

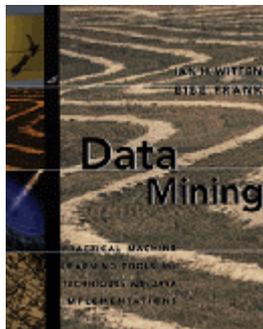
# Neuerscheinungen – Bücher, die mir auffielen

## Folge 2

Gottfried Vossen, Universität Münster

In dieser Rubrik möchte ich Ihnen wieder Bücher vorstellen, die mir in letzter Zeit auf den Tisch gekommen sind und die mir lesenswert (oder zumindest erwähnenswert) erscheinen (Hinweise in eigener Sache – siehe unten – seien auch erlaubt). Die dabei jeweils präsentierte Auswahl erhebt weder Anspruch auf Vollständigkeit noch auf Einhaltung einer bestimmten Systematik; der Natur meiner persönlichen Interessen entsprechend entstammen allerdings die meisten der genannten Bücher den Bereichen Datenbanken und Informationssysteme.

Ich beginne heute mit einer nicht nur für Informatiker interessanten Neuerscheinung zum Thema Data Mining. Unter dem Begriff *Data Mining* versteht man eine Sammlung von Techniken, mit denen sich aus großen Datenbeständen Zusammenhänge, letztlich Wissen ableiten lässt, das zuvor in dieser Form nicht bekannt ist. Das Gebiet hat in den letzten Jahren sowohl im Datenbankbereich wie auch in der Statistik und in der Künstlichen Intelligenz einen enormen Aufschwung erlebt; man erhofft sich z.B. in E-Commerce-Anwendungen vom Einsatz von Data Mining-Techniken ein Erkennen von Kunden und deren Verhalten, um entsprechend darauf reagieren zu können. Das Buch von Ian H. Witten und Eibe Frank mit dem Titel *Data Mining – Practical Machine Learning Tools and Techniques with Java Implementations* (Morgan Kaufmann 2000, ISBN 1-55860-552-5) ist ein kompakter und aktueller Abriß der wichtigsten Techniken, die im Data Mining zum Einsatz kommen.

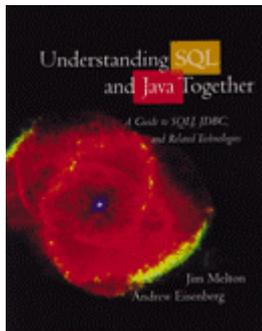


Es motiviert zunächst Data Mining und maschinelles Lernen anhand von Beispielen wie der Wettervorhersage für ein fiktives Spiel oder dem Verschreiben von Kontaktlinsen. Es behandelt die wesentlichen Techniken wie Entscheidungstabellen und -bäume, Assoziationsregeln, Klassifikation, Cluster-Bildung, statistische Modellierung, lineare Modelle und bettet diese ein in den Prozess der Wissensgewinnung durch Data Mining. Dieser auch als *Knowledge Discovery in Databases* (KDD) bekannte Prozess erfordert nämlich mehr als nur algorithmische Techniken.

Es ist zum Beispiel wichtig, die zu untersuchenden Daten geeignet vor zu verarbeiten, und auch nach erfolgtem Mining ist eine geeignete Interpretation der Ergebnisse wesentlich. Witten und Frank gehen auch auf Implementierungstechniken ein und bieten sogar fertige Algorithmen im Web an, die im Buch beschrieben sind. Der Leser kann dadurch selbst mit den beschriebenen Verfahren experimentieren und Erfahrungen sammeln. Das Buch sei jedem empfohlen, der an einer übersichtlichen Darstellung des Gebietes interessiert und dem an Hinweisen auf relevanten Quellen aus dem Umfeld des Data Mining gelegen ist.

Eine Neuigkeit für den Datenbankprogrammierer ist das neue (und seit längerem erwartete) Buch von Jim Melton und Andrew Eisenberg mit dem Titel *Understanding SQL and Java Together – A Guide to SQLJ, JDBC, and Related Technologies* (Morgan Kaufmann 2000, ISBN1-55860-562-2). Dies ist die erste ausführliche Darstellung von statischem SQL eingebettet in Java-Programme (SQLJ) sowie von verwandten Techniken (JDBC und ähnlichem). Das Zusammenwachsen bzw. die Integration von SQL und Java ermöglicht eine

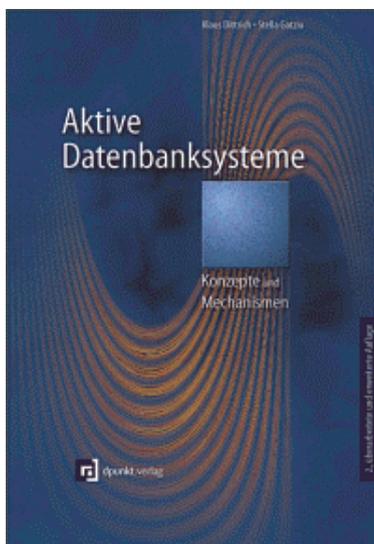
Manipulation der Tabellen einer relationalen (SQL-) Datenbank von Java-Programmen aus, ferner eine Form von Persistenz in Java, ein Schreiben von *Stored Procedures* in SQLJ und eine Ablage von Java-Objekten in relationalen Tabellen.



