



Data/Information Quality Management

Seminar WI/Informationsmanagement
im Sommersemester 2002

Markus Berberov, Roman Eder, Peter Gerstbach
11.6.2002

Inhalt

- ❑ Daten und Datenqualität
 - ❑ Einführung und Definition
 - ❑ Methoden die Datenqualität zu verbessern
- ❑ Data Quality Management
- ❑ Knowledge Discovery in Databases

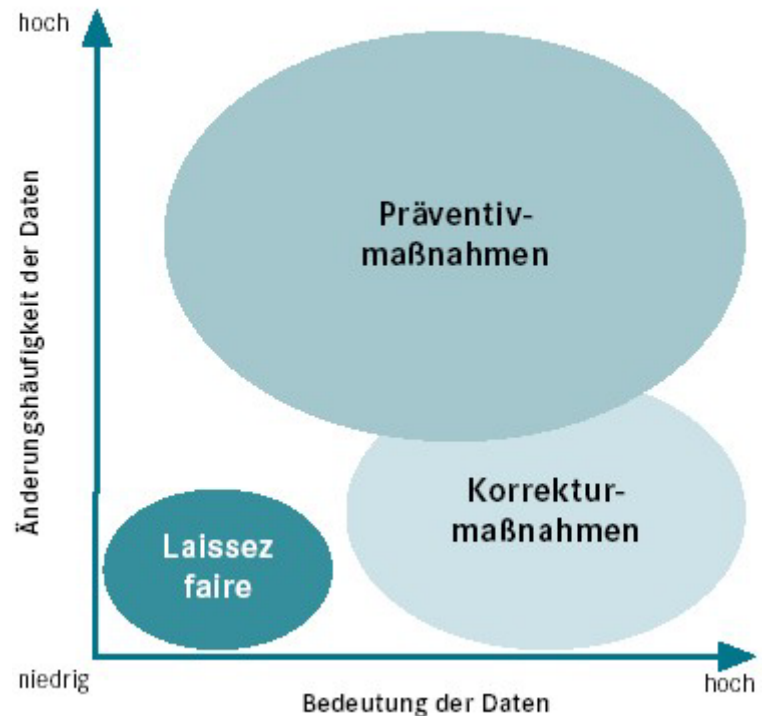
Daten/Datenqualität: Einführung und Definition

- ❑ Qualität der Daten
 - ❑ Thomas C. Redman: Data A is better quality than Data B if it meets customer's demands better
- ❑ Anforderung
 - ❑ Relevanz
 - ❑ Korrektheit, Konsistenz
 - ❑ Aktualität
 - ❑ Verfügbarkeit, Zugänglichkeit
 - ❑ Sicherheit
- ❑ Motivation
 - ❑ höherer Profit, Kundenzufriedenheit und Arbeitsklima

Methoden die Datenqualität zu verbessern

- ❑ Korrekturmaßnahmen
 - ❑ Vergleich mit Umwelt
 - ❑ Database Bashing (Vergl. mehrerer Datenquellen)
 - ❑ Data Edits (Computerroutinen)
- ❑ Präventivmaßnahmen
 - ❑ Prozesskontrolle und Verbesserung
 - ❑ Prozessdesign
- ❑ Laissez faire

Maßnahmenportfolio nach verschiedenen Kriterien



Inhalt

- ❑ Daten und Datenqualität
- ❑ Data Quality Management
 - ❑ Einführung und Definition
 - ❑ Total Data Quality Management
 - ❑ Kosten und Return of Investment
- ❑ Knowledge Discovery in Databases

