansätze zur Beurteilung der Güte der automatisch generierten Modelle und
- grundlegende IT-Konzepte zur Datenhaltung.

Darüber hinaus sind u.a. Fähigkeiten auszuprägen, die relevant sind

- für eine Benutzung von Data Mining-Werkzeugen und den in ihnen realisierten Varianten der Standardverfahren des maschinellen Lernens und der Statistik und
- für ein erfolgreiches Durchlaufen der einzelnen Phasen des Data Mining-Prozesses bei der Bearbeitung praxisrelevanter Aufgaben, inkl. der Interpretation der Ergebnisse und der Beurteilung ihrer Qualität und Aussagekraft.

## 2 DaMiT – Der Data Mining-Tutor

Das DaMiT-Konsortium, bestehend aus 10 Universitäten und Fachhochschulen aus ganz Deutschland, ist angetreten, diese hochkomplexe Materie in einem elektronischen Tutorsystem verfügbar zu machen. Die Benutzer sollen in die Lage versetzt werden, mittels DaMiT sowohl die theoretischen Hintergründe zu erlernen als auch praktische Fähigkeiten im Umgang mit realen Data Mining-Problemen und -Werkzeugen zu erwerben.

DaMiT stellt natürlich die klassischen Lehrbuchinhalte zum Data Mining (d.h. Grundlagen, Algorithmen und Prozesswissen) zur Verfügung. Ein Online-System bietet je-

doch – verglichen mit einem Lehrbuch – eine Reihe zusätzlicher Möglichkeiten, von denen einige wichtige in DaMiT bereits genutzt werden. So können aktuelle Lerninhalte unmittelbar in das System integriert und wissenschaftliche Erkenntnisse sofort nach deren Bekanntwerden zur Verfügung gestellt werden. Mehr noch, die relevanten Originalquellen können darüber hinaus gleichzeitig zugänglich gemacht werden.<sup>1</sup> So kann etwa ein Student PDF-Files der Originalliteratur herunterladen. Während sich Lehrbücher – nicht zuletzt aus Platz- und Marketinggründen – darauf beschränken, ein bestimmtes Standardrepertoire vollständig und dafür weniger in die Tiefe gehend abzudecken oder nur einige ausgewählte Themen sehr detailliert zu behandeln, bietet ein E-Learning-System Raum für beides. In DaMiT findet man neben einer umfassenden Darstellung der klassischen Data Mining-Verfahren beispielsweise auch die JSM- und die FCA-Methode [BDF<sup>+</sup>03; GW99], die in Deutschland nahezu unbekannt sind. Darüber hinaus werden im Rahmen von sogenannten Fallstudien Anwendungen des Data Mining detailliert vorgestellt, die in den einschlägigen Lehrbüchern komplett fehlen. In DaMiT werden u.a. Anwendungen im Bereich des Music Mining diskutiert. Dabei geht es darum, die Besonderheiten in der Spielweise einzelner Pianisten aufzudecken und die gewonnenen Erkenntnisse auszunutzen, um Computern beizubringen, wie sie ausdrucksvoller musizieren können (vgl. [Wid02]). Daß sich der Lernende die auf diesem Weg erzielten Verbesserungen in der Spielweise von Computern auch noch anhören kann, soll an dieser Stelle nicht unerwähnt bleiben.

Online-Lernen spielt seinen Reiz jedoch erst in Verbindung mit Computeranwendungen so richtig aus. In DaMiT kann man – was mit keinem Buch geht – aktuelle Verfahren des Data Mining ausprobieren. Der Benutzer hat die Möglichkeit, sowohl Implementierungen einzelner Data Mining-Verfahren als auch komplette Data Mining-Werkzeuge herunterzuladen und den Umgang mit ihnen zu trainieren. Des weiteren stehen Datensätze zur Illustration der Stärken und Schwächen einzelner Data Mining-Verfahren sowie umfangreiche Datensätze zum Training der einzelnen Phasen des Data Mining-Prozesses und zur Erprobung des Umgangs mit Data Mining-Werkzeugen zur Verfügung. Auch bei der Entscheidung, welches Data Mining-Werkzeug eingesetzt werden soll, bietet DaMiT Unterstützung: Für viele kommerzielle Systeme findet sich eine Beschreibung ihres Leistungsumfanges in DaMiT.

Data Mining läßt sich nur anhand praktischer Beispiele erlernen. In DaMiT steht das Prinzip des Learning by Doing im Mittelpunkt. Simulationen und Animationen laden zum Ausprobieren ein, in Videos wird die Bedienung von Data Mining-Werkzeugen erklärt. Anhand komplexer, praxisnaher Wettbewerbsaufgaben (den sogenannten „Competitive Exercises“) wird den Lernenden das Zusammenspiel der verschiedenen Phasen des Data Mining-Prozesses nahegebracht. Die Qualität der eingereichten Lösungen kann – da es im allgemeinen weder richtige noch falsche und insbesondere nicht das beste Modell gibt – nur im Vergleich mit den Lösungen anderer Lerner und mit Musterlösungen bestimmt werden. Der so entstehen-

<sup>1</sup>In diesem Zusammenhang waren Copyright-Fragen zu berücksichtigen, die sich aus dem beabsichtigten Verwendungszweck (Forschung/Lehre bzw. Wirtschaft) ergeben und in unterschiedlichen Zugangsmöglichkeiten (und deren Kontrolle) resultieren. Diese Aspekte wurden im realisierten Rollen- und Zugangskonzept umgesetzt.

de Wettbewerbscharakter bietet einen weiteren Anreiz für die Bearbeitung der Aufgabe.

Die Spannweite der potentiellen Benutzer ist sehr groß. Auf der einen Seite steht der klassische akademische Adressatenkreis, zunächst vornehmlich Studenten der Informatik und angrenzender Disziplinen. Diese Zielgruppe wird DaMiT in der Form des „Blended Learning“ etwa als Ergänzung zu klassischen Vorlesungen (sowohl für Spezialvorlesungen über Data Mining oder Maschinelles Lernen bzw. für Grundvorlesungen über Künstliche Intelligenz) benutzen. Dem gegenüber stehen die Benutzer, die sich stärker problemorientiert der Thematik Data Mining nähern und vor allem an Methodenwissen interessiert sind. Zu dieser Gruppe zählen etwa Geistes- und Sozialwissenschaftler aus dem akademischen Umfeld und Benutzer aus der betrieblichen Praxis, bei denen man nicht notwendig einen akademischen Hintergrund voraussetzen kann. Schließlich gibt es noch andere Benutzergruppen, die adressiert werden, etwa Entscheidungsträger aus der Wirtschaft, die sich nur einen groben Überblick über die Möglichkeiten und Potentiale des Data Mining verschaffen wollen. DaMiT unterstützt alle diese unterschiedlichen Bedürfnisse, indem es angepaßte Zugänge zu den entsprechend der unterschiedlichen Bedürfnisse der Benutzergruppen aufgearbeiteten Inhalten bereitstellt.

