



Johann Wolfgang Goethe-Universität Frankfurt am Main

Anwendung von Data Mining

[University Home](#)

[University
Overview](#)

[Suche](#)

[Dept. of
Computer Science](#)

[Databases and
Information
Systems](#)

Praktikum - Winter 2002/2003

Dr. Christoph Schommer, Dipl. Math. Norbert Wartig

Erste Vorbesprechung: Freitag, den 18.10.2002, 15:00 Uhr c.t.,
Zeppelinallee 6-8, Zi. 107

A. Meetings

Die erste Phase beginnt am 18. Oktober 2002 mit dem Kickoff Meeting. Weitere Termine sind:

- Meeting am 1. November 2002: [Fragen](#)
- Meeting am 15. November 2002: [Klausurfragen Summer 2002](#)
- Meeting am 29. November 2002: [Übungen am Rechner](#)
- Meeting am 13. Dezember 2002 (Antestat)

Die [Literaturliste](#) enthält Bücher, die mittlerweile auch in der Informatik-Bibliothek verfügbar sind. Die zweite Phase beginnt am 16. Dezember 2002. Die Termine sind:

- 8. Januar 2003 (Besprechung der Analyseaufgabe und Übergabe der Aufgaben)
- 24. Januar 2003 (Data Preprocessing für die Produktion des Data Mart)
- 13./14. Februar 2003 (Data Mining; Fragen zur Dokumentation)
- 28. Februar 2003 (Data Mining; Vorbereitung für Abschlusspräsentation)
- 14. März 2003 (Abschlusspräsentation; **SR 11, Robert-Mayer Str. 11-15**)

Bitte füllen Sie den [Feedback](#)-Bogen aus!

Die [Abtestat](#)-veranstaltung beginnt am **14.3. im SR11 um 9ct.** (bitte pünktlich sein!). Nach einer kurzen Einführung (Welcome Session) beginnen die einzelnen Vorträge zu je einer Stunde:

- **09:30**-10:30 Gruppe 1
- 10:35-11:35 Gruppe 2
- 11:40-12:40 Gruppe 3
- 12:40-14:15 Pause/Lunch-time
- **14:15**-15:15 Gruppe 4
- 15:20-16:20 Gruppe 5
- 16:25-17:25 Gruppe 6

- 17:30-17:40 Wrap-Up (Zusammenfassung)

Welche Praktikumsgruppe welcher Gruppe entspricht, wird durch **Auslosen ohne Zurücklegen** bestimmt. Es wäre wünschenswert und darüberhinaus sehr fair, wenn alle Teilnehmerinnen und Teilnehmer an jedem Vortrag partizipieren. Falls aus wichtigem Grunde eine Teilnahme nicht oder nur teilweise möglich ist, möge Er oder Sie sich bitte bei mir oder Norbert Wartig melden.

Bitte beachten Sie: die jeweilige Gruppe agiert als Expertenteam für Ihre Domäne. Ich erwarte von dem Auditorium rege Teilnahme, in keinem Falle aber soll die Veranstaltung zu einer Reihe von Monologen bzw. Nickfolgen degenerieren.

Denken Sie auch daran, mir die **Karten für die RBI-Räumlichkeiten** zurück zu geben. Das **Abgabedatum für die Dokumentation ist Dienstag, der 25. März (Stichtag)**. Die Scheinvergabe erfolgt ab April 2003, falls die Dokumentation rechtzeitig eingetroffen, in Ordnung und der Vortrag gelungen ist.

Im Anschluss an die Präsentationen (14. März) findet eine **Praktikumsabschlussfeier** statt! Für Getränke wird gesorgt, es wäre schön, wenn eine jede Gruppe hierzu diverse Salate, Brote, etc. mitbringen könnte. Möglicherweise können wir auch ein internationales Gericht zusammen stellen.

B. Inhalt

Ziel des Praktikums ist das Schreiben eines Skripts. Dieses soll neben den theoretischen Grundlagen des Knowledge Discovery 3-4 anwendungsorientierte Dokumentationen von realen Projekten enthalten. Beide Teile sollen letztendlich zusammengefasst und in ein Buchformat gebracht werden. Die Publikation in einem Buchverlag kann verfolgt werden.

Das Praktikum wird als **Projekt** angesehen, es soll eine Brücke darstellen zwischen theoretischem Fundament und industrieller Anwendung (Universität / Industrie). Zu diesem Zweck werden zwei Projektabschnitte definiert: im ersten Teil sollen die Grundlagen für das Projekt erlernt werden, im zweiten Teil im Rahmen von Projektgruppen (Redaktionsteam, Analysegruppen) eine Aufgabe selbständig bearbeitet werden (Problematik, Generelles Vorgehen, Strategien, Ergebnisse, Handlungsempfehlungen).