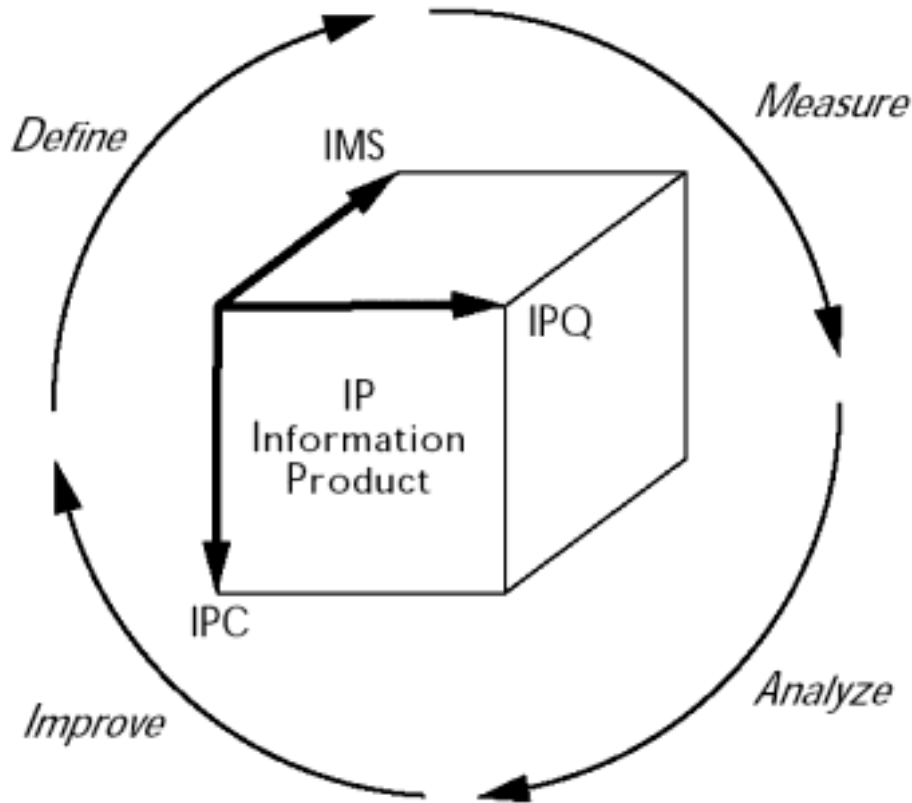
Data Quality Management: Einführung und Definition

- ❑ Instrument der Unternehmensleitung um Kosten-Nutzen-Faktor der Produktion zu optimieren

- ❑ Qualität als Marktfunktion
 - ❑ Produkt/Dienstleistung angeboten
 - ❑ Zusammenspiel: Kunde - Markt - Staat

- ❑ internen und externen Nutzen
 - ❑ Transparenz innerhalb der Organisation, Optimierung, Qualitätsbewusstsein fördern
 - ❑ Kundenzufriedenheit, Konkurrenzfähigkeit

Total Data Quality Management (TDQM)