Das Buch beginnt mit einer Java-Einführung für den SQL-Programmierer, gefolgt von einer SQL-Einführung für den Java-Programmierer. Ferner werden JDBC 1.0 und 2.0 sowie die Teile 0, 1 und 2 des mittlerweile spezifizierten SQLJ ausführlich behandelt. In Teil 0 geht es um die Einbettung von SQL-Anweisungen in Java-Programme. In Teil 1 steht die Verwendung von in einer SQL-Datenbank gespeicherten Java-Methoden innerhalb von SQL-Code, in Teil 2 die Definition und Speicherung von Daten, die Instanzen von Java-Klassen sind, im Vordergrund.

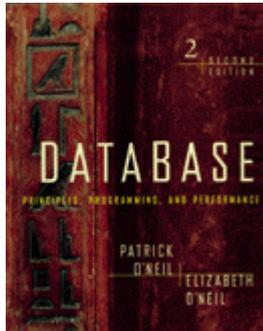
SQLJ Teil 0 wurde 1998, Teil 1 1999 vom ANSI standardisiert. Das Buch richtet sich an den im Umgang mit Datenbankanbindungen (wie OLE oder ODBC) erfahrenen Programmierer, macht allerdings auch keinen Hehl daraus, dass die hier beschriebene Technologie noch keineswegs völlig stabil ist. Die Programmbeispiele sowie eine Reihe relevanter Tools (wie *Oracle JDeveloper*, *Sybase SQL Anywhere* oder *Informix Cloudscape*) werden auf CD-ROM mitgeliefert, so dass der Leser einerseits die Beispiele leichter nachvollziehen und andererseits von der derzeitigen Vielfalt von Produkten und Werkzeugen einen guten Eindruck bekommen kann.

Auch der deutschsprachige Buchmarkt hat für den Datenbankinteressierten Neues zu bieten: Von Klaus R. Dittrich und Stella Gatzju ist gerade ***Aktive Datenbanksysteme – Konzepte und Mechanismen*** in der 2. Auflage erschienen (dpunkt.verlag 2000, ISBN 3-932588-19-3). Dieses Buch war ja ursprünglich Bestandteil der bei Thomson erschienenen TAT-Reihe, die allerdings nicht erfolgreich war. Dass das gleiche für einzelne Titel aus dieser Reihe nicht zutrifft, beweist dieser Band. Er ist nach wie vor die einzige deutschsprachige Einführung in das Gebiet der aktiven Datenbanksysteme und behandelt die wichtigsten Aspekte dieser. Grundsätzlich geht es bei aktiven Datenbanksystemen darum, sie von passiven Systemen, die lediglich auf Anstoß von außen (also z.B. eine Benutzeranfrage) reagieren, zu aktiven zu machen, die das Eintreten bestimmter Ereignisse beobachten, gegebenenfalls Bedingungen überprüfen und schließlich Aktionen selbständig auslösen können.



Grundkonzepte aktiver Datenbanken sind insbesondere die *Event-Condition-Action*-Regeln (ECA-Regeln), mit denen sich Situationen, Ereignisse, Bedingungen und Aktionen beschreiben lassen. Für diese werden Möglichkeiten der Spezifikation und der Bearbeitung dargestellt. Daran schließen sich Ausführungen zur Systemarchitektur, über Implementierungsaspekte, Anwendungen und Systeme an, wobei auch auf die entsprechenden Entwicklungen im Rahmen von SQL:99 eingegangen wird. Das Buch ist kompakt geschrieben, Sachverhalte werden prägnant an einem durchgehenden Beispiel verdeutlicht, und am Ende der Lektüre weiß der Leser das Gebiet und die dort ablaufenden Aktivitäten einzuordnen.