Die Attraktivität eines E-Learning-Systems steigt in dem Maße, wie es gelingt, dem Lernenden genau die für ihn relevante Information bereitzustellen. DaMiT ist ein adaptives Tutorssystem, das sich in gewissen Grenzen dem Lernstil des Benutzers anpaßt, indem es je nach Wunsch den Inhalt mehr formal oder mehr informal, mehr beispiel- oder mehr theorieorientiert darbietet. Abhängig vom Wissensstand des Lernenden werden bestimmte Themenkomplexe vorgeschlagen, die als zum Verständnis des gewählten Lerninhaltes notwendig erachtet werden. Die angestrebte benutzerorientierte Darbietung der Inhalte erfordert eine passende Systemarchitektur und ein adäquates Konzept zur Repräsentation und Annotation der Inhalte. Alle DaMiT-Inhalte sind in XML repräsentiert und in einer Datenbank abgelegt. Die Annotation der Inhalte erfolgt in Anlehnung an den Standard des IMS Global Learning Consortium (Details finden sich in [Mem03]). Inhalte, die dem Benutzer angeboten werden, werden dynamisch und zwar in Übereinstimmung mit den Benutzerpräferenzen erzeugt und angezeigt. Das E-Learning-System DaMiT wurde gemäß der Model/View/Control-Architektur aufgebaut und realisiert eine strikte Trennung der Daten von den Sichten auf die Daten.

DaMiT ermöglicht eine sichere Kommunikation durch eine Public-Key Infrastruktur (PKI). Beispielsweise wird über die PKI eine sichere Online-Registrierung neuer Anwender ermöglicht. Beim Login kann sich dann der Anwender sicher über SSL authentifizieren. Die Kommunikation zwischen dem Anwender und DaMiT kann durch Verschlüsselung in SSL-Kanälen vertraulich gestaltet werden, ebenso wie der E-Mail-Austausch. Signierte Applets erlauben die vertrauenswürdige Ausführung von Animationen und Simulationen auf dem eigenen Rechner. Schließlich lassen sich Aufgabenlösungen signiert einreichen, dies erlaubt einen Herkunftsnachweis und bildet die Grundlage für eine wirkliche curriculare Einbindung.

Einige der genannten Aspekte, nämlich die Zugangsmöglichkeiten, die didaktischen Konzeptionen, die Adaptivität, das Learning by Doing sowie die „Competitive Exercises“ werden im folgenden genauer diskutiert.

### 3 Zugang zum Inhalt

DaMiT bietet unterschiedliche Zugänge zu den Inhalten.

Zunächst ist der Lehrstoff in sogenannte Lektionen aufgeteilt. Diese Lektionen sind, ähnlich den Kapiteln in einem Buch, inhaltlich relativ abgeschlossen und definieren ein klares Lernziel. Die Lektionen sind, analog dem Inhaltsverzeichnis in einem Lehrbuch, hierarchisch zu einem Kurs zusammengefaßt. Die Lehrbuchmetapher greift jedoch zu kurz, da Teile von oder ganze Lektionen an mehreren Stellen in diesem Inhaltsbaum auftreten können, wenn sie beispielsweise theoretische Grundlagen für verschiedene Data Mining-Verfahren bereitstellen.

Die Lernenden können nun aus dem Inhaltsbaum (der genaugenommen ein Graph ist) eine Lektion auswählen und bearbeiten. Das System prüft anhand des Benutzermodells, inwieweit bestimmte Vorkenntnisse, d.h. andere Lektionen oder Teile davon, nötig sind, und schlägt dem Lernenden bei Bedarf vor, diese Inhalte zuvor durchzuarbeiten.

Der Zugang über den hierarchisch strukturierten Inhaltsbaum adressiert all diejenigen, die gewohnt sind, mit Büchern zu lernen. Alternativ steht hier auch der Zugang über das Glossar zur Verfügung. In einer Stichwortliste sind die wichtigsten Begriffe erläutert und mit Verweisen zu den relevanten Lektionen versehen, so daß der Lernende direkt zu den benötigten Lerninhalten navigieren kann.

Abbildung 2 auf der nächsten Seite illustriert den Zugang über das Glossar am Beispiel des Glossareintrags „ID3“, über den es möglich ist, direkt in die Lektion „Entscheidungsbaumlernen“ einzusteigen.

Benutzer aus der beruflichen Weiterbildung bzw. im lebenslangen Lernen ziehen meist einen problem- und nicht den inhaltsorientierten Zugang vor. DaMiT bietet eine repräsentative Reihe von Fallstudien an, die jeweils ein typisches Data Mining-Problem adressieren. Neben dem Problem wird ein möglicher Lösungsweg beschrieben, wobei quasi ganz nebenbei in diejenigen Lektionen verwiesen wird, die die theoretischen Hintergründe bzw. die angewendeten Methoden behandeln.

Zur leichteren Navigation finden sich auf einer Übersichtsseite kurze Zusammenfassungen mit Verweisen zu den relevanten Inhalten, so daß ein problemorientierter Benutzer sehr schnell die für ihn interessanten Fallstudien erkennen und zu den Inhalten verzweigen kann.

Schließlich bietet DaMiT noch eine Übersichtsseite mit den verfügbaren Softwarewerkzeugen an. Von hier gelangt man einerseits zu den Werkzeugen selbst sowie zu Erläuterungen zu ihrer Bedienung, andererseits auch zu den Lektionen, die Wissen über die in den Werkzeugen eingesetzten Verfahren vermitteln. Dieser Einstieg wird vor allem von Praktikern bevorzugt, die sich gezielt mit einem bestimmten Werkzeug und den dort verfügbaren Methoden vertraut machen wollen.

Je nach Hintergrund und Vorliebe des Benutzers erschließen sich die Inhalte von DaMiT auf unterschiedlichem Wege. DaMiT ist somit in dem Sinne adaptierbar, daß der Benutzer seine bevorzugte Sicht auf die Inhalte leben kann.

Um sich auf den Benutzer einstellen und seinen Lernerfolg messen zu können, müssen Daten über ihn gesammelt werden, die auch die jeweilige Sitzung überdauern müssen. Es ist also essentiell für die Adaptationsfähigkeit, den Benutzer identifizieren zu können. Hierzu bietet DaMiT den üblichen Anmeldemechanismus mittels Name und Paßwort an. Um die unterschiedlichen Hemmschwellen von Erst-

