

# DaMiT: Ein adaptives Tutorssystem für Data Mining

Gunter Grieser

Technische Universität Darmstadt

Fachbereich Informatik

Alexanderstr. 10, 64283 Darmstadt

e-mail: grieser@informatik.tu-darmstadt.de

Steffen Lange

Deutsches Forschungszentrum für Künstliche Intelligenz

Stuhlsatzenhausweg 3

66123 Saarbrücken

e-mail: lange@dfki.de

Martin Memmel

Universität Kaiserslautern

Fachbereich Informatik

PF 3049, 67653 Kaiserslautern

e-mail: memmel@informatik.uni-kl.de

## Kurzfassung

Das Tutorssystem DaMiT stellt Wissen über das Gebiet des Data-Mining auf bisher nicht dagewesene Weise zur Verfügung. Es repräsentiert Content – nämlich Wissen über die 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 modernste IT-Sicherheitsinfrastrukturen.

## 1 Data Mining

Mit dem Tutorssystem DaMiT kann man Data-Mining lernen. Was ist eigentlich Data-Mining und wie kann man es am besten lehren? Diesen Fragen ist der folgende Abschnitt gewidmet.

Die Antworten auf diese Fragen sollen verdeutlichen, 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?

Das klassische Beispiel für erfolgreiches Data-Mining ist die Entdeckung der Gesetze zu den Bewegungen der Planeten – heute als Keplersche Gesetze bekannt. Johannes Kepler gelang es, Gesetzmäßigkeiten in den von Tycho Brahe, einem heute weitestgehend vergessenen dänischen Astronomen,

der pikanterweise Keplers Chef war, über Jahre hinweg gesammelten Daten aufzudecken und ein einfaches Modell zur Beschreibung der Planetenbahnen aufzustellen.

Beim Data-Mining geht es darum, Modelle zu bilden, die Regularitäten und Zusammenhänge in großen Datenmengen zu 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 Maschine 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 ihre Qualität und Nützlichkeit und akzeptiert bzw. verwirft sie. Durch eine adäquate Auswahl und Vorverarbeitung der zu berücksichtigenden Daten und die geschickte Auswahl der zur automatischen Modellbildung einzusetzenden Methode greift der Mensch steuernd ein und legt fest, welche Modelle überhaupt in die engere Wahl kommen.

Data-Mining findet in unterschiedlichen Bereichen statt – im Bereich des Customer Relationship Management (CRM), in der Medizin und der Pharmaforschung, aber auch bei der Terroristenfahndung in den USA, wo gegenwärtig gezielt Bevölkerungs- und Einwanderungsdaten mit Methoden des Data-Mining analysiert werden.

Data-Mining ist Kunst und Wissenschaft zugleich. Einerseits bedarf es eines reichhaltigen Reservoirs an theoretisch fundierten Methoden und Methodologien zur Datenvorverarbeitung, automatischen Modellbildung, quantitativen Bewertung von Modellen und Unterstützung des Menschen bei der Analyse ihrer Qualität. Im Prozeß der interaktiven Modellbildung hat der Mensch 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 ein tieferes Verständnis 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 Lehrinhalt unterschiedliche Gruppen von Lernenden zu adressieren: solche, die im Rahmen ihrer akademischen Ausbildung an Hochschulen und Universitäten theoretisch fundiertes anwendungsnahes Wissen erwerben wollen, und solche, die im Rahmen der beruflichen Weiterbildung und des lebenslangen Lernens mit dieser Thematik konfrontiert sind. Der Zugang zu und die Darstellung der Lehrinhalte ist sinnvollerweise an die Bedürfnisse dieser recht heterogenen Gruppen anzupassen.

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

- die relevante Grundlagen aus den Gebieten Mathematik, Statistik und theoretische Informatik,
- die prinzipielle Funktionsweise und die Vor- und Nachteile von Standardalgorithmen des maschinellen Lernens, einem Teilgebiet der Künstlichen Intelligenz, und der Statistik sowie über adäquate Ansätze zur Kombination solcher Verfahren,
- über mögliche Ansätze zur Beurteilung der Güte der automatisch generierten Modelle und
- über grundlegende IT-Konzepte zur Datenhaltung.

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

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

Letzteres setzt voraus, daß in einem gewissen Umfang Domänenwissen, etwa aus dem Bereich des CRM vermittelt wird.

## 2 DaMiT – Der Data-Mining-Tutor

Das DaMiT-Konsortium, bestehend aus 10 Universitäten und Fachhochschulen aus ganz Deutschland, ist nun 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 wie auch praktische Fähigkeiten im Umgang mit realen Data-Mining-Problemen und -werkzeugen zu erwerben.

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 relativ algorithmenorientierte Zielgruppe wird DaMiT im Zusammenhang mit klassischen Vorlesungsinhalten, sei es nun eine Spezialvorlesung über Maschinelles Lernen oder eine allgemeine Einführung in die Künstliche Intelligenz, benutzen.

Dem gegenüber stehen die problemorientierten Benutzergruppen, die vor allem an Methodenwissen interessiert sind. Hierbei handelt es sich einerseits um Nichtinformatiker aus dem akademischen Umfeld; auf der anderen Seite stehen Praktiker aus der betrieblichen Praxis, bei denen man nicht notwendig einen akademischen Hintergrund voraussetzen kann.

Schließlich gibt es noch Benutzergruppen, die, z.B. als Entscheidungsträger in der Wirtschaft, sich einen groben Überblick über die Potenzen und Risiken des Data-Mining verschaffen wollen.

DaMiT unterstützt alle diese unterschiedlichen Bedürfnisse, indem es unterschiedliche Zugänge in für die jeweilige Benutzergruppe geeigneten didaktischen Aufbereitungen anbietet. In gewissen Grenzen paßt es sich dem Lernstil des Benutzers an, 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.

Wie bereits in der Einleitung besprochen sind praktische Fähigkeiten und Erfahrungen für erfolg-

reiches Data-Mining essentiell. DaMiT bietet die Möglichkeit, den Umgang mit Algorithmen und kompletten Werkzeugen zu trainieren. Anhand komplexer, praxisnaher Wettbewerbsaufgaben wird den Lernenden das Zusammenspiel der verschiedenen Phasen des Data-Mining-Prozesses nahegebracht.

Neben den fachlichen Fragen spielen jedoch auch rechtliche Aspekte beim E-Learning eine große Rolle. Beispielsweise können nicht alle Inhalte außerhalb des akademischen Umfelds genutzt werden, da Copyrightfragen berührt werden.

All diese Aspekte des DaMiT-Systems diskutieren wir in den folgenden Abschnitten nun detaillierter.

### 2.1 Zugang zum Inhalt

DaMiT bietet verschiedene Zugangswege zu den enthaltenen 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, beispielsweise wenn sie 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.

Dieser 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 Lehrinhalten navigieren kann.

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 in die einzelnen Lektionen, beispielsweise auf die theoretischen Hintergründe oder die angewendeten Methoden, quasi ganz nebenbei verwiesen wird.

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 zutreffenden Fälle erkennen und zu den Lehrinhalten 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 das 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 zu 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 Erstbenutzern 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 lassen. 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 was den Inhaltzugang als auch den Umgang mit Systemfunktionalitäten betrifft. Beispielsweise dürfen nicht persönlich bekannte Benutzer in den Diskussionsforen nicht schreiben, um Mißbrauch vorzubeugen. Außerdem dürfen beispielsweise 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 Bezahlungssystem 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 also in dem Sinne adaptierbar, daß derjenige, der es in seiner Lehre einsetzt, 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. Dies diskutieren wir im nächsten Abschnitt.

## 2.2 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 Internet-basierter 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 reichsspezifische 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 übliche 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 zugrundeliegende didaktische Konzeptionen, die dadurch selbst zum Diskussion- und Evaluationsgegenstand werden.

Storyboards werden in Form von hierarchischen Graphen repräsentiert. Die Knoten repräsentieren einzelne Szenen, die wiederum aus Teilszenen bestehen können. Die Kanten beschreiben antizipierte Übergänge zwischen einzelnen Szenen.

Die Idee der Storyboards trägt noch weiter. Eine Analyse unterschiedlicher Storyboards zeigt, daß – zur Realisierung derselben didaktischer 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. So kann zusätzliche Unterstützung bei der Herstellung didaktisch adäquat aufbereiteten Lehrmaterials geleistet werden.

## 2.3 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.

Bei der Entwicklung eines Konzepts zur Ermöglichung von Adaptivität sind vier Dimensionen zu beachten (vgl. [Spe99]): die Adaptationsmittel<sup>1</sup>, die Adaptationsformen<sup>2</sup>, der Adaptationsprozeß<sup>3</sup> und letztlich das Adaptationsziel<sup>4</sup>.

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. Dabei kann der Benutzer zwischen einer beispielorientierten oder theorieorientierten und einer formalen oder informalen Darstellung von Lerninhalten 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. Beispielerorientierte und theorieorientierte Darstellungen unterscheiden sich dadurch, daß zur Vermittlung derselben Lehrinhalte verschiedene Lernelemente verwendet werden. Beispielsweise werden in einer beispielorientierten Darstellung eines Lernmoduls verstärkt Lernelemente mit einem illustrierenden bzw. erläuternden Charakter, Lernelemente

<sup>1</sup> Wo soll die Adaptivität zum Tragen kommen?

<sup>2</sup> Aufgrund welcher Daten erfolgt eine Anpassung?

<sup>3</sup> Wie erfolgt die Anpassung, benutzer- oder systemgesteuert?