Legend

IP: Information Product

IPC: IP Characteristics

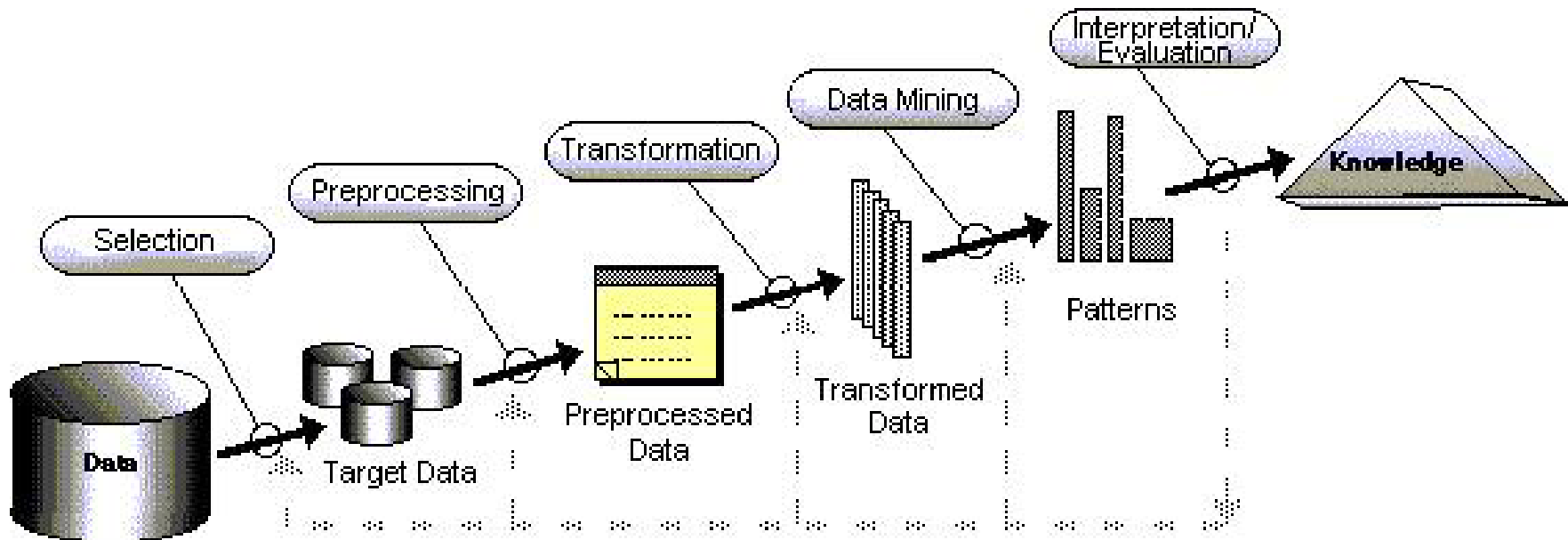
IPQ: IP Quality

IMS: Information Manufacturing System

Inhalt

- ❑ Daten und Datenqualität
- ❑ Data Quality Management
- ❑ Knowledge Discovery in Databases (KDD)
 - ❑ KDD-Prozess
 - ❑ Begriffe
 - ❑ Data Mining Technologien
 - ❑ Data Mining Werkzeuge
 - ❑ Text und Web Mining
 - ❑ Vergleich der Mining-Arten

KDD-Prozess



KDD Begriffe

- ❑ Knowledge Discovery
 - ❑ Mehrstufiger Findungsprozess (von Rohdatenauswahl bis zur Ergebnissinterpretation).
- ❑ Data Mining
 - ❑ Data Mining ist ein Teilschritt des KDD-Prozesses, der aus Algorithmen besteht, die aus einer vorgegebenen Datenbasis eine Menge von Datenmustern liefern.
- ❑ Data Warehouse
 - ❑ Riesige Datenbank, die alle Informationen und Daten eines Unternehmens in eine sinnvolle Struktur bringt.
- ❑ Data Marts
 - ❑ Stellt ein Subsystem eines Data Warehouse dar.

Data Mining Technologien

- ❑ Klassifikation
- ❑ Clustering
- ❑ Assoziation
- ❑ Sequenzanalyse
- ❑ Generalisierung
- ❑ Neuronale Netze