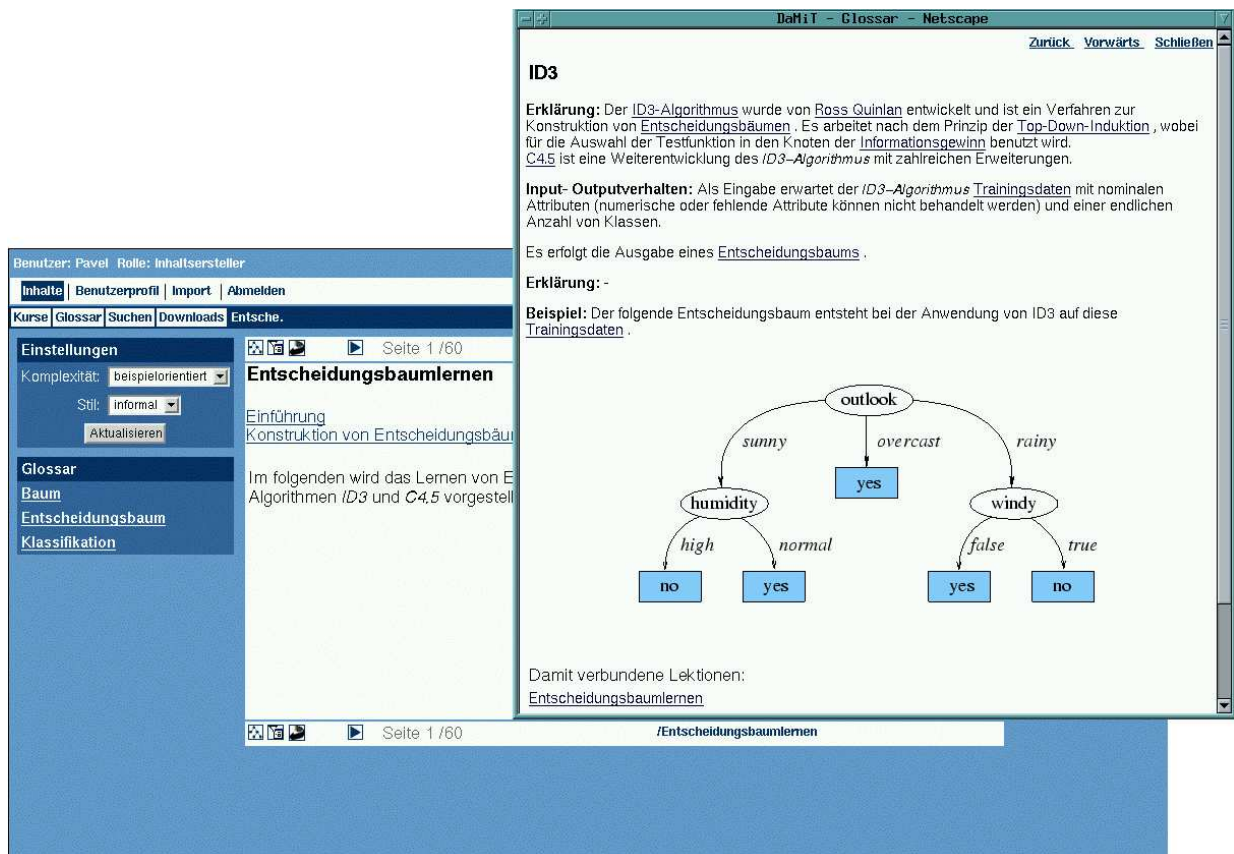


Abb. 2: Zugang zum Inhalt über einen Glossareintrag

benutzern abzubauen, existieren drei verschiedene Niveaus des Zugangs. Zunächst ist ein Gastlogin möglich, bei dem der Benutzer keinerlei Daten über sich preisgeben muß. Hierbei gehen jedoch alle Daten am Ende der Sitzung verloren. In der mittleren Stufe kann der Benutzer einen beliebigen Loginnamen wählen und ist darüber auch später wieder identifizierbar, muß aber keinerlei persönliche Daten preisgeben. In der höchsten Stufe ist der Benutzer persönlich bekannt, hierzu muß er sich persönlich registrieren. Dies erlaubt es, einen speziellen Status, wie z.B. Student, rechtssicher zu vergeben. Außerdem wird auf diese Weise gesichert, daß im System abgelegte Prüfungsleistungen anerkannt werden können.

Technisch gesehen werden den Logins bestimmte Rollen zugeordnet. Aus der Rolle ergeben sich dann bestimmte Rechte, sowohl den Inhaltszugang als auch den Umgang mit Systemfunktionalitäten betreffend. Beispielsweise ist nur persönlich bekannten Benutzern gestattet, Beiträge in die Diskussionsforen einzustellen. Darüber hinaus dürfen – aus Gründen des Copyright – Kopien wissenschaftlicher Artikel zunächst nur an Studenten verteilt werden. Des weiteren enthält DaMiT auch Inhalt von kommerzieller Bedeutung, beispielsweise Studien über den Vergleich von Data Mining-Werkzeugen oder direkten Zugang zu den Werkzeugen selbst. Manche Inhalte bzw. Funktionalitäten sind nur für Inhaber einer bestimmten Rolle verfügbar, andere können käuflich erworben werden. Hierzu ist in DaMiT ein elektronisches Bezahlssystem integriert. In späteren Ausbaustufen könnte DaMiT beispielsweise problemlos als Plattform zum Verkauf von Problemlösungen oder Werkzeugen benutzt werden.

Ein Lehrender, der DaMiT für seine spezielle Vorlesung

einsetzen will, kann das System ganz einfach seinen Bedürfnissen anpassen. Hierzu setzt er die Lektionen bzw. Teile davon beliebig zusammen und erzeugt so einen neuen Kurs. Selbstverständlich kann er auch neue Lektionen erzeugen bzw. die vorhandenen abändern. Schließlich kann er eine neue Benutzerrolle anlegen, so daß alle diese Benutzer seinen Kurs ausschließlich oder zusätzlich zum DaMiT-Kurs besuchen können. DaMiT ist auch in dem Sinne adaptierbar, daß ein Lehrender die Inhalte an seine Bedürfnisse anpassen kann.

Je nachdem, ob der Benutzer einen problem- oder inhaltsorientierten Zugang bevorzugt, müssen die Inhalte auf verschiedene Weise didaktisch aufbereitet werden. Diese didaktischen Unterschiede werden jedoch nicht an der Benutzerrolle festgemacht, sondern den Benutzervorlieben angepaßt.

#### 4 Didaktische Konzeptionen

Um komplexe Lerninhalte wie Data Mining zu vermitteln, ist eine sehr gute didaktische Konzeption notwendig. Dies gilt insbesondere dann, wenn mit Hilfe Neuer Medien und Technologien der Lernprozeß unterstützt und die Vorteile gegenüber klassischen Vermittlungsformen ausgeschöpft werden sollen. Da im Rahmen des Einsatzes des DaMiT-Systems Lernende nicht nur Wissen vermittelt bekommen sollen, sondern dieses auch auf praxisrelevante und eigene Probleme transferieren sollen, sind hierfür besondere didaktische Konzeptionen notwendig. Insbesondere der Einsatz internetbasierter Lernformen erfordert eine Erweiterung und Verbesserung bislang entwickelter Lernformen.

DaMiT ist ein tutorielles System, das interaktive Lernobjekte beinhaltet. Lernobjekte sind klassische Text- und

Multimedia-Objekte zur Darstellung von Data Mining-Inhalten oder Daten.