- IBM DB2 Intelligent Miner User Manual ([Eng](#), [Deu](#))
- IBM DB2 Association Visualizer ([Eng](#), [Deu](#))
- IBM DB2 Intelligent Miner for Data - Stepwise Polynom Regression ([Eng](#))
- IBM DB2 Universal Database for Windows: [User Manual](#)
- IBM DB2 Universal Database for Windows: [Get Started](#)
- IBM DB2 Universal Database for Windows: [Glossary](#)
- IBM DB2 Intelligent Miner for Data: [Application in banking](#)
- IBM DB2 Intelligent Miner for Data: [Application in health](#)
- IBM DB2 Intelligent Miner for Data: [Application in Retail](#)
- IBM Intelligent Miner for Data: [Other Applications](#)
- Business Intelligence: [Tutorial](#)

C. Gruppenzusammensetzung

Am Praktikum nehmen StudentInnen unterschiedlicher Couleur teil. So sind neben InformatikerInnen mit Nebenfächern wie Mathematik, Wirtschaft und Psychologie auch BioInformatikerInnen dabei. Folgende Gruppen haben sich gebildet (Die Sprecher sind fett markiert):

- Gruppe 1: Bioinformatik (Fr. Wendel, Fr. Schmidt, **Hr. Radestock**, Hr. Junglas): Untersuchung von Daten aus der Bioinformatik (**Hefe**)
- Gruppe 2: Medizin (**Fr. Burst**, Fr. Christian, Fr. Sun): Suche nach Indikatoren für das Auftreten von **Hepatitis B und C**
- Gruppe 3: Handel (**Fr. Uhde**, Hr. Marekwica, Hr. Mounir): Warenkorbanalyse von **Handelsdaten aus der Tabakbranche**
- Gruppe 4: Sport (**Hr. Güner**, Hr. Durak, Hr. Yildirim): Kick-Stream-Analyse im Fussball anhand von **Fussball Bundesligadaten der Saison 1998/1999 (Bayern München)**
- Gruppe 5: VWL (**Hr. Abu-Gaber**, Hr. Dzaferovic, Hr. Abu Shahrour): Risiko-scoring **volkswirtschaftlicher Daten**
- Gruppe 6: Web (**Hr. Nesil**, Hr. Hassani, Hr. Alsan): Weblog Analyse von Bestelldaten der **gazelle.com**
- Gruppe 7: Redaktion (**Hr. Osman**, Hr. Arabzadeh)

Momentan sind dem Praktikum 21 StudentInnen assoziiert. Diese haben folgende Aufgaben zu erfüllen:

- Einarbeitung in die jeweilige Domäne
- Einarbeitung in DB2 Universal Database und DB2 Intelligent Miner
- Analyses einer Data Mining-relevanten Fragestellung (inkl. Schreiben eines Projektberichts und Präsentation der Ergebnisse) oder Mitwirken am Schreiben des theoretischen Teils des vorlesungsbegleitenden Knowledge Discovery-Skriptes

Den Schein erhält die-/derjenige, der die festgelegten Meilensteine (Antestat, Abtestat) erreicht. Ein **Antestat** wird Mitte Dezember (13.12.), ein **Abtestat** Mitte März durchgeführt. Die Arbeit wird mit der Aushändigung der fertigen Dokumentation sowie eines Vortrags im Umfang von etwa einer Stunde abgeschlossen.

Die Voraussetzungen für das Praktikum sind

- Kenntnisse in SQL
- Kenntnis von TeX bzw. LaTeX (für das Redaktionsteam)
- Regelmässige Teilnahme an der Vorlesung 'Data Mining' sowie Schein (D)
- Freude am Arbeiten im Gebiet Knowledge Discovery

Der Zeitplan, organisatorische Dinge, Gruppeneinteilungen, Anforderungen für das Antestat, etc. werden am 18.10.2002 (Kickoff-Meeting) bekannt gegeben.

D. Werkzeuge



Datenbank: DB2 Universal Database



Data Mining: Intelligent Miner for Data

E. Referenzen

- Vorlesung 'Data Mining' vom [Sommer 2002](#)
- [Literaturliste](#)

F. Sonstiges

- Sprechstunde: Freitags, n.V. (Zi. 107, Zeppelinallee 6-8)

Last Update: April 20, 2002