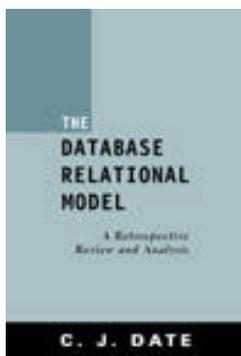
Ebenfalls in zweiter Auflage ist inzwischen das Buch von Patrick O’Neil und Elizabeth O’Neil, *Databases – Principles, Programming, and Performance* (Morgan Kaufmann 2001, ISBN 1-55860-438-3) erschienen. Dieses Buch ist inzwischen ein Standardwerk an vielen amerikanischen Hochschulen und behandelt die Theorie und Praxis von Datenbankentwurf und –implementierung. Dazu gehören einerseits Standardthemen wie Grundlagen des relatio-



nenalen Datenmodells und SQL, Datenbankentwurf, Integrität, Sichten, Sicherheit und Datenbank-Kataloge. Dazu gehören vor allem jedoch implementierungsrelevante Themen wie Indexierung, Anfrageverarbeitung, Transaktionsverarbeitung und parallele sowie verteilte Datenbanken. Gerade in diesen Themen ist das Buch sehr ausführlich, mit vielen praktischen Beispielen ausgestattet und sehr aktuell. Die zweite Auflage widmet sich an neuen Themen vor allem objekt-relationalen Datenbanken, also der Integration von SQL und Objekten, wie man sie heute in gän-

gigen Produkten wie Oracle8 oder Informix vorfindet. Diesen beiden Systemen ist dann auch ein separates Kapitel gewidmet, da sie die Verbindung von Objekten und relationalen Datenbanken auf sehr unterschiedliche, aber instruktive Weise vollziehen.

Wer glaubt, dass Chris Date lange nichts von sich hat hören lassen (und wer vielleicht jetzt den nächsten Band seiner berüchtigten „*Selected Writings*“ erwartet), darf sich auf *The Database Relational Model – A Retrospective Review and Analysis* (Addison Wesley Longman 2001, ISBN 0-201-61294-1) freuen.

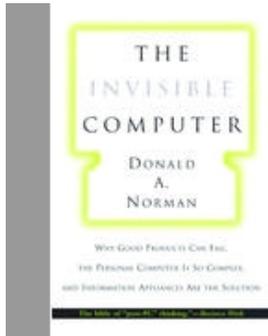


Es handelt sich um ein kleines Bändchen mit gerade mal 152 Seiten (damit aber immer noch dicker als das Buch von Dittrich und Gatzju), welchem Date selbst den Untertitel „*A historical account and assessment of E.F. Codd’s contribution to the field of database technology*“ gibt. Ähnlich den *Writings* besteht das Buch aus einer Serie von 12 Artikeln, die Date bereits zuvor in einem einschlägigen Magazin, hier der *Intelligent Enterprise*, veröffentlicht hat. Er behandelt in den 12 entsprechenden Buchkapiteln der Reihe nach zahlreiche Codd’sche Papiere, wobei er deren wesentliche Beiträge informal erläutert.

Naturgemäß beginnt dieser historische Abriss mit den Arbeiten aus den Jahren 1969 und 1970, in denen Codd die Tabellensicht auf Datenbanken vorgeschlagen und die Logik-nahe Begründung dieser Sicht vorgenommen hat. Es folgen die Relationenalgebra und der Relationenkalkül, Normalformen und Erweiterungen des ursprünglichen Modells, insbesondere das RM/T-Modell. Natürlich ergeben sich für den Experten hier keine neuen Erkenntnisse; dennoch ist diese Zusammenstellung ganz nett zu lesen insbesondere für den, der sich mit Datenbanken erst jetzt zu beschäftigen beginnt (und der das Aufarbeiten vor allem älterer Literatur daher gerne vor sich her schiebt). Das Gebiet hat mittlerweile immerhin eine 30jährige Geschichte, und sowohl Date als auch Codd waren und sind zentrale Figuren dieser Geschichte.

Wer noch etwas Urlaubslektüre sucht, dabei jedoch auf der Informatik und dem Computer nahes Gedankengut nicht völlig verzichten möchte, dem empfehle ich das Buch *The Invisible Computer – Why Good Products can Fail, The Personal Computer is so Complex, and*

***Information Appliances are the Solution*** von Donald A. Norman (The MIT Press 1999, ISBN 0-262-64041-4).



Wie der Autor im Vorwort schreibt, will er einen realistischen Blick auf die Welt der Technologie werfen, die Frage angehen, warum gute Produkte sich als Flop erweisen und schlechte erfolgreich sein können; er will begründen, warum der PC heutzutage so komplex ist und dass nur ein Überdenken der ihm unterliegenden grundlegenden Ideen dies nachhaltig ändern kann. Er beginnt bei Thomas Edison und seinem Phonographen: Hier ging es lediglich um eine technische Neuerung ohne Rücksicht auf Kundenwünsche.