Die Strukturierung der Lernmodule wird mit einem Storyboard vorgenommen (vgl. [TD01]). Hierbei werden Organisation, Ablauf, benötigte Lernhilfen, Sozialformen und methodische Einzelelemente der Lehr-Lern-Prozeß-Struktur berücksichtigt. Lernziele werden explizit dargestellt, wobei allgemeinste Lernziele, Bestimmung weiterer Qualifikationen (z.B. Kommunikationsfähigkeit) und bereichsspezifische Konkretisierungen der allgemeinsten Lernziele und der weiteren Qualifikationen unterschieden werden. Für die Lernkontrolle werden spezifische interaktive Übungen, Praktika und Tests bereitgestellt, die die für das Data Mining üblichen Inhalte abdecken. Die Anwendung reicht dabei von tutoriellen bis hin zu kooperativen und kollaborativen Lernkontrollen.

Bei der Entwicklung eines Storyboards antizipiert ein Content-Ersteller mögliche Interaktionen des Lernenden mit dem System. Der Lernende ist natürlich nicht gezwungen, sich den Inhalt gemäß der „vorweggenommenen Pfade“ durch das System anzueignen. Die resultierenden Storyboards dienen als Grundlage für die Materialauswahl und die Gestaltung der einzelnen Lernobjekte. Sie explizieren die zugrundeliegenden didaktischen Konzeptionen, die dadurch selbst zum Diskussions- und Evaluationsgegenstand werden.

Storyboards werden in Form von hierarchischen kanten- und knotenbewerteten Graphen repräsentiert. Die annotierten Knoten repräsentieren einzelne Szenen, die wiederum aus Teilszenen bestehen können. Die annotierten Kanten beschreiben antizipierte Übergänge zwischen einzelnen Szenen (siehe etwa auch [Deg03] und [Mem03]).

Die Idee der Storyboards trägt noch weiter. Eine Analyse unterschiedlicher Storyboards zeigt, daß – zur Realisierung derselben didaktischen Absichten – häufig ähnlich strukturierte Teile in den Storyboards unterschiedlicher Content-Ersteller auftauchen. Solche Teile können – ähnlich wie Design-Patterns im modernen Software-Engineering – mit Hilfe von allgemeinen Mustern beschrieben werden, die beim Design neuer Storyboards wiederverwendet werden können. Um solcherart Muster verwalten und wiederverwenden zu können, bedarf es adäquater Werkzeuge. So kann zusätzliche Unterstützung bei der Herstellung didaktisch adäquat aufbereiteten Lehrmaterials geleistet werden.

## 5 Adaptivität

Ein zentrales Element eines erfolgreichen E-Learning-Angebots ist seine Adaptivität. Ein traditioneller „one size fits all“-Ansatz wird den Ansprüchen nicht gerecht. Vielmehr müssen die individuellen Erfahrungen, die Lernziele, die Voraussetzungen und das Vorwissen berücksichtigt werden.

[Spe99] unterscheidet vier Dimensionen, die bei der Entwicklung eines Adaptivitätskonzepten zu beachten sind:

- die Adaptationsmittel: Wo soll die Adaptivität zum Tragen kommen?
- die Adaptationsformen: Aufgrund welcher Daten erfolgt eine Anpassung?
- der Adaptationsprozeß: Wie erfolgt die Anpassung, benutzer- oder systemgesteuert?
- das Adaptationsziel: Warum erfolgt eine Anpassung?

Das DaMiT-System bietet sowohl funktionale als auch inhaltliche Adaptivität. Die funktionale Adaptivität spiegelt sich dadurch wider, daß bestimmte Funktionalitäten

nur dann zur Verfügung stehen, wenn der Benutzer in einer bestimmten Benutzerrolle agiert. Eine inhaltliche Adaptivität wird durch ein am IMS-Standard orientiertes Metadatenkonzept für Inhalte und Benutzer ermöglicht. Benutzer agieren in einer bestimmten Benutzerrolle, sie können Lernziele festlegen und einstellen, in welcher Art und Weise der Inhalt präsentiert werden soll. Der Benutzer kann zwischen einer beispilorientierten oder theorieorientierten und einer formalen oder informalen Darstellung der Inhalte wählen.

Der Unterschied zwischen einer formalen und informalen Darstellung spiegelt sich auf der Ebene der einzelnen Lernelemente wider, u.a. im Umfang der Darstellung und im verwendeten Vokabular.

Abbildung 3 auf der nächsten Seite illustriert den Unterschied zwischen der formalen und der informalen Darstellung am Beispiel eines Lernelements aus der Lektion „Entscheidungsbaumlernen“, in dem es um die Definition des Begriffs „Entscheidungsbaum“ geht.

Beispilorientierte und theorieorientierte Darstellungen unterscheiden sich dadurch, daß zur Vermittlung derselben Lehrinhalte verschiedene Lernelemente verwendet werden. Beispielsweise werden in einer beispilorientierten Darstellung eines Lernmoduls verstärkt Lernelemente mit einem illustrierenden bzw. erläuternden Charakter, Lernelemente mit einem eher theoretischen Charakter dagegen nur vereinzelt benutzt.

Der Benutzer des Systems kann gezielt auf einzelne Lernmodule zugreifen. Anhand der mit den Lernmodulen verknüpften Metadaten lassen sich alle zum Verständnis erforderlichen Lernelemente ermitteln. In Abhängigkeit vom prognostizierten Wissensstand des Benutzers werden daraus Vorschläge für eine – an den Bedürfnissen und Erfahrungen des Benutzers orientierte – Auswahl und Anordnung der relevanten Lerninhalte abgeleitet.

## 6 Learning by Doing

Beim Data Mining geht es darum, Modelle zur Erklärung von Zusammenhängen in großen Datenmengen zu finden. Versucht man, dies in einer Präsenzveranstaltung zu lehren, so wird man sich auf Spielbeispiele beschränken müssen.

Um DaMiT zu benutzen, muß man an einem Computer sitzen – und dies eröffnet großartige Möglichkeiten für die Wissensvermittlung. Der Lernende kann nun aktiv ins Geschehen eingreifen, er kann die Simulation eines Algorithmus steuern, unterbrechen, wiederholen, so oft wie er will. Er kann sich komplexe Beispiele für komplizierte Algorithmen vorrechnen lassen, und dies in der Gewißheit, keinem Flüchtigkeitsfehler aufzusitzen.

Die Verwendung eines Applets, welches ein Verfahren zum Lernen von Entscheidungsbaum über regulären Patterns realisiert, wird in Abbildung 4 gezeigt. Der Lernende kann positive und negative Beispiele eingeben bzw. automatisch erzeugen lassen. Er kann sich den vom implementierten Lernverfahren anhand dieser Beispiele erzeugten Entscheidungsbaum über regulären Patterns zusammen mit einer Beurteilung seiner Klassifikationsgüte auf diesen Beispielen anzeigen lassen.

Etwas allgemeiner betrachtet bieten solcherart Applets dem Lernenden die Möglichkeit, in einer für ihn neuartigen Form mit dem Computer zu interagieren: Der Lernende schlüpft im Prinzip in die Rolle des Lehrers. Er stellt dem Computer selbständig konstruierte Aufgaben, die dieser dann zu lösen hat. Der Lernende kann den Schwierigkeitsgrad der Aufgaben variieren, die vom Computer ge-

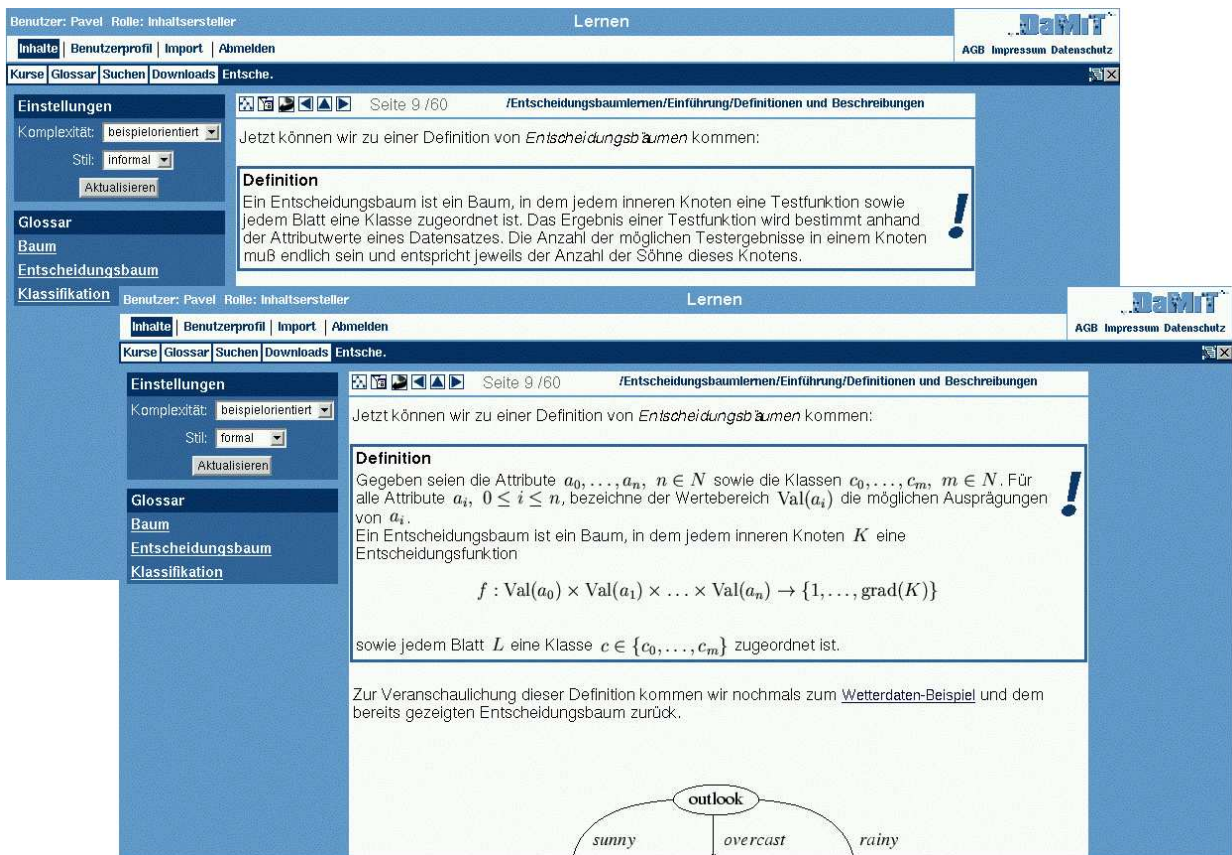


Abb. 3: Formale und informale Darstellung einer Definition

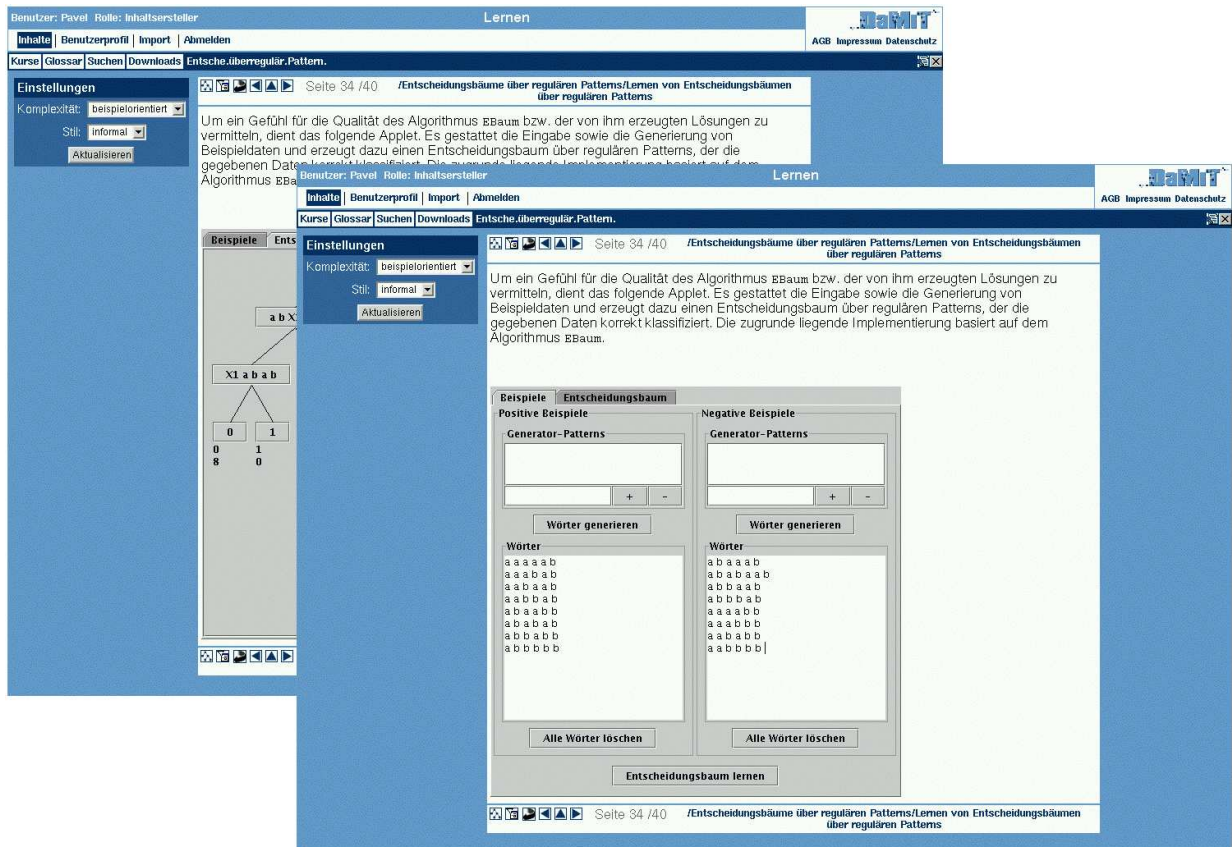


Abb. 4: Applet zur Generierung von Entscheidungsbaum

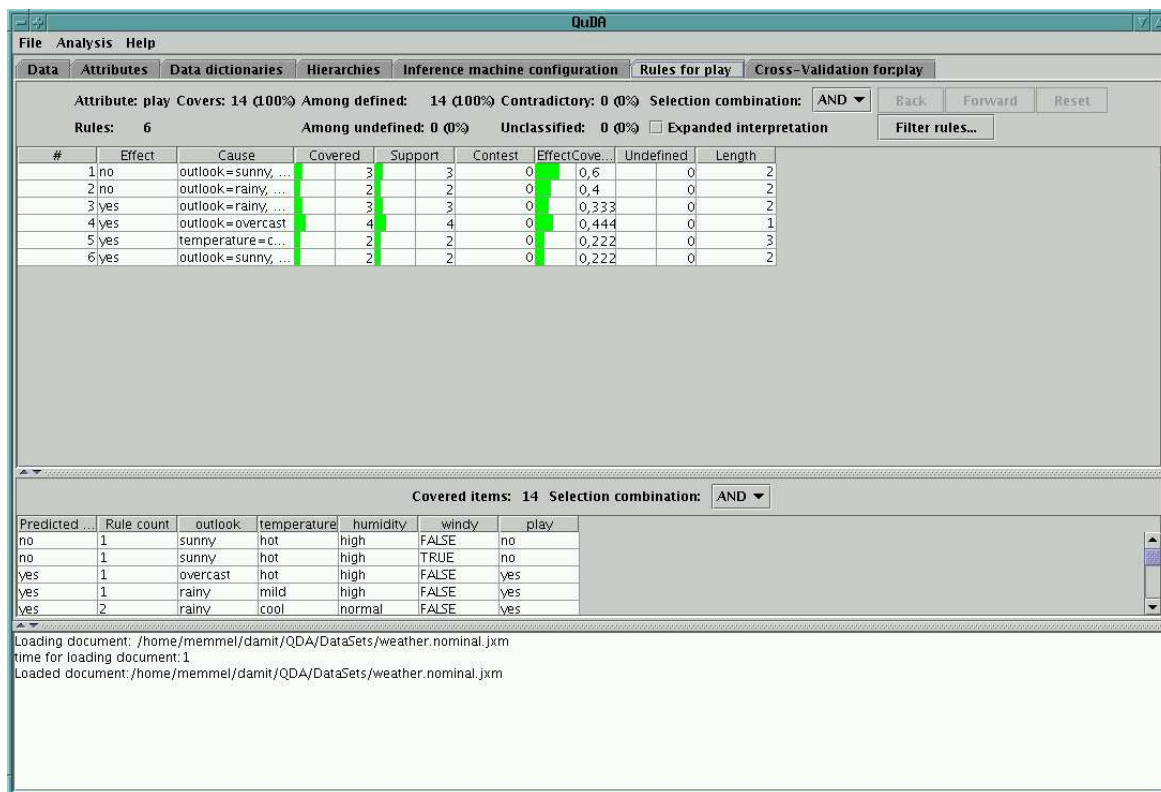


Abb. 5: Das im DaMiT-System angebotene Tool "QuDA"

fundenen Lösungen analysieren und mit seinen eigenen Erwartungen vergleichen. Solcherart Applets bieten ihm also eine alternative Möglichkeit, die Arbeitsweise und die Eigenheiten des implementierten Lernverfahrens besser verstehen zu können.

Die gleichzeitige Anregung verschiedener Perzeptionskanäle erlaubt es, viel Information sehr kompakt zu vermitteln. Der natürliche Spieltrieb sorgt dafür, daß der Benutzer auch wirklich konzentriert die angebotenen Möglichkeiten nutzt.

DaMiT unterstützt das „Learning by Doing“ auf verschiedene Weise. Zunächst wird die Inhaltsvermittlung, beispielsweise von Algorithmen, durch sogenannte interaktive Illustrationen begleitet. Hierbei handelt es sich um Java-Applets, die zunächst wie eine normale Abbildung wirken. Der Lernende kann dann aber beispielsweise die Simulation eines Algorithmus ablaufen oder einen Algorithmus auf bestimmten Daten „rechnen“ lassen. Der Benutzer kann (und muß) aktiv eingreifen, da keine vorgeplanten Sequenzen abgespielt werden. Der Lernende kann Parameter ändern, Ein- oder Ausgabewerte vorgeben etc. Die Applets werden nicht nur zur Illustration benutzt, sondern auch in die Übungsaufgaben integriert, etwa um Parameter herauszufinden, die zu bestimmten Effekten bei der Abarbeitung der Algorithmen führen.

DaMiT bietet eine Reihe professioneller Data Mining-Tools und Daten aus realen Anwendungen an, die in die Wissensvermittlung integriert sind. Die Benutzer lernen neben den jeweiligen Methoden auch gleich noch kennen, welche Varianten davon in realen Werkzeugen verfügbar sind und können sie bei der Verarbeitung dieser Daten erproben.

Abbildung 5 zeigt die Arbeit mit dem an der TU Darmstadt entwickelten Data Mining-Tool „QuDA“ [GY03], das

den Benutzern von DaMiT zum Download angeboten wird.

In späteren Versionen können Firmen ihre Daten in DaMiT für eine Analyse zur Verfügung stellen, je nach Bedarf allen oder nur ausgewählten Benutzern. DaMiT kann somit quasi als Marktplatz für wirtschaftlich verwertbare Lösungen dienen. Für die Benutzer ist neben dem kommerziellen Aspekt hier besonders attraktiv, daß sie ihre Urheberschaft an den Lösungen nachweisen können.

## 7 Competitive Exercises

Beim Data Mining sind interaktiv Modelle zu bilden. Allen praktisch relevanten Aufgaben ist gemein, daß es nicht „das richtige“ bzw. „die richtigen“ Modelle gibt. Die Aufgaben sind meist sehr komplex und unterschiedliche, zum Teil aus vielen Einzelschritten bestehende Wege können zum Ziel führen. Jeder dieser Wege hat zudem spezielle Stärken und Schwächen. Diesen Besonderheiten ist bei der Konzeption der Übungsaufgaben und Tests Rechnung zu tragen. Relativ kleine, abgeschlossene Übungsaufgaben und Tests wie Multiple-Choice-Questions und Fill-In-Blank-Questions werden den Erfordernissen nicht gerecht.

Im DaMiT-System findet sich deshalb neben den klassischen Übungsaufgaben ein auf die speziellen Anforderungen des Gebiets Data Mining unmittelbar zugeschnittener Aufgabentyp, die sogenannten „Competitive Exercises“ (siehe [SD02]).