<sup>4</sup> Warum erfolgt eine Anpassung?

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 antizipierten Wissensstand des Benutzers werden daraus Vorschläge für eine – an den Bedürfnissen und Erfahrungen des Benutzers orientierte – Auswahl und Anordnung der relevanten Lerninhalte abgeleitet.

## 2.4 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 ein Spielbeispiel 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, sooft und wie er will. Er kann sich beliebig komplexe Beispiele für komplizierte Algorithmen vorrechnen lassen, und dies in der Gewißheit, keinem Flüchtigkeitsfehler aufzusitzen. Die gleichzeitige Anregung verschiedener Perzeptionskanäle erlaubt es, viel Information sehr kompakt zu vermitteln. Der natürliche Spieltrieb sorgt schließlich 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 herausfinden, die zu bestimmten Effekten bei der Abarbeitung der Algorithmen führen.

Des weiteren bietet DaMiT eine Reihe professioneller Data-Mining-Tools und Daten aus realen Anwendungen an, die in die Wissensvermittlung integriert sind. Die Benutzer lernen neben den eigentlichen Methoden auch gleich noch kennen, welche Varianten in realen Werkzeugen zur Verfügung stehen und können diese bei der Verarbeitung realer Daten erproben.

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 Urhebererschaft an den Lösungen nachweisen können.

## 2.5 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 zugeschnittener Aufgabentyp, die sogenannten „Competitive Exercises“. Bei der Bearbeitung einer solchen Aufgabe hat der Lernende folgende Schritte zu absolvieren:

1. Vertrautmachen mit der Aufgabenstellung, Herunterladen der Trainingsdaten und Auswahl eines Data-Mining-Werkzeuges,
2. die Trainingsdaten offline mit dem Data-Mining-Werkzeug zu bearbeiten und ein Modell zu generieren (dies ist der eigentliche Data-Mining-Schritt, der normalerweise aus mehreren Iterationen besteht),
3. das generierte Modell als PMML-Datei<sup>5</sup> zu exportieren und

<sup>5</sup>PMML ist ein XML-basierter Standard zur Beschreibung von Modellen, die beim Data-Mining entstehen. Dieser Standard wird von den meisten Data-Mining-Tools unterstützt.

4. das PMML-Modell einzureichen (ins DaMiT-System hochzuladen).

Beim Einreichen hat der Benutzer die Möglichkeit, sein Modell digital zu signieren und verschlüsselt zu übertragen. Das ist insbesondere von Bedeutung, wenn eine prüfungsrelevante Leistung erbracht werden soll und etwa ein Urhebernachweis erforderlich ist.

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 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 die 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 nicht so sehr die Lösung, sondern vor allem die Diskussion eines Lösungsweges und von Alternativen im Vordergrund.

### 3 Zusammenfassung

In dem vom BMBF im Rahmen des Zukunftsinvestitionsprogramms geförderte Projekt DaMiT wird Wissen über das Gebiet des Data-Mining auf bisher nicht dagewesenen Weise angeboten. Es werden die Grundlagen und Prinzipien des Data-Mining sowie Verfahren des Data-Mining inkl. ausgearbeiteter Fallstudien und einer Bewertung von kommerziellen Systemen zur Verfügung gestellt. 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 Stammvorlesungen über das Gebiet der Künstlichen Intelligenz. Außerdem wird gerade der Einsatz in der Weiterbildung in Unternehmen vorbereitet.

DaMiT ist ein integriertes System, welches – getrieben vom Paradigma des „Learning by Doing“ – das Studium des Data Mining organisch mit seinen Anwendungen verbindet. Hierzu bietet DaMiT personalisierten Inhalt über verschiedene Zugangsmodi und über Adaptation des Contents in jeweils adäquaten didaktischen Aufarbeitungen an. Für die komplexen Anwendungsübungen der Lernenden hat DaMiT ein eigenes Konzept hervorgebracht, entwickelt und implementiert: 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 einem integrierten e-Payment Funktionalität und angebunden an modernste IT-Sicherheitsinfrastrukturen der Partner, seien es nun Universitäten und Hochschulen oder Unternehmen, die DaMiT einsetzen.

Auf die jeweils aktuelle Version des Tutorsystems kann im Internet unter [damit.dfki.de](http://damit.dfki.de) zugegriffen werden.

### Literatur

- [Spe99] Marcus Specht. *Adaptive Methoden in computerbasierten Lehr/Lernsystemen*. PhD thesis, Universität Trier, 1999.
- [TD01] Bernhard Thalheim and Antje Düsterhöft. Conceptual Modelling of Internet Sites. In H.S. Kunii, S. Jajodia, and A. Solberg, editors, *Proc. ER'2001*, volume 2224 of *Lecture Notes in Computer Science*. Springer-Verlag, 2001.
- [WF99] Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers, 1999.