Data Mining Werkzeuge

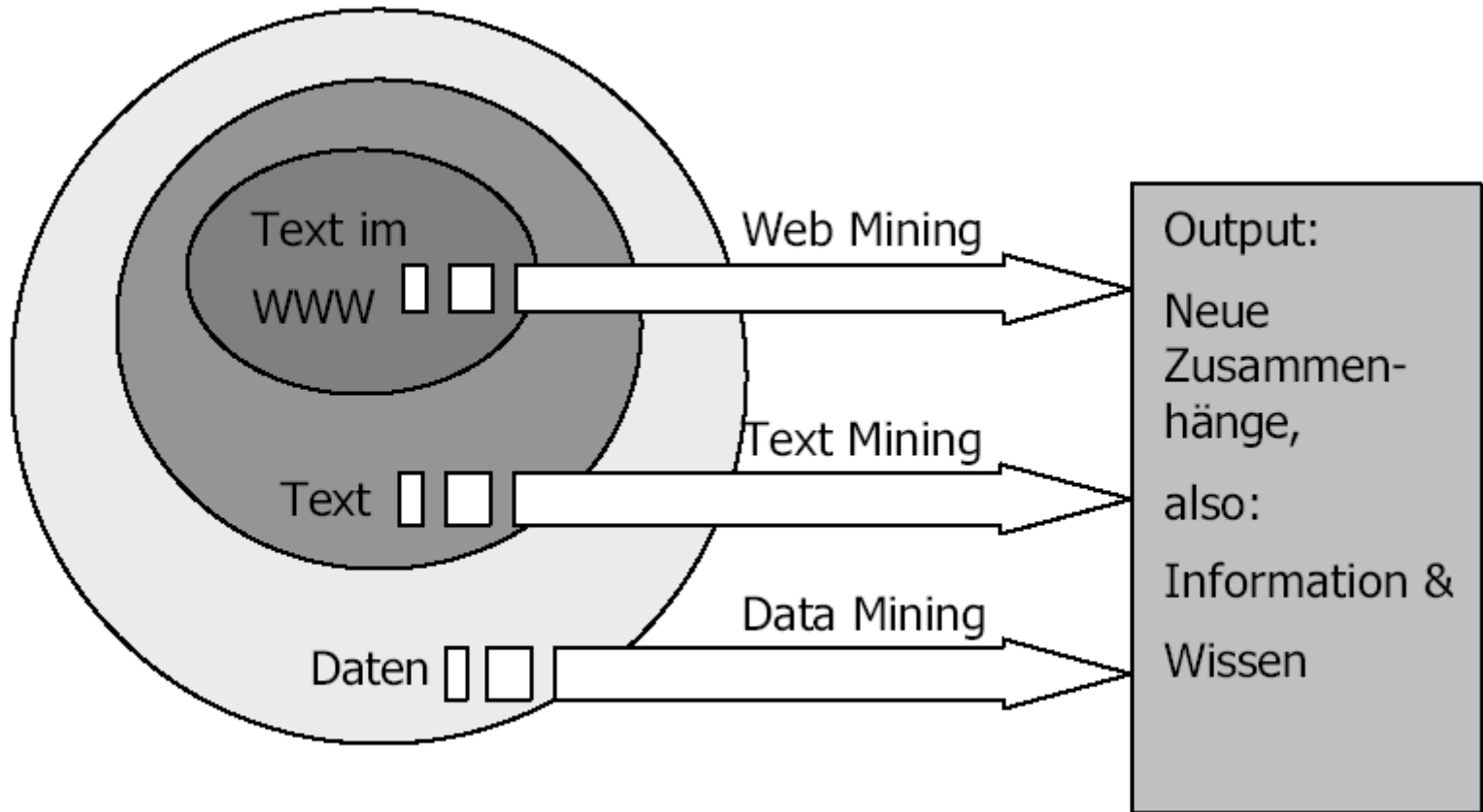
- ❑ Softwarepakete führen Data Mining-Analysen durch
- ❑ Beispiel
 - ❑ IBM Intelligent Miner
 - ❑ SPSS Clementine
 - ❑ SAS Enterprise Miner
- ❑ Kriterien
 - ❑ Systemanforderungen und Architekturen
 - ❑ Unterstützte Datenquellen
 - ❑ Unterstützte Data Mining Technologien
 - ❑ Anwendung

Text und Web Mining

- ❑ Text Mining
 - ❑ linguistische Analyse
 - ❑ Gemeinsamkeiten in Dokumenten finden

- ❑ Web Mining
 - ❑ Muster und Strukturerkennung im Internet
 - ❑ Web Content Mining
 - ❑ Web Structure Mining
 - ❑ Web Usage Mining

Vergleich der Mining-Arten



Zusammenfassung

- ❑ unvollständige und falsche Daten verursachen Kosten
- ❑ DQM versucht diese Probleme zu lösen
- ❑ mögliche Ansätze:
 - ❑ Datenkorrektur und Prävention
 - ❑ organisatorische Maßnahmen
 - ❑ Knowledge Discovery/Data Mining: viele unübersichtliche Daten auf wenige, interpretierbare Muster reduzieren
- ❑ Wettbewerbsvorteil erlangen
- ❑ DQM wird in Zukunft stark an Bedeutung gewinnen

Fragen