Abbildung 6 soll einen Eindruck vom Charakter der in DaMiT eingebundenen „Competitive Exercises“ vermitteln. Die zu bearbeitende Aufgabe war Teil des Data Mining Cup 2001. Der Data Mining-Cup ist ein jährlich stattfindender internationaler studentischer Wettbewerb, in dem Studierende nationaler und internationaler Universitäten, Fachhochschulen und Berufsakademien aufgerufen sind,

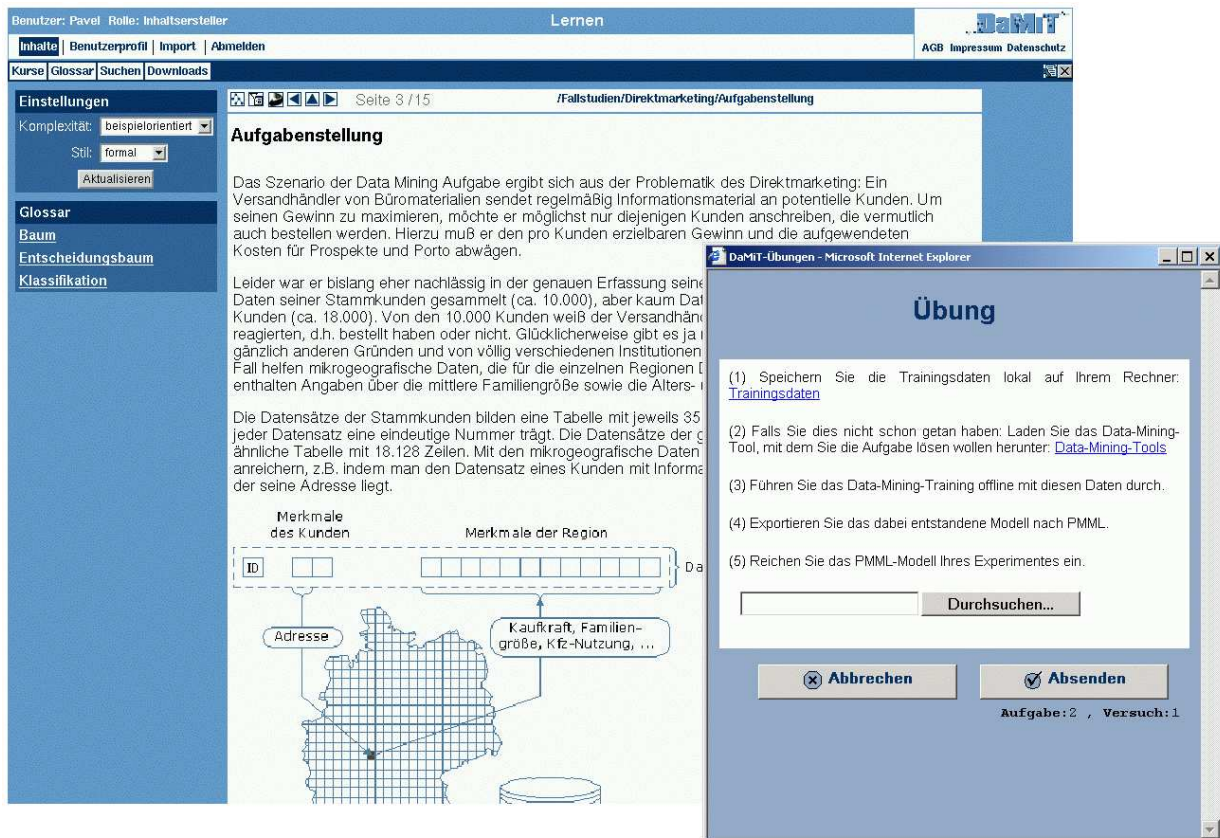


Abb. 6: Eine „Competitive Exercise“ im DaMiT-Lernsystem

sich an der Lösung eines konkreten Data Mining-Problems zu beteiligen.

Die Bearbeitung einer „Competitive Exercise“ stellt vielfältige Anforderungen an die Interaktion. Der Lernende hat sukzessive die folgenden Arbeitsschritte zu absolvieren:

1. sich mit der Aufgabenstellung vertraut machen, die Trainingsdaten herunterladen und ein Data Mining-Werkzeug auswählen,
2. die Trainingsdaten offline mit dem Data Mining-Werkzeug bearbeiten und ein Modell generieren (dies ist der eigentliche Data Mining-Schritt, der normalerweise aus mehreren Iterationen besteht),
3. das Modell als PMML<sup>2</sup>-Datei exportieren und
4. das PMML-Modell einreichen (in das DaMiT-System hochladen).

Beim Einreichen hat der Benutzer die Möglichkeit, seine Lösung digital zu signieren und verschlüsselt zu übertragen. Das ist von Bedeutung, wenn eine curriculare Einbindung erfolgen soll und prüfungsrelevante Leistungen zu erbringen sind.

Das System schätzt die Güte des eingereichten Modells ab, indem automatisch überprüft wird, wie gut das Modell zur Erklärung von bekannten, bei der Erzeugung des Modells nicht verwendeten Testdaten taugt. Da es ist nicht das bzw. die richtige(n) Modell(e) gibt, kann die Qualität der eingereichten Lösung jedoch nur im Vergleich mit

<sup>2</sup>Die *Predictive Markup Modeling Language* ist ein XML-basierter Standard zur Beschreibung von Data Mining-Modellen (siehe auch <http://www.dmg.org>). Er wird von den meisten Data Mining-Tools unterstützt.

bereits bekannten Lösungen ermittelt werden. Neben der Problematik, daß eine Lösung, die heute Spitze ist, morgen vielleicht nur noch mittelmäßig ist, hat diese Konstellation auch einen gewissen Charme. Lernende stehen auf ganz natürliche Weise in einem Wettbewerb, der – geschickt transparent gemacht – zu einer zusätzlichen Lernmotivation werden kann. Darüber hinaus bietet eine solche Wettbewerbssituation einen exzellenten Ausgangspunkt für eine produktive, inhaltsgetriebene Diskussion zwischen den Lernenden. In einem an eine „Competitive Exercise“ angebotenen Forum können Probleme und Lösungsvarianten diskutiert sowie Erfahrungen ausgetauscht werden. Kommunikatives und kollaboratives Arbeiten wird hier nicht nur unterstützt, sondern auch provoziert. Dieser Arbeitsstil ist gerade beim E-Learning von besonderer Bedeutung.

Schließlich erhalten die Lernenden, nachdem sie eine Lösung eingereicht haben, Zugriff auf eine Beispiellösung. Hier steht jedoch nicht so sehr die Lösung, sondern vor allem die Diskussion eines Lösungsweges und von Alternativen im Vordergrund.

## 8 Zusammenfassung

In dem vom BMBF im Rahmen des Zukunftsinvestitionsprogramms geförderten Projekt DaMiT wird Wissen über das Gebiet des Data Mining auf bisher nicht dagewesene Weise zur Verfügung gestellt. Grundlagen und Prinzipien des Data Mining sowie Verfahren des Data Mining, ausgearbeitete Fallstudien und Bewertungen kommerzieller Systeme werden innerhalb eines E-Learning-Systems angeboten. DaMiT ist zur Zeit deutschlandweit in der akademischen Ausbildung in Form des „Blended Learning“ im Einsatz. Das Einsatzspektrum



reicht von Spezialvorlesungen über Data Mining und Maschinelles Lernen bis zu Grundvorlesungen zur Künstlichen Intelligenz. Außerdem wird gerade der Einsatz in der Weiterbildung in Unternehmen vorbereitet.

Die Möglichkeiten des E-Learning-Systems DaMiT lassen sich – stark vereinfacht – wie folgt zusammenfassen:

- Mit DaMiT lernt man die Grundlagen und Anwendungen des Data Mining kennen.
- DaMiT ermöglicht berufliche Weiterbildung und lebenslanges Lernen.
- Mit DaMiT ist man immer aktuell darüber informiert, was Data Mining heute kann.
- DaMiT informiert über kommerzielle Systeme und deren Leistungsumfang.
- Mit DaMiT sind exemplarische Anwendungen von wirtschaftlicher Relevanz möglich.

DaMiT stellt nicht nur klassische Lehrbuchinhalte zur Verfügung. Es ist hochaktuell und bietet detaillierte Informationen zu Nicht-Standard-Verfahren und -Ansätzen des Data Mining. Sorgfältig aufgearbeitete Fallstudien unterstützen einen problemorientierten Zugang zum Thema Data Mining.

Data Mining kann nur anhand praktischer Beispiele erlernt werden. Die Applets zur Illustration von Data Mining-Verfahren und das Angebot, komplette Data Mining-Werkzeuge und diverse Datensätze herunterzuladen, gestatten es dem Benutzer, praktische Erfahrungen zu sammeln.

Will man mit einem E-Learning-System Lernende erreichen, genügt es nicht, Lehrbücher – seien sie noch so aktuell und didaktisch geschickt aufbereitet – ins Netz zu stellen. Der Lernende muß auch über andere Perzeptionskanäle angesprochen werden. Hierzu dienen Simulationen, Animationen und Videos. Im Rahmen von „Competitive Exercises“ werden die Lernenden mit praktisch relevanten Aufgaben konfrontiert. Die Lösungen der Lernenden werden mit denen anderer verglichen. Dadurch erhält der Lernende die Möglichkeit, die Qualität seiner Arbeit zu beurteilen und eigene Qualitätsmaßstäbe zu entwickeln. Ferner werden auf ganz natürliche Art und Weise Anknüpfungspunkte für die Kommunikation und Kooperation zwischen den Lernenden geboten.

Das DaMiT-System paßt sich an die Bedürfnisse des Benutzers an. Das spiegelt sich u.a. in den zur Verfügung gestellten Zugängen zum Inhalt, den angebotenen Funktionalitäten und in der Präsentation der Inhalte wider.

Die konsequente Realisierung des DaMiT-Systems gemäß der Model/View/Control-Architektur, der durchgängige Einsatz von Datenbanktechnologien, das Respektieren internationaler Standards sowie die realisierte Anbindung an ein E-Payment-System und an eine moderne Sicherheitsinfrastruktur garantieren, daß DaMiT auch in Zukunft sowohl curricular im Rahmen der akademischen Ausbildung als auch in der betrieblichen Lehre und im lebenslangen Lernen gewinnbringend eingesetzt werden kann. Fragen der Nachhaltigkeit, wie sie sich mit Blick auf die zukünftige Nutzung von E-Learning-Systemen stellen, werden im Beitrag [Deg03] am Beispiel von DaMiT diskutiert.

Die jeweils aktuelle Version von DaMiT ist unter der URL <http://damit.dfki.de> verfügbar.

## Literatur

- [BDF<sup>+</sup>03] Blinova, V.G., D.A. Dobrynin, V.K. Finn, S.O. Kuznetsov und E.S. Pankratova: *Toxicology Analysis by Means of the JSM-Method*. Bioinformatics, 19(10):1201–1207, 2003.
- [Deg03] Degel, G.: *Nachhaltigkeit von e-Learning Projekten. Erfahrungen und Erwartungen am Beispiel DaMiT*. In: Jantke, Wittig und Jörg (Hrsg): *Von e-Learning bis e-Payment 2003. Das Internet als sicherer Marktplatz*, S. 34–41. Akademische Verlags-Gesellschaft, 2003.
- [GLM03] Grieser, Gunter, Steffen Lange und Martin Memmel: *DaMiT: Ein adaptives Tutorssystem für Data-Mining*. In: Jantke, Wittig und Herrmann (Hrsg): *Von e-Learning bis e-Payment 2003. Das Internet als sicherer Marktplatz*, S. 192–203. Akademische Verlags-Gesellschaft, 2003.
- [GW99] Ganter, B. und R. Wille: *Formal Concept Analysis: Mathematical foundations*. Springer-Verlag, 1999.
- [GY03] Grigoriev, P. und S. Yevtushenko: *Elements of an Agile Discovery Environment*. In: Grieser, Tanaka und Yamamoto (Hrsg): *Proc. DS'2003, Lecture Notes in Artificial Intelligence* 2843, S. 309–316. Springer-Verlag, 2003.
- [Mem03] Memmel, M.: *Bausteine eines durchgängigen Workflows für die Inhaltserstellung in e-Learning-Systemen*. In: Jantke, Wittig und Herrmann (Hrsg): *Von e-Learning bis e-Payment 2003. Das Internet als sicherer Marktplatz*, S. 133–142. Akademische Verlags-Gesellschaft, 2003.
- [Rei02] Reidenbach, D.: *A negative result on inductive inference of extended pattern languages*. In: Cesa-Bianchi, Numao und Reischuk (Hrsg): *Proc. ALT'2002, Lecture Notes in Artificial Intelligence* 2533, S. 308–320. Springer-Verlag, 2002.
- [SD02] Strutz, J. und G. Degel: *Offene Übungsaufgaben und Praktika im e-Learning*. In: Jantke, Herrmann und Wittig (Hrsg): *Von e-Learning bis e-Payment. Das Internet als sicherer Marktplatz*, S. 410–420. Akademische Verlagsgesellschaft Aka, 2002.
- [Spe99] Specht, M.: *Adaptive Methoden in computerbasierten Lehr/Lernsystemen*. Doktorarbeit, Universität Trier, 1999.
- [TD01] Thalheim, B. und A. Düsterhöft: *Conceptual Modelling of Internet Sites*. In: Kunii, H.S., S. Jajodia und A. Solvberg (Hrsg): *Proc. ER'2001, Lecture Notes in Computer Science* 2224. Springer-Verlag, 2001.
- [WF99] Witten, I.H. und E. Frank: *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers, 1999.
- [Wid02] Widmer, G.: *In Search of the Horowitz Factor: Interim Report on a Musical Discovery Project*. In: Lange, Satoh und Smith (Hrsg): *Proc. DS'2002, Lecture Notes in Computer Science* 2534, S. 13–32. Springer-Verlag, 2002.