Als Folge daraus sind Edisons Firmen sämtlich gescheitert. Wenn man sich dann überlegt, welche Veränderungen ein Produkt durchläuft, wenn es „erwachsen“ wird, wird klar, dass ein Produkt, dessen Entwicklung im Wesentlichen abgeschlossen ist, den Menschen zum Mittelpunkt der Betrachtung machen muss, um erfolgreich zu sein. Norman untersucht diese These am Beispiel des PCs und folgert, dass der digitale PC nicht zum analogen Menschen passen kann. Abhilfe sieht er in den *Information Appliances*, also intelligenten informationsverarbeitenden Geräten, die den Menschen in hoher Zahl umgeben, über deren Anwesenheit er aber nicht nachzudenken braucht.

Zum Schluss wieder einige Kurzhinweise auf weitere Neuerscheinungen, die ich mir noch nicht so intensiv angesehen habe (oder zu denen ich – aus offensichtlichen Gründen – hier nichts weiter schreiben möchte):

- M. Merz: ***Electronic Commerce – Marktmodelle, Anwendungen und Technologien***; dpunkt.verlag 1999, ISBN 3-932588-31-2

Dieses Buch ist (wie zahlreiche andere Neuerscheinungen in diesen Tagen auch) dem Thema elektronischer Handel gewidmet, hebt sich jedoch von vielen Konkurrenten wohltuend ab. Dem Autor gelingt eine interessante Mischung aus technischen und betriebswirtschaftlichen Themen (u.a. neue ökonomische Modelle, Smart Cards, elektronisches Geld, Modelle wie B2C und B2B) sowie eine Übersicht über am Markt befindliche B2C-Software (die natürlich wegen der Schnellebigkeit des Gebietes nicht aktuell sein kann). Merz hat sowohl einen breiten Background (aus Informatik und BWL) als auch Erfahrungen im Consulting, in Industrieprojekten und in der universitären Lehre. Als Vorlesungsgrundlage, aber auch als Einstieg für den Anwender ist das Buch sehr zu empfehlen.

- R.G.G. Cattell et al.: ***The Object Data Standard ODMG 3.0*** ; Morgan Kaufmann 2000, ISBN 1-55860-647-4

Für diejenigen, die die Standardisierungs-Arbeiten der *Object Database Management Group* (ODMG) weiter verfolgen möchten oder müssen, vermerkt der Chronist ein Erscheinen der dritten Auflage des Standards für Objektmodelle. Neu oder ergänzt bzw. verbessert sind z.B. das Java-Binding oder Mappings zwischen Objekten und Relationen. Das Buch ist nach wie vor die definitive Referenz für die *Object Query Language* (OQL), für die es ja inzwischen eine Reihe von Implementierungen gibt..

- Th. Michel: *XML kompakt – Eine praktische Einführung*; Carl Hanser Verlag 1999, ISBN 3-446-21302-3

Es ist kein Geheimnis mehr, dass XML im Zusammenhang mit dem Internet und z.B. dem Electronic Commerce eine zunehmend wichtigere Rolle spielt (siehe hierzu auch meine Surftipps). Der Buchmarkt reagiert auf eine solche Erkenntnis üblicherweise mit einer Flut von Neuerscheinungen, und so kann man sich derzeit auch vor XML-Titeln nicht mehr retten. Hier die Spreu vom Weizen zu trennen ist nicht einfach, zumal sich weite Teile des XML-Umfelds noch rasant ändern. Das Buch von Michel gehört aus meiner Sicht zu den gelungenen Büchern über XML, was sicher auch daran liegt, dass es sich auf das XML-Konzept, DTDs und XML-Dokumente beschränkt (also z.B. in die aktuelle Diskussion über eine XML-Anfragesprache gar nicht erst einsteigt). Es behandelt daneben Namenräume, Schemata, DOM und XLink nur am Rande und bringt eine Menge relevantes Material auf CD-ROM mit.

- G. Vossen, K.-U. Witt: *Grundlagen der Theoretischen Informatik mit Anwendungen*; Vieweg-Verlag 2000, ISBN 3-528-03147-6

In diesem Buch geht es den Autoren um eine Einführung in die Theoretische Informatik, die deren Anwendungen möglichst unmittelbar erkennen lässt. Exemplarisch sei lediglich der Zusammenhang zwischen dem Sprachstandard XML und kontextfreien Grammatiken mit regulären Ableitungen erwähnt, der bei der Festlegung einer DTD (Document Type Definition) wesentlich ist. Das Buch ist aus Vorlesungen entstanden, welche die Autoren in anwendungsorientierten Studiengängen an Unis und FHs gehalten haben. Um das Erkennen des Buches zu erleichtern, darf ich abschließend sein Cover zeigen:

