Module 6:
Knowledge Management

Inferences from data,
Data mining
Knowledge portals

KBL Srivastava

1-Topics on learning from data

- Understand the “Learning” Concept
- How to go for data visualization
- Data Neural Networks as learning models
  - The Basics
  - Supervised and Unsupervised Learning
  - Business Applications
  - Relative fit with KM
- Association Rules- Market bases analysis, PETCO
- Classification Trees
- Implications for Knowledge Management

Importance of learning

- Learning is an iterative process where the final model results from the combination of prior knowledge and newly discovered information
- Learning tools are critical for development of a knowledge management environment.
- Data driven tools create the model based on patterns inferred from the data.
- Thus it is important to understand the concept of learning from data.
- The goal is to improve the quality of the communication and decision making in the firm.

The concept of learning and Goals of learning

- Learning is a process of filtering ideas and transforming them into valid knowledge having the force to guide decisions.
  - The unifying concept of learning is the specific mechanism that helps companies determine the kind of knowledge required for decision making.
- Goals of the learning process- knowledge that can be used in business decision making

1. Discovering new patterns in the data
2. Verifying hypothesis formed from previously accumulated al-world knowledge
3. Predicting future values, trends, and behavior

The Concept of learning and Knowledge validation

Knowledge validation is a two step process- Model validation and consensual approval.

- Model validation involves testing the logical structure of a conceptual or operational model for internal consistency and assessing the results for external consistency with the observable facts of the real world.
- Consensual approval means approval of a special reference group or the user of the results

Two approaches to learning models-

Top down- one starts with the hypothesis derived from observation, intuition or prior knowledge- generate ideas, develop models and evaluate them for validation

Bottom up- no hypothesis testing, learning techniques are used to discover new patterns by findings key relationship in the data

Data Visualization

- Exploring the data means looking visually for groups or trends that are meaningful and useful for the decision maker

It includes:

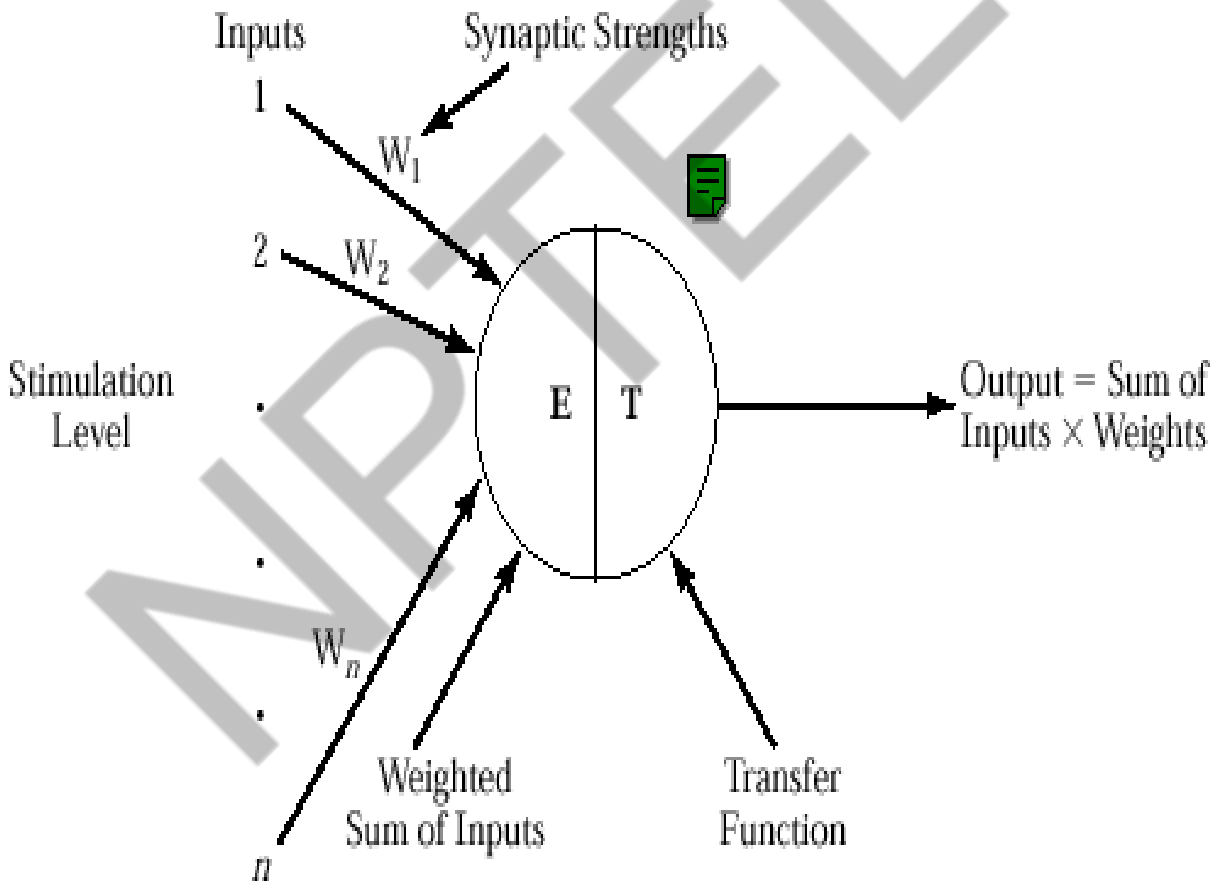
- Distribution of key attributes (e.g., target attribute of a prediction task)
- Identification of outlier points that are significantly outside *expected range* of the results
- Identification of initial hypothesis and predictive measures
- Extraction of interesting grouping data subsets for further investigation

Neural networks as learning models

- Modeled after the human brain network
- The technology attempt to simulate biological information processing via massive networks of processing elements called neurons.
- Neural nets and computers are different.
- Neural nets are neither digital nor serial, they are analog or parallel. They learn by examples not by programmed rules or instructions. Digital computers do not evolve.
- Neurons evaluates **inputs**, performs a weighted sum, and compares result to a ***threshold*** (transfer function) level. If sum is greater than threshold, the neuron fires.
- Interconnecting or combining neurons with other neurons form a layer of nodes or a neural network.

A Neuron Model

(source: Awad and Ghaziri: Knowledge Management, 2007)

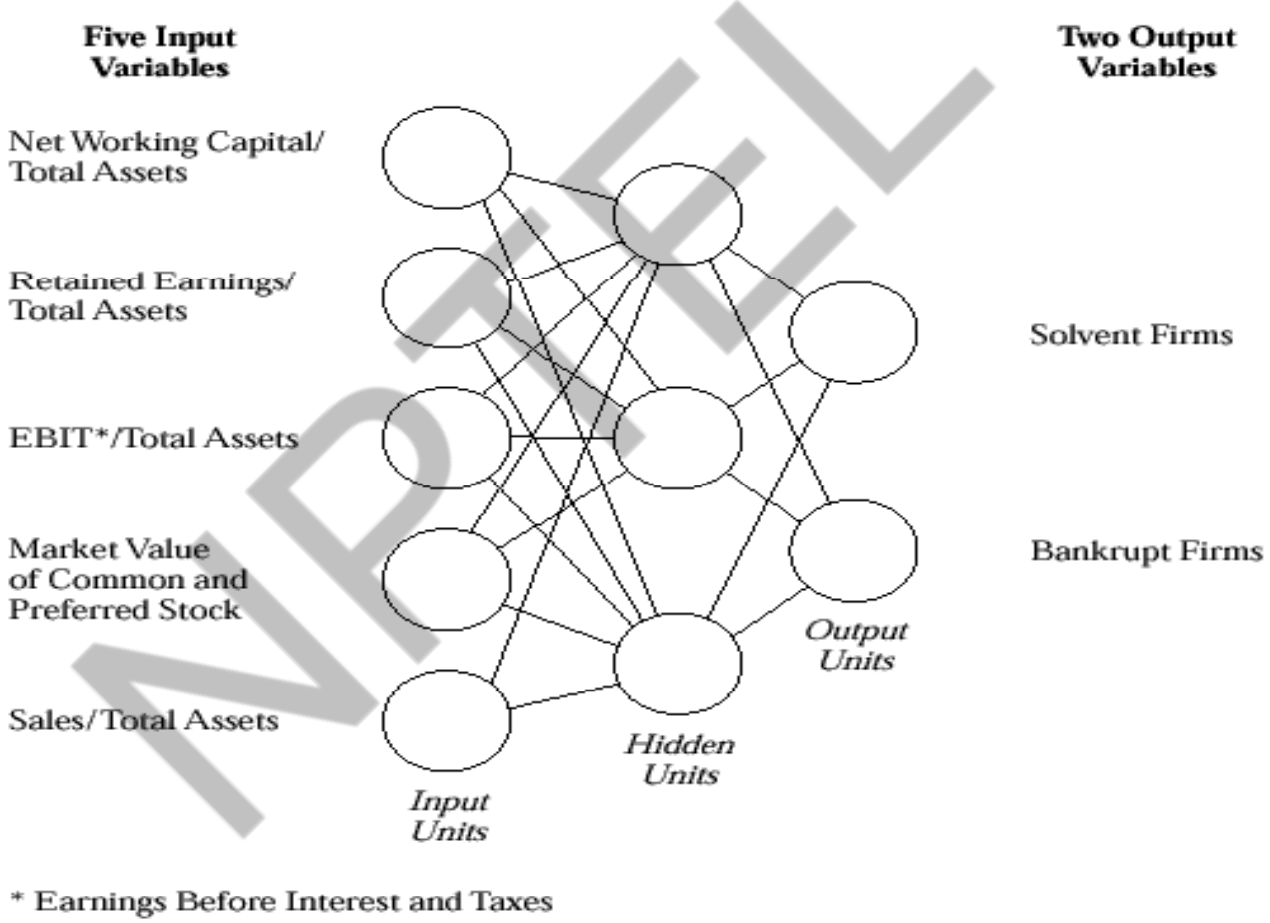


Supervised and unsupervised learning

- **Supervised learning process** requires a teacher represented by a **training set** of examples.
- Each element in a training set is a pair of input and desirable output.
- Network makes successive passes through the examples and the weights adjust toward the goal state. The network has learned to associate a set of input patterns with a specific output.

A Supervised Neural Network Model

Example: Neural network models predicting a firm's bankruptcy



Source: Awad and Ghaziri: Knowledge Management, 2007)

Unsupervised Learning (Self Supervised)

- In **unsupervised** learning, no external factors influence adjustment of the input's weights.
- The neural networks has no advance indication of correct or incorrect answers.
- Adjusts solely through direct confrontation with new experience .
- This is also known as self organization.

Business Applications

- Neural networks are applied in situations having a need for pattern recognition, where the data are dynamic.
- This technology have been applied in all sector.
- Business sector has experienced significant success in application of neural networks.

Business Application

- **Risk management** Appraising commercial loan applications

The network trained on thousands of applications, half of which were approved and the other half rejected by the bank's loan officers

From this much experience, the neural net learned to pick risks that constitute a bad loan

Identifies loan applicants who are likely to default on their payments

Business Application: Some examples

- **Predicting Foreign Exchange Fluctuations:**
  - A set of relevant indicators were identified, then used as inputs to a neural network
  - The system was trained for exchange rates of the US dollar against Swiss franc and Japanese yen, using data from first 6 months of 1990. Then it was tested over an 8-11= 1week period
  - Results revealed return on capital of about 20%

Business Application

- **Mortgage Appraisals:**

Neural network uses the data in the mortgage loan application

It estimates value of the property based on the immediate neighborhood, the city, and the country

The system comes up with a valuation for the property and a risk analysis for the loan

Relative fit with KM

- Neural net exhibits high accuracy and response speeds.
- High input preprocessed data is often required for building a neural net.
- A neural net must start all over with every new application.
- A critical condition for using neural nets in KM is the level of knowledge needed to apply the technology. For example in case based reasoning require a high level of user knowledge (fast solution), but expert and rule based knowledge requires low level of knowledge (slow solution).

Association Rules

- Association rules techniques generate a set of rules to help understand relationship that may exit in data. The main rules are as follows:
- **Boolean Rule**: If a rule consists of examining the presence or absence of items, it is a Boolean Rule
- For example, if a customer buys a PC and a 14” monitor, then he will buy a printer. Presence of items (a PC and 14” monitor) implies presence of the printer in the customer’s buying list

Association Rules

- **Quantitative Rule**: In this rule, instead of considering the presence or absence of items, we consider quantitative values of items
- For example, if a customer earns between Rs 40,000 and Rs 50,000 and owns an apartment worth between Rs250,000 and Rs500,000, he will buy a Car.

Association Rules

- Multi dimensional Rule:
- A single dimensional rule, because it refers to a single attribute, “buying”
- If a customer lives in a big city and earns more than Rs 35,000, then he will buy a cellular phone
  - This rule involves 3 attributes: living, earning, and buying. Therefore, it is a multidimensional rule

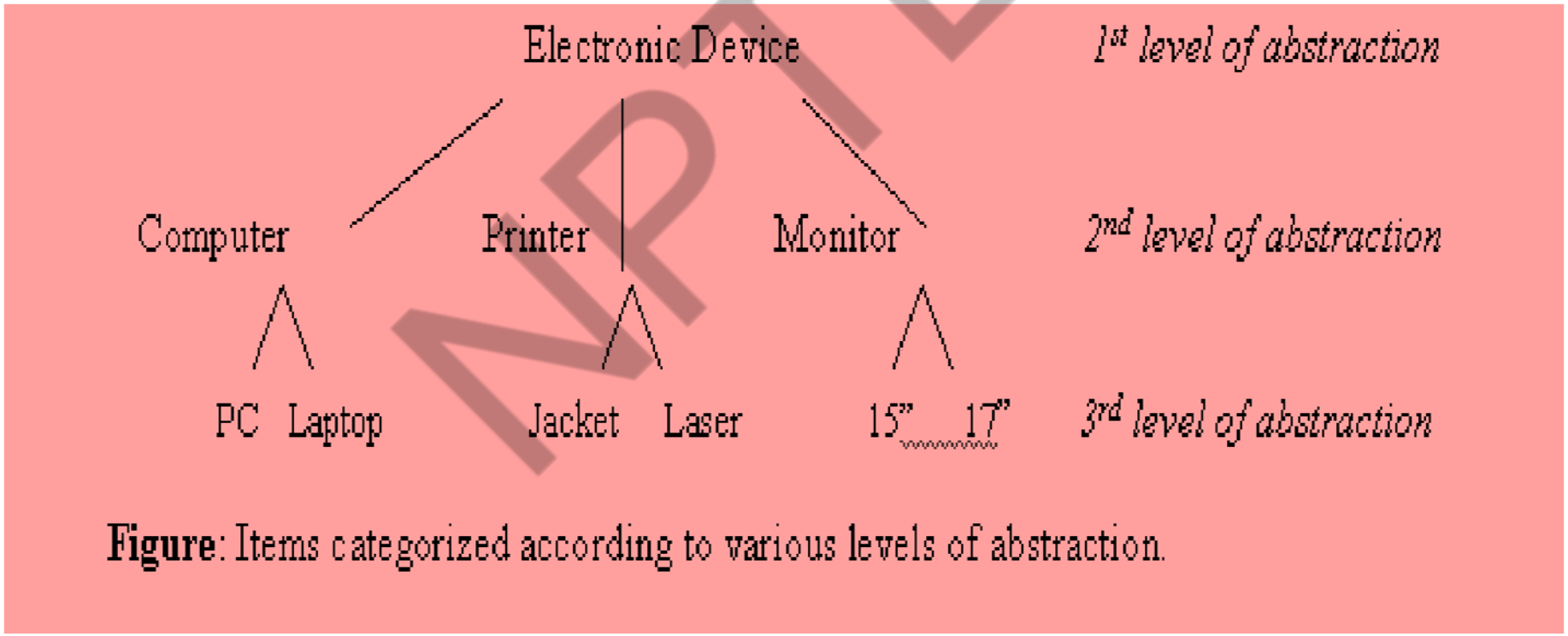


Figure: Items categorized according to various levels of abstraction.

Source: Awad and Ghaziri: Knowledge Management, 2007)

Association Rules

- Statements of the form. When a customer buys a PC, in 70% of the cases he or she will buy a printer; it happens in 14% of all purchases. This means an association rule consisting of 4 elements:
- Rule body (condition of the rule): When a customer buys a PC
- A confidence level: In 70 % of cases
- A rule head (result of the rule): He or she will buy a printer
- A support (how often items in the rule body occur together as a % of the total transactions): It happens in 14% of all purchases

Market based analysis

- Example of PETCO: GIS and business analytics)
- Petco is a leading national pet specialty retailer, with more than 1,200 Petco and Unleashed by Petco store locations nationwide.
- Petco staff members began using an Esri GIS-based site selection solution to mitigate the risks associated with expanding its network of stores.
- As the number of stores has grown, so have the risks of selecting inappropriate or marginal locations or new stores that have the potential to cannibalize the sales of existing stores.
- Petco sought to improve its ability to assess both the sales potential for new locations and any risks. Investing in the Esri solution provided scientific analysis that gave Petco leaders more confidence in their decisions.

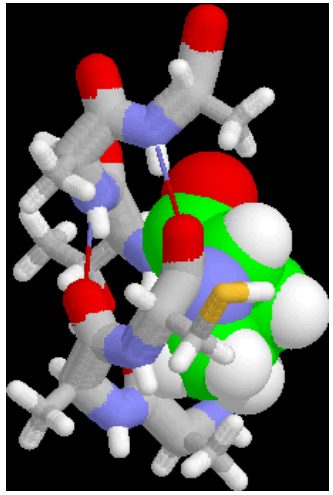
Classification Trees

- Classification tress are powerful tools for classification and prediction. Rules are explained to enable people to understand or a databases is created to see that records fall into a category.
- The concept of tree is derived from graph theory .
- A tree is a network of nodes connected by areas called branches to avoid loops in the network. There is a root node- starting node and the ending nodes are called leaf nodes. These root and leaf nodes are separated by intermediate node organizations in layers called levels.

Examples of Classification Task

- Predicting tumor cells as benign or malignant
- Classifying credit card transactions as legitimate or fraudulent

- Classifying secondary structures of protein as alpha-helix, beta-sheet, or random coil



- Categorizing news stories as finance, weather, entertainment, sports, etc

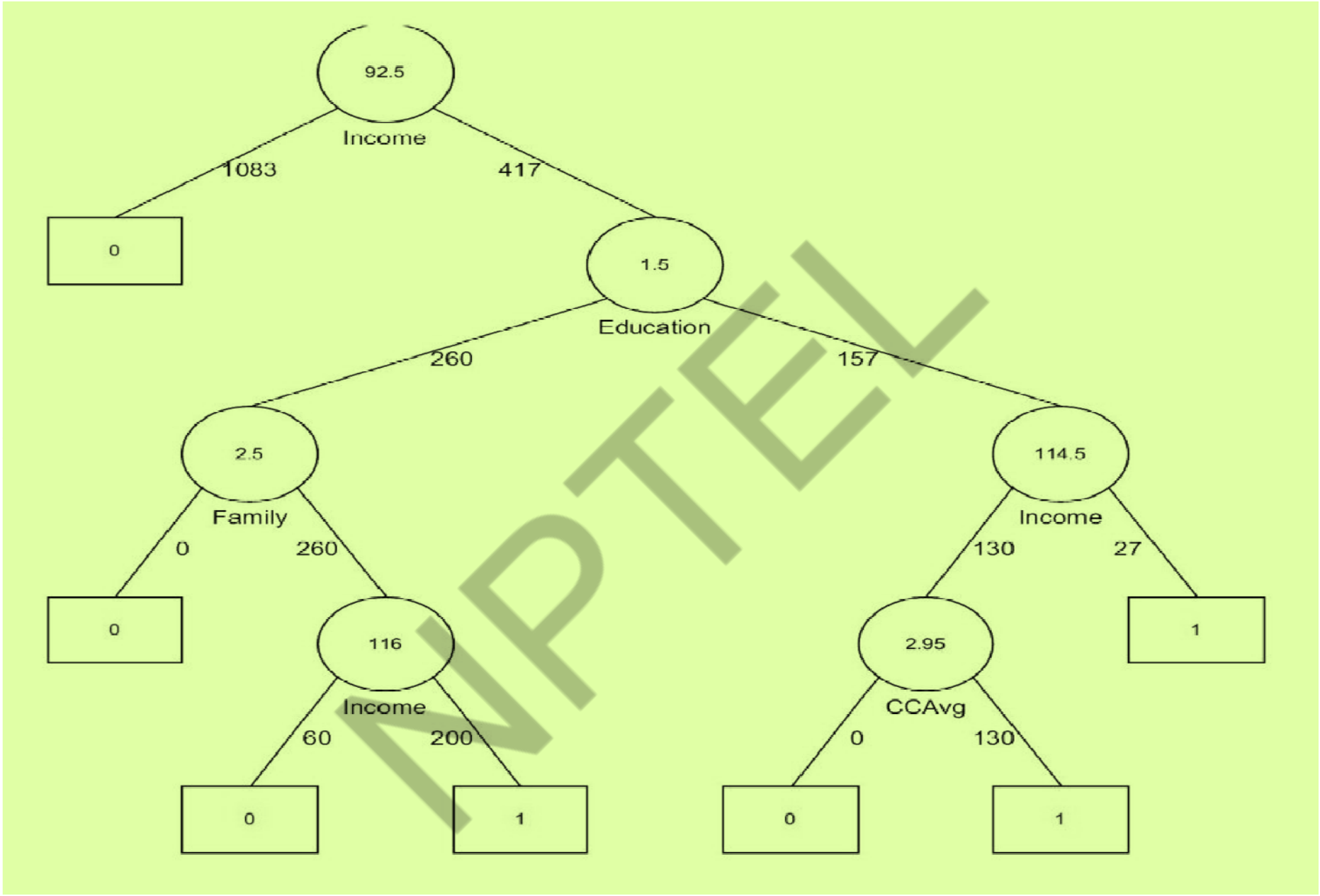
Classification Trees: Rules

Goal: Classify or predict an outcome based on a set of predictors

The output is a set of **rules**

Example:

- Goal: classify a record as "will accept credit card offer" or "will not accept"
- Rule might be "IF (Income > 92.5) AND (Education < 1.5) AND (Family <= 2.5) THEN Class = 0 (nonacceptor)
- Also called CART, Decision Trees, or just Trees
- Rules are represented by tree diagrams



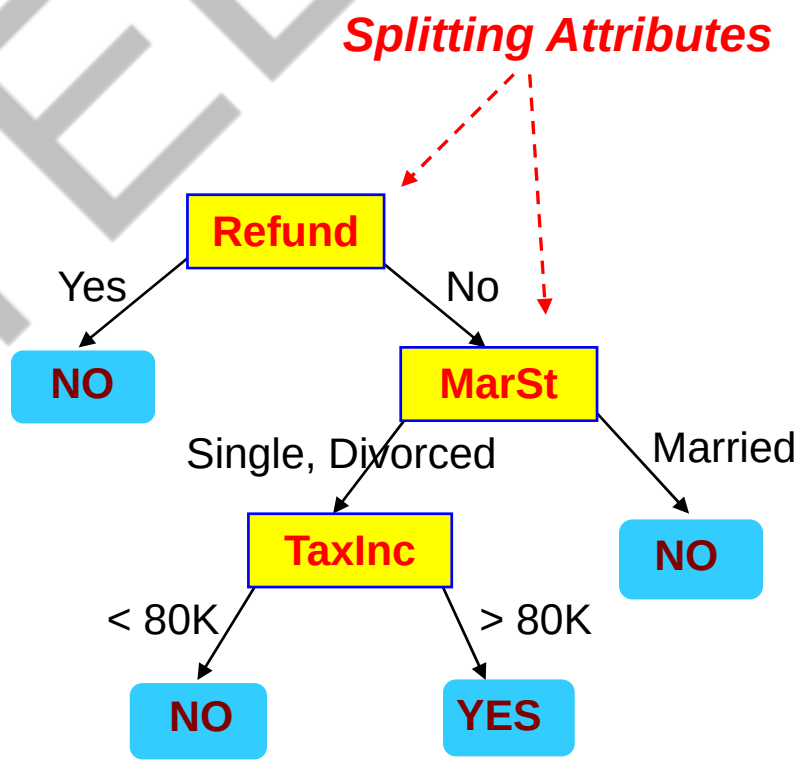
Source: Awad and Ghaziri: Knowledge Management, 2007)

Example of a Decision Tree

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

categorical categorical continuous class

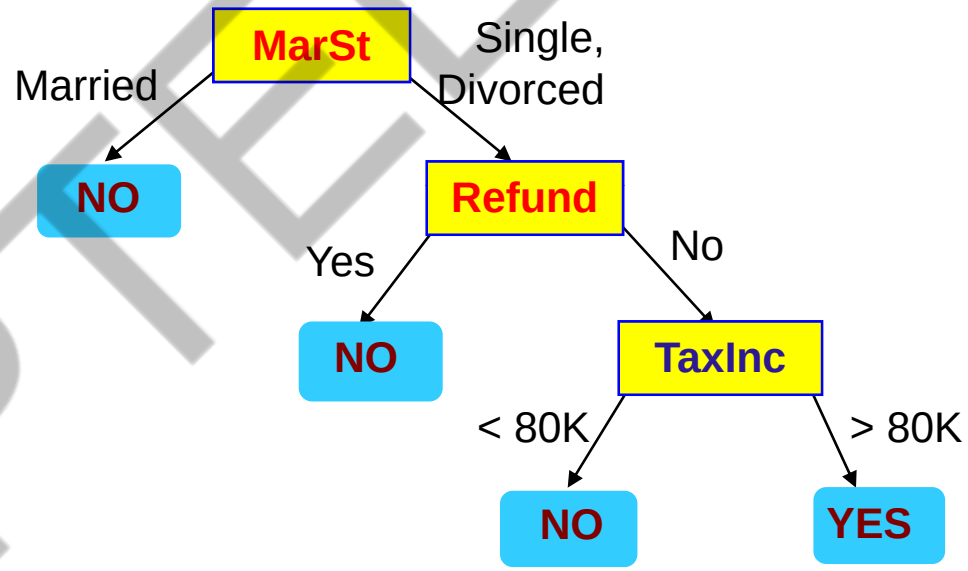
Training Data



Model: Decision Tree

Another Example of Decision Tree

Tid	Refund	Marital Status	Taxable Income	Cheat
		categorical	categorical	continuous	class
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



There could be more than one tree that fits the same data!

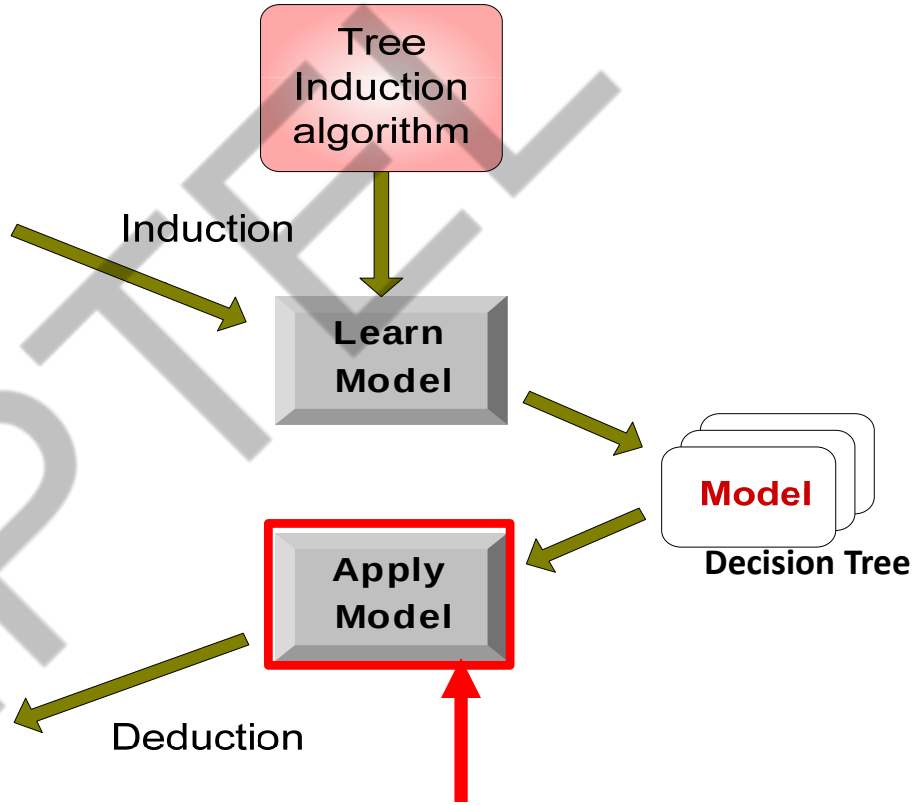
Decision Tree Classification Task

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

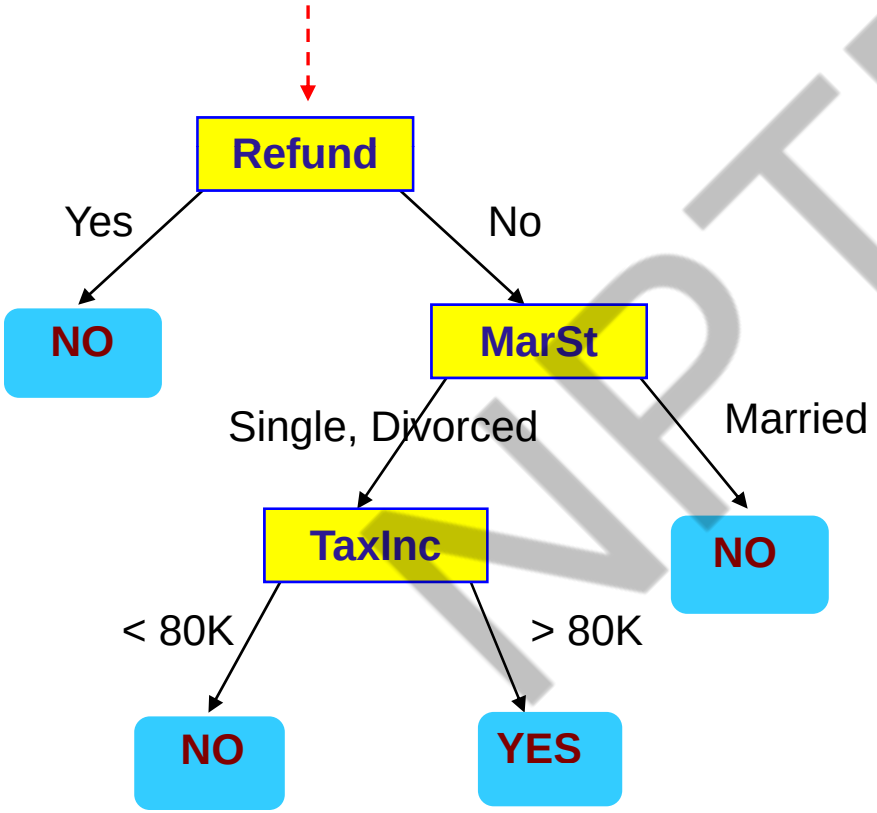
Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set



Apply Model to Test Data

Start from the root of tree.



Test Data

Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?

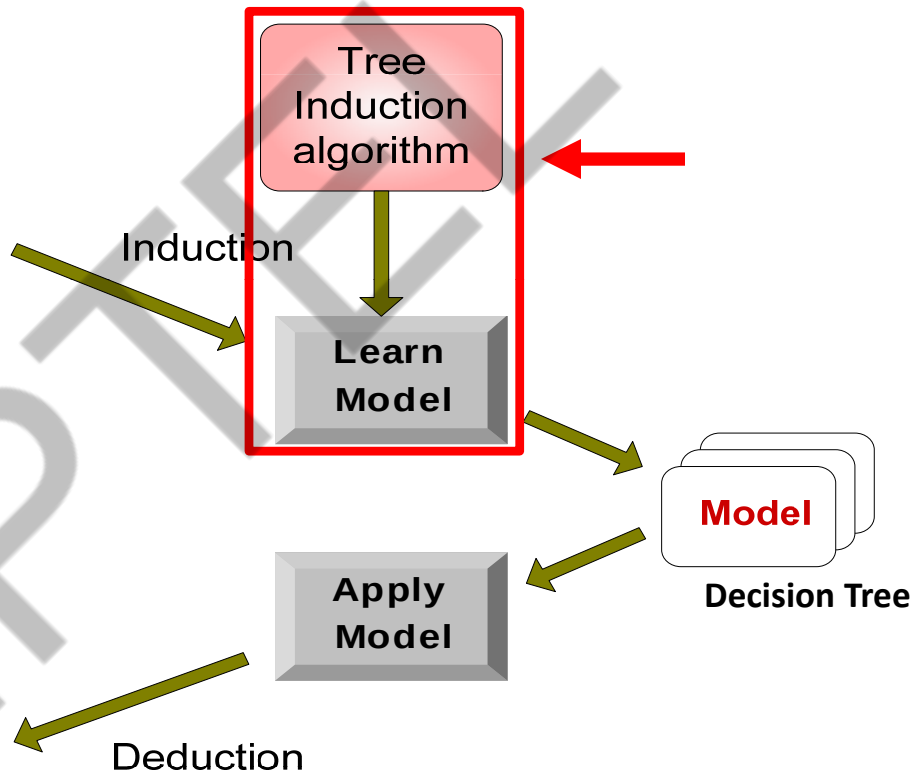
Decision Tree Classification Task

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set



2- Data Mining

- Topics covered

- **Defining data mining**
- **Data mining and business intelligence**
- **Business drivers**
- **Technical drivers- Role of statistics, machine learning, data warehouses, OLAP (on line analytical processing), DM architecture**
- **DM virtual cycle – business understanding and developing DM application**
- **Data management- sources, taxonomy, preparation, model building, parameter setting, tuning, deployment and post deployment phase action**
- **DM in practice**
- **Role of DM in customer relationship management**
- **Application for KM**

Data Mining

- A body of scientific knowledge accumulated through decades of forming well-established disciplines, such as statistics, machine learning, and artificial intelligence
- A technology evolving from high volume transaction systems, data warehouses, and the Internet
- A business community forced by an intensive competitive environment to innovate and integrate new ideas, concepts, and tools to improve operations and DM quality

Data mining

- Data mining (knowledge discovery in databases):
  - Extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from data in large databases

- Alternative names and their "inside stories":
  - Knowledge discovery(mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, business intelligence, etc

Why data mining- motivation

- Data explosion problem
  - Automated data collection tools and mature database technology lead to tremendous amounts of data stored in databases, data warehouses and other information repositories
- We are drowning in data, but starving for knowledge!
- Solution: Data warehousing and data mining
  - Data warehousing and on-line analytical processing
  - Extraction of interesting knowledge (rules, regularities, patterns, constraints) from data in large databases

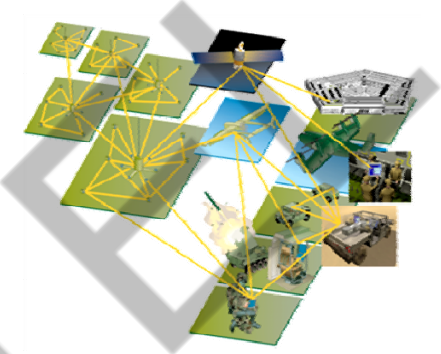
Why Mine Data? Commercial Viewpoint

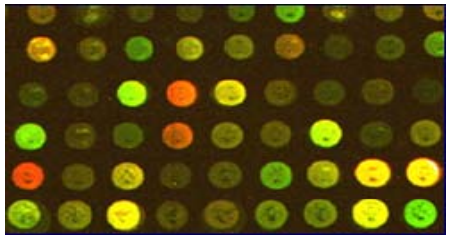
- Lots of data is being collected and warehoused
  - Web data, e-commerce
  - purchases at department/ grocery stores
  - Bank/Credit Card transactions

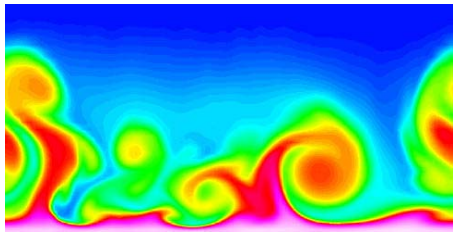
- Computers have become cheaper and more powerful
- Competitive Pressure is Strong
  - Provide better, customized services for an *edge* (e.g. in Customer Relationship Management)

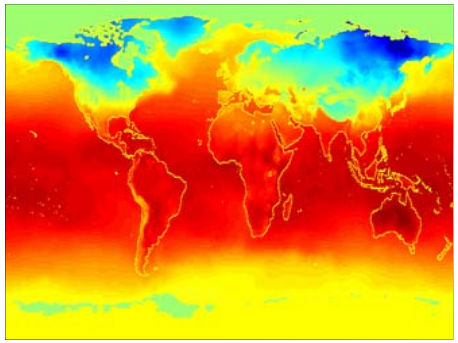
Why Mine Data? Scientific Viewpoint

- Data collected and stored at enormous speeds (GB/hour)
  - remote sensors on a satellite
  - telescopes scanning the skies
  - microarrays generating gene expression data
  - scientific simulations generating terabytes of data
- Traditional techniques infeasible for raw data
- Data mining may help scientists
  - in classifying and segmenting data
  - in Hypothesis Formation









Example of data mining

- Database-

- Find all credit applicants with last names of Sinha.
- Identify customers who have purchased more than 30000 rupees in the last months
- Find all customers who have purchased chocolate ice-creams.
- **Data Mining-**
- Find all credit applicants with poor risk (classification)
- Identify customers with similar buying habits (clustering)
- Find all items which are frequently purchased with ice cream (association rules).

Data Mining: Classification Schemes

- Decisions in data mining
  - Kinds of databases to be mined
  - Kinds of knowledge to be discovered
  - Kinds of techniques utilized
  - Kinds of applications adapted
- Data mining tasks
  - Descriptive data mining
  - Predictive data mining

Decisions in Data Mining

- **Databases to be mined**
  - Relational, transactional, object-oriented, object-relational, active, spatial, time-series, text, multi-media, heterogeneous, legacy, WWW, etc.
- **Knowledge to be mined**
  - Characterization, discrimination, association, classification, clustering, trend, deviation and outlier analysis, etc.
  - Multiple/integrated functions and mining at multiple levels
- **Techniques utilized**
  - Database-oriented, data warehouse (OLAP), machine learning, statistics, visualization, neural network, etc.
- **Applications adapted**
  - Retail, telecommunication, banking, fraud analysis, DNA mining, stock market analysis, Web mining, Weblog analysis, etc.

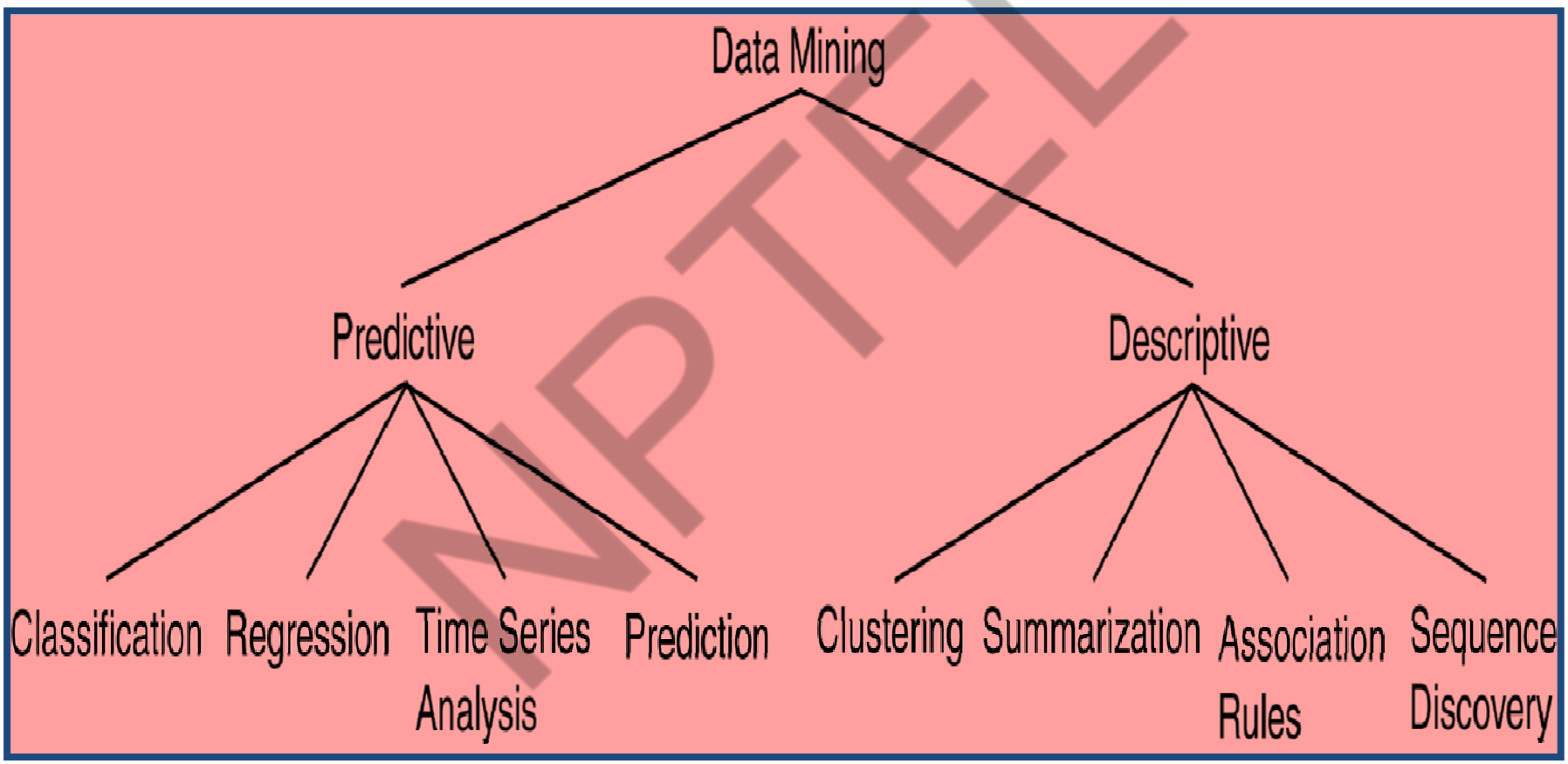
Data Mining Tasks

- Prediction Tasks
  - Use some variables to predict unknown or future values of other variables
- Description Tasks
  - Find human-interpretable patterns that describe the data.

Common data mining tasks

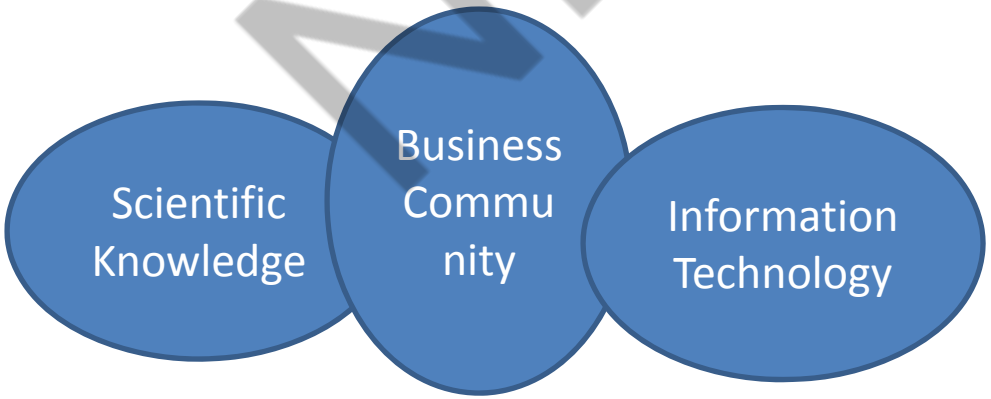
- Classification [Predictive]
- Clustering [Descriptive]
- Association Rule Discovery [Descriptive]
- Sequential Pattern Discovery [Descriptive]
- Regression [Predictive]
- Deviation Detection [Predictive]

Data Mining Models and Tasks



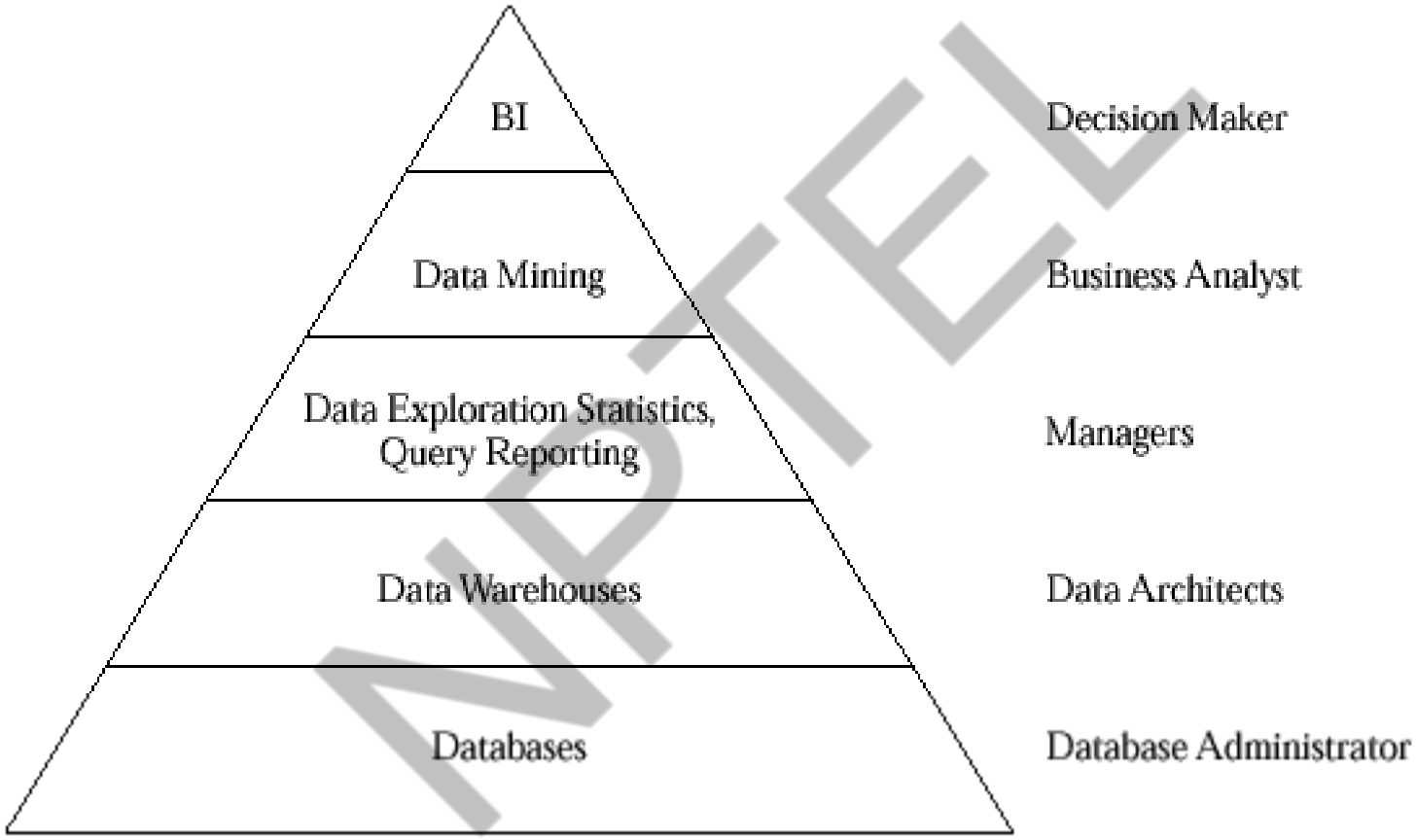
Data mining and business intelligence

- Data mining helps in producing new knowledge and discovering new patterns to describe the data using intelligent automated systems.
- BI is a global processes, techniques, and tools that support business decision making based on IT.
- Approaches can range from a simple spreadsheet to an advanced decision support system
- DM and the three bodies of knowledge-



Source: Awad and Ghaziri: Knowledge Management, 2007)

Data mining as a component of BI



Source: Awad and Ghaziri: Knowledge Management, 2007)

Business Drivers of data mining

- ***Competition.*** Successfully competing in today's economy requires an understanding of customer needs and behavior and flexibility to respond to market demands and competitors' challenges
- ***Information glut.*** Today's manager is confronted with increasingly large volumes of data, collected and stored in various databases and constitutes a challenge for decision making
- ***Serving knowledge workers efficiently.*** Databases provide the firm with transactional memory. Memory is of little use without intelligence--capacity to acquire and apply knowledge

Knowledge worker types and needs

Type	Need
• Knowledge analysts	• Sophisticated tools for their investigations and research analysis
• Knowledge users	• **Review** and analysis of data
• Knowledge consumers	• Easy access to information

Source: Awad and Ghaziri: Knowledge Management, 2007)

Technical Drivers

- Objective of DM is to optimize use of available data and reduce risk of making wrong decisions
- Statistics and machine learning considered the analytical foundations upon which DM was developed

- **ROLE OF STATISTICS**

  With databases organizing data records of hundreds of attributes, the statistics methodology of the hypothesize-and-test paradigm becomes a time-consuming process. DM helps automate formulation and discovery of new hypotheses

- **MACHINE LEARNING**

  A scientific discipline considered a sub-field of artificial intelligence. In contrast, data mining is a business process concerned with finding understandable knowledge from very large real-world databases

- **DATA WAREHOUSES**

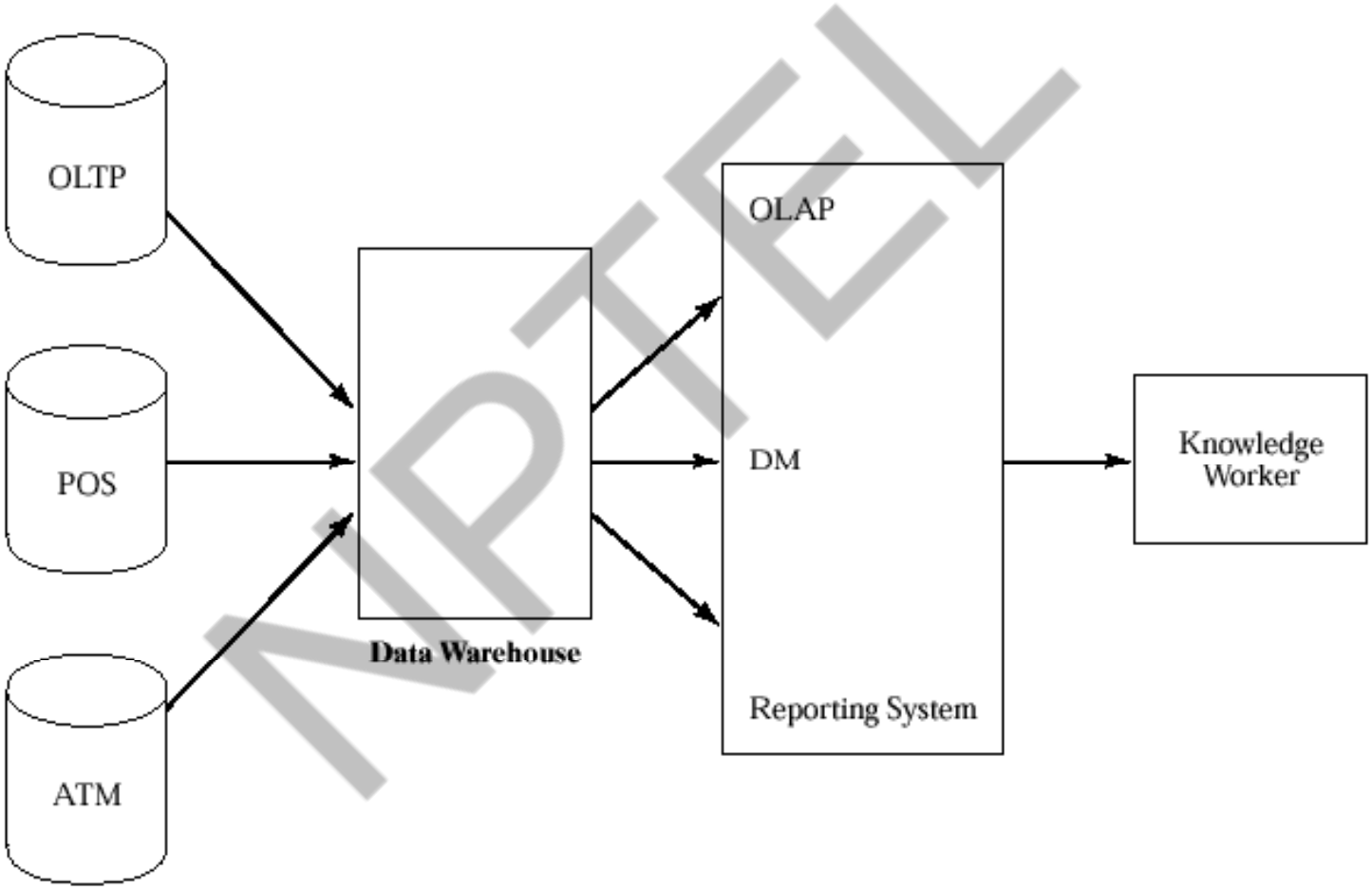
  Extracting and transforming operational data into informational or analytical data and loading it into a central data warehouse

- Major features of DW:
  - **Subject oriented-** Organized around subjects such as customers vendor's products compared to operational systems organized around business functions
  - **Integration-** data are loaded from different operation system may be inconsistent, and integration help in brining consistency in naming and measuring the variables
  - **Time variant-** data may be collected from different time period
  - **Nonvolatile**- no redundancy between operational data houses and data warehouses

OLAP (*Online Analytical Processing)*

- Uses computing power and graphical interfaces to manipulate data easily and quickly at user's convenience
- Focus is showing data along several dimensions. Manager should be able to drill down into the ultimate detail of a transaction and zoom up for a general view
- Strengths- Visualization, easy to use interactive tool, helps to understand the data
- Limitations- does not find pattern automatically, not a powerful analytical tool.

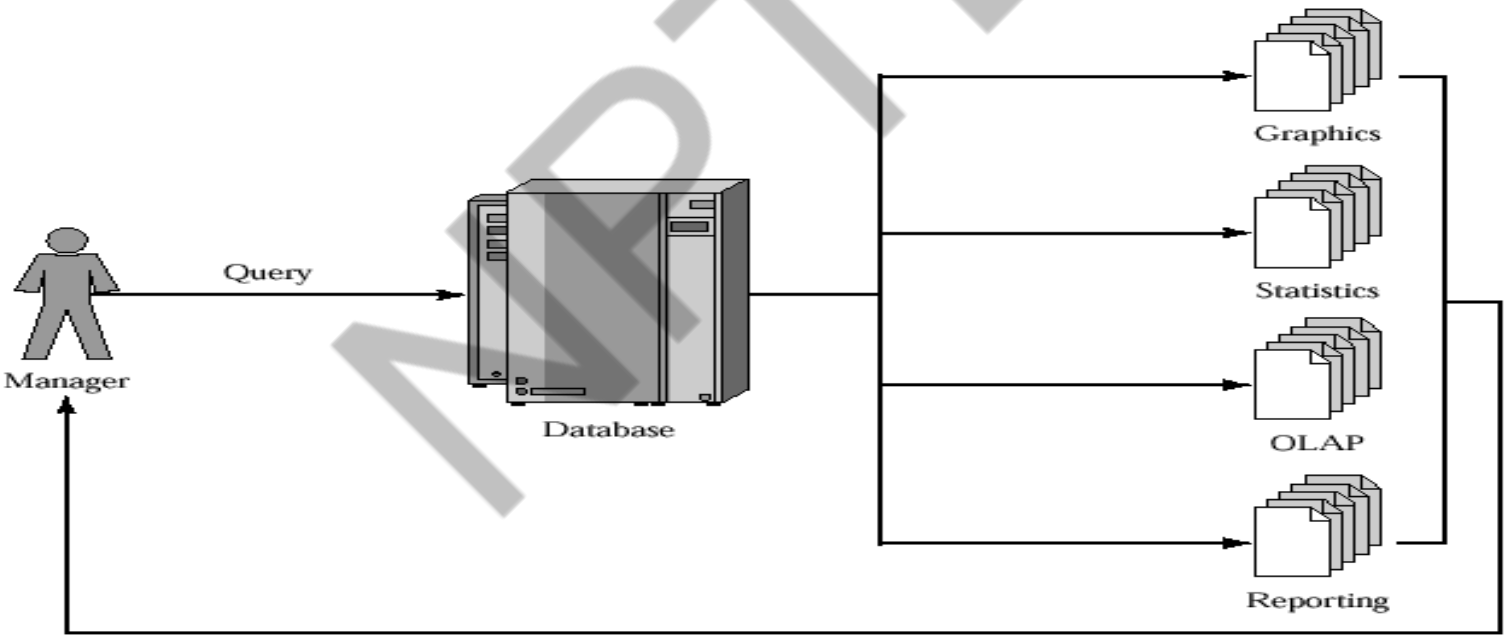
Data Warehouse and DM Technological Framework:



Source: Awad and Ghaziri: Knowledge Management, 2007)

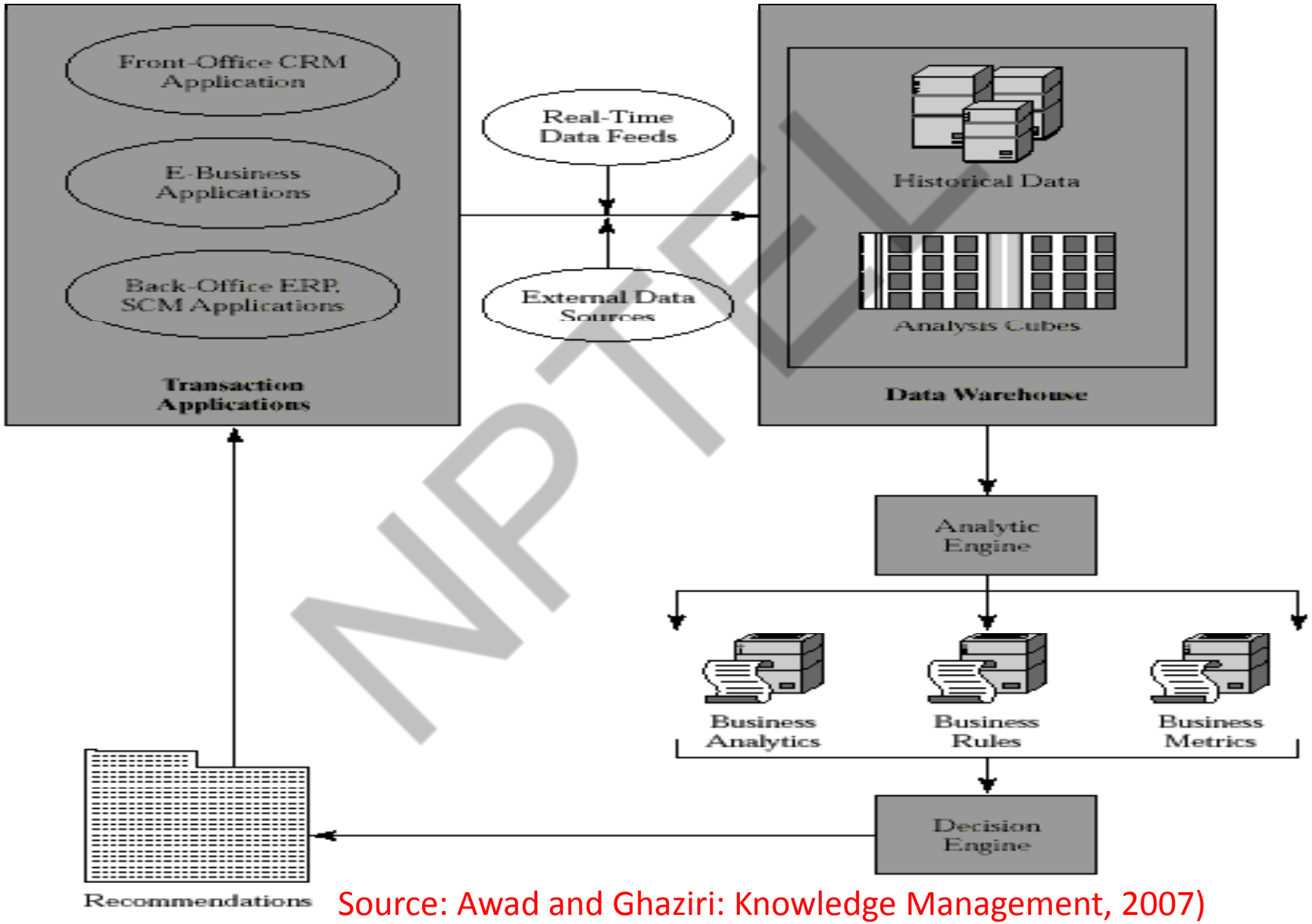
Evolution of the Decision Making Architecture

- Integration of decision processing into the overall business process achieved by building a closed-loop system, where decision processing applications output delivered to users in the form of recommended actions
- In the e-business environment, many companies look to extend closed-loop processing to *automatically* adjust business operations based on messages generated by a decision engine
- Conventional approach to corporate decision making-



Source: Awad and Ghaziri: Knowledge Management, 2007)

DM in the Context of the New Corporate Business Model



Source: Awad and Ghaziri: Knowledge Management, 2007)

DM Virtuous Cycle

- Harnessing power of data and transforming it into added value for the entire organization

- **Capabilities:**

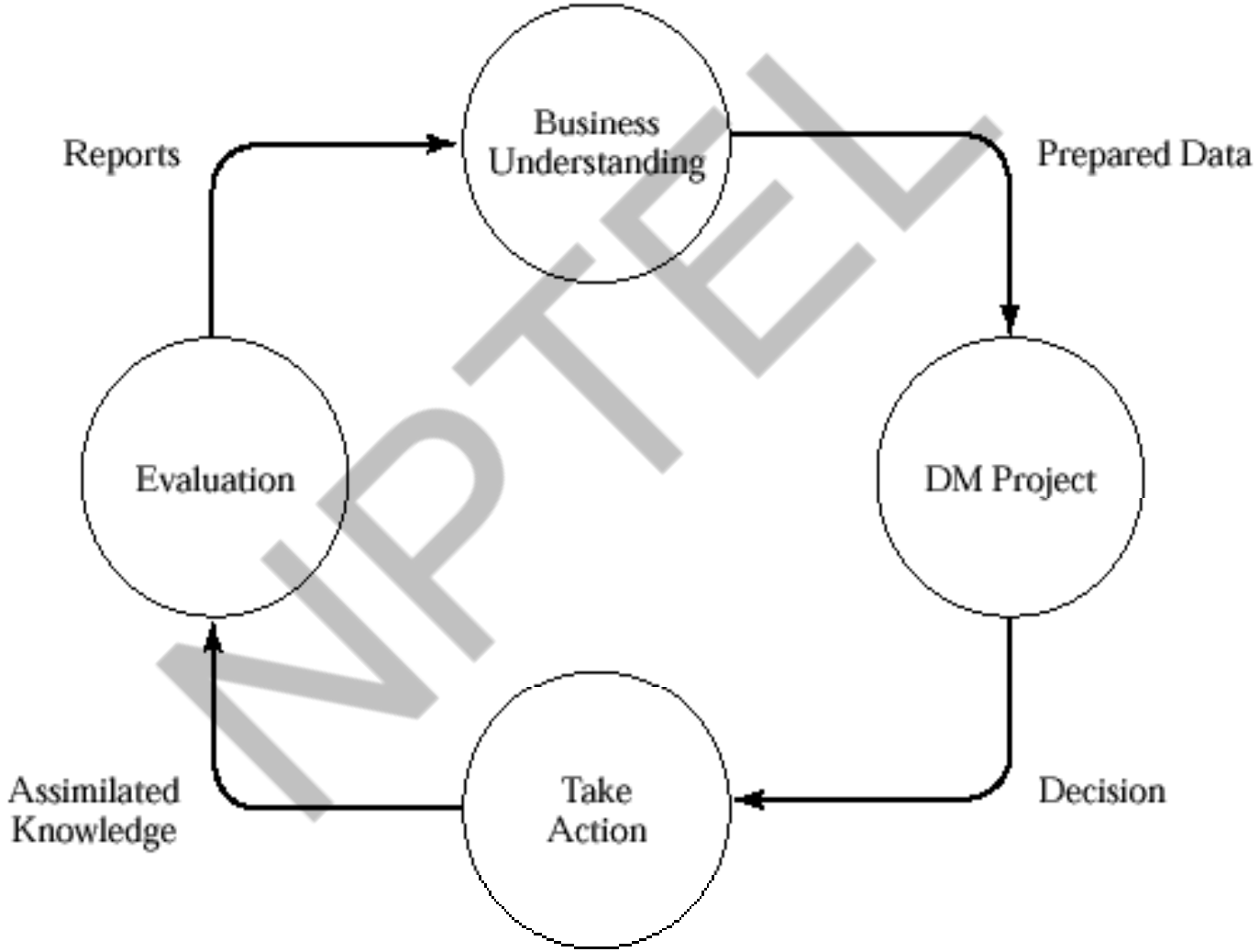
  Response to extracted patterns

  Selection of the right action

  Learning from past actions

  Turning action into business valu

DM Virtuous Cycle



Source: Awad and Ghaziri: Knowledge Management, 2007)

BUSINESS UNDERSTANDING

- The first step in the virtuous DM cycle is identifying the business opportunity--defining the problems faced by the firm
- The goal is to identify areas where data can provide values
- Defining the problem should involve technical and business experts

DEVELOP THE DM APPLICATION

- Define the adequate data-mining tasks
- Organize data for analysis
- Use the right DM technique to build the data model
- Validate the model

DM tasks

Clustering:

- Clustering finds groups that are very different from each other, but whose members are similar to each other
- One does not know what the clusters will be at the start or by what attributes the data will be clustered

DM tasks

Classification

The classification function identifies characteristics of the group to which each case belongs

Affinity grouping

A descriptive approach to exploring data that can help identify relationships among values in a database. The two most common approaches are:

- Association discovery
- Sequence discovery

Different Industries and the DM Goals

DM application in Customer services

Business Challenge--understanding that individual preferences of customers is the key to satisfying them

DM Goals

- Customer acquisition profile
- Customer-centric selling
- Online shopping
- Staffing level prediction
- Targeting market
- Customer retention profiling
- Inquiry routing
- Scenario notification
- Web mining for prospects

DM application in Financial services industry

Business Challenge--Retaining customer loyalty is of the utmost importance to this industry

DM Goals

- Focused statistical and DM applications are prevalent
- Risk management for all types of credit and fraud detection

DM application in Health-care business

Business Challenge

- Keeping pace with rate of technological and medical advancement
- Cost is a constant issue in ever-changing market

DM Goals

- Early DM activities have focused on financially oriented applications
- Predictive models have been applied to predict length of stay, total charges, and even mortality

DM application in Telecommunications industry

Business Challenge

- Keeping pace with rate of technological change
- Deregulation is changing business landscape, resulting in competition from a wide range of service providers
- Finding and retaining customers

DM Goals

- Customer profiling
- Subscription fraud and credit applications are utilized throughout the industry
- Concerns about privacy and security are likely to result in DM applications targeted to these areas

Data Management

- The most challenging part of a DM project is the data management stage
- At least 40 percent of a DM project is spent on this stage

Sources of data:

- Flat files
- Relational databases
- Data warehouses
- Geographical databases
- Time series databases
- World Wide Web

TAXONOMY OF DATA

Data can be found in several forms:

- Business transactions
- Scientific data
- Medical data
- Personal data
- Text and documents
- Web repositories

DATA PREPARATION

- Evaluating data quality
- Handling missing data
- Processing outliers
- Normalizing data
- Quantifying data

MODEL BUILDING

The first step is to select the modeling technique.

The most popular techniques:

- Association rules
- Classification trees
- Neural networks

PARAMETER SETTINGS AND TUNING

- The set of initial and intermediate parameters must be recorded for eventual use and comparison
- The selection of parameters must also be explained
- The process of reaching the final values must be documented
- Testing and validating samples are used for this task

MODEL TESTING AND ANALYSIS OF RESULTS

- Reviewing business objectives and success criteria
- Assessing success of the DM project to ensure all business objectives have been incorporated
- identifying factors that have been overlooked
- Understanding data-mining results
- Interpreting the results
- Comparing results with common sense and knowledge base to detect any worthwhile discoveries

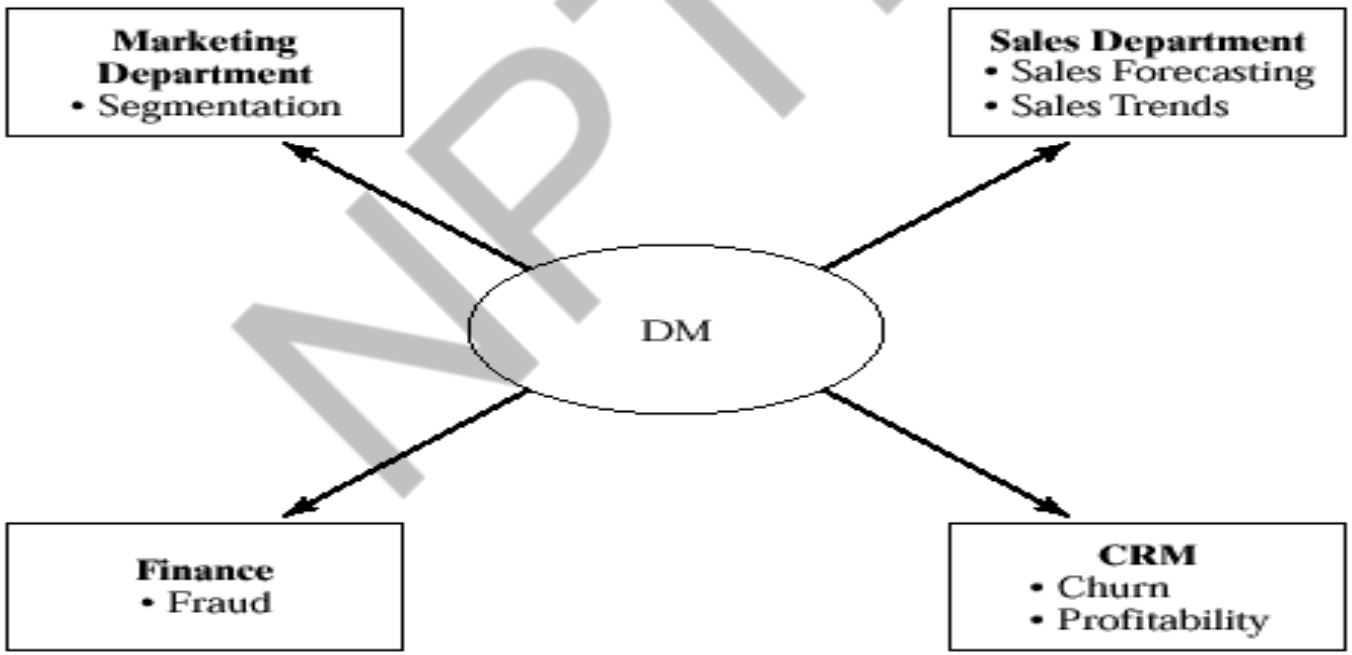
TAKING ACTION AND DEPLOYMENT

- Summarizing deployable results
- Identifying users of the discovered knowledge and finding out how to deliver and propagate it
- Defining a performance measure to monitor obtained benefits by implementation of DM results

DM Applications

- Market management applications
- Sales applications
- Risk management applications
- Web applications
- Text mining

DM as a Required Component for Various Applications in Different Departments

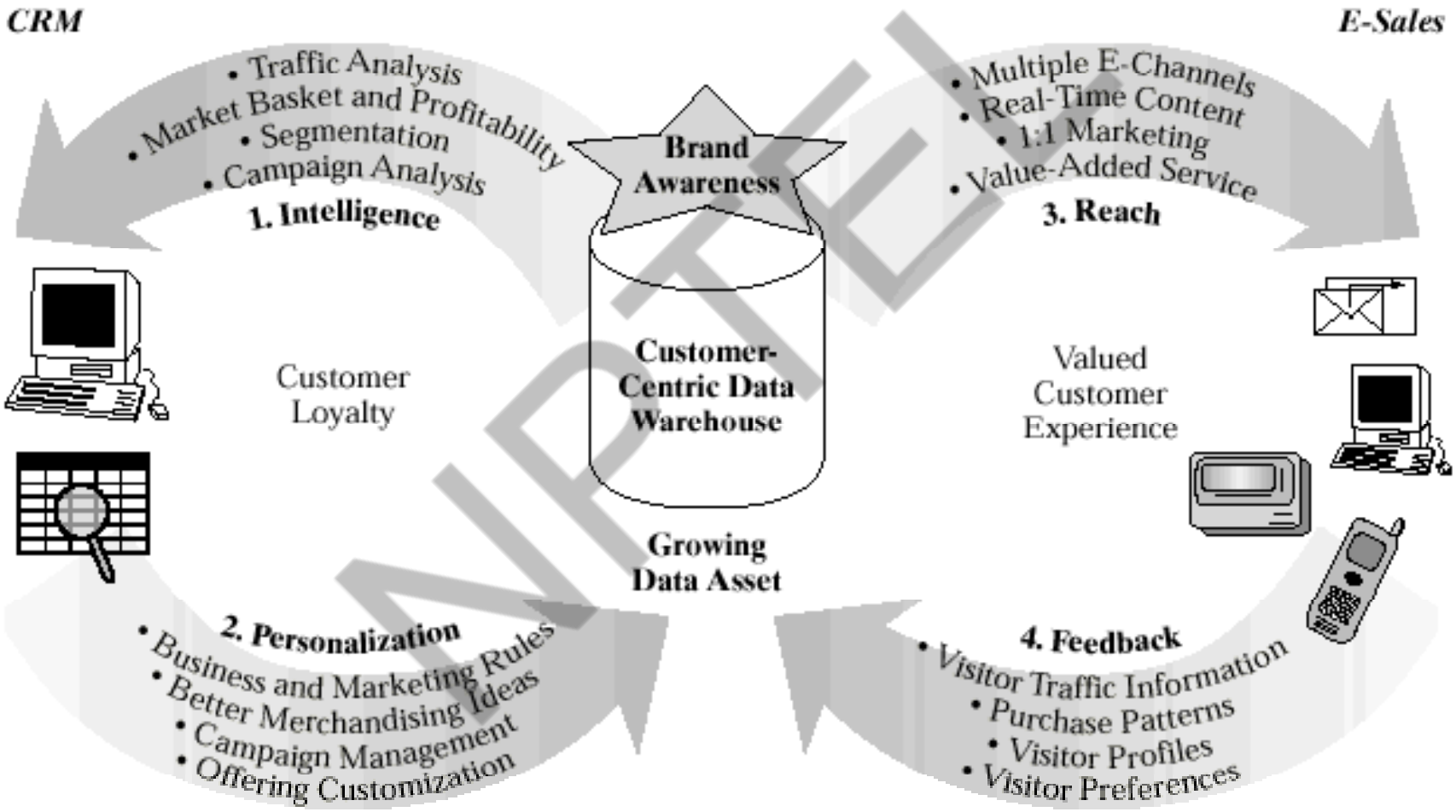


Source: Awad and Ghaziri: Knowledge Management, 2007)

INTEGRATING DM, CRM, AND E-BUSINESS

- DM applications are the first line in understanding the customer and an integral key to segmenting the market
- An intelligent e-business system enhances CRM by enabling a level of responsiveness and proactive customer care not achievable through other channels
- Through personalization, corporations can build successful 1-to-1 relationships with customers

Customer-Centric Data Warehousing Enables Next Generation E-Commerce

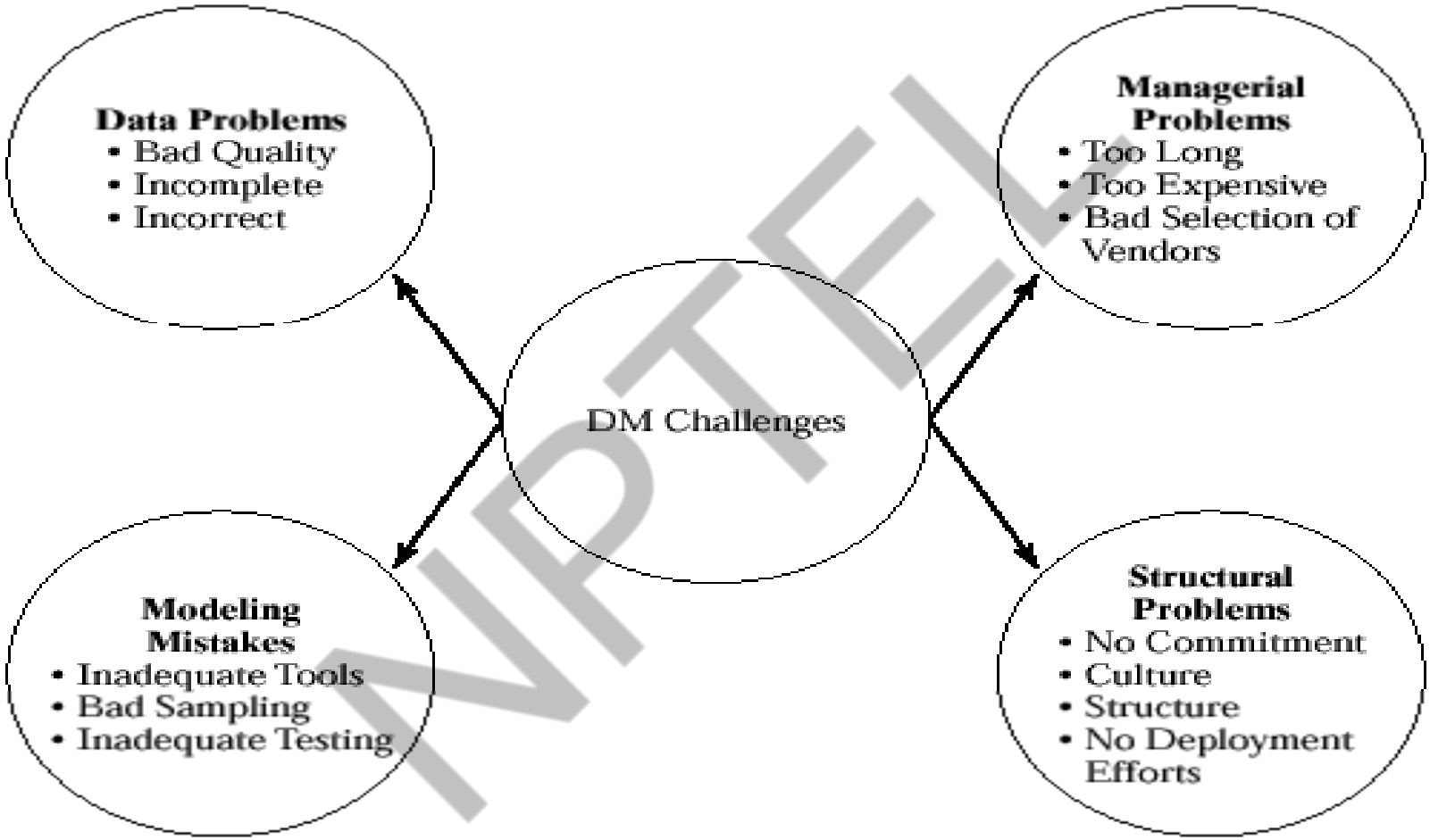


Source: Awad and Ghaziri: Knowledge Management, 2007)

Implications for Knowledge Management

- A DM project is definitely not a straightforward project
- While conducting such a project, companies may face many problems, obstacles, and pitfalls that prevent them from gaining returns on investing in a DM project

Data-Mining Challenges



Source: Awad and Ghaziri: Knowledge Management, 2007)

Insufficient understanding of business needs

- Companies should have a business justification for DM
- The organization will only reap the benefits of DM if there is a real business case to answer
- It should not be an experiment for experiment's sake

Careless handling of data

- Overquantifying data
- Miscoding data
- Analyzing without taking precautions against sampling errors
- Loss of precision due to improper rounding of data values
- Incorrectly handling missing values

Invalidly validating the data-mining model

- Most data miners will face either abundant amounts of testing data or extremely scare amounts
- It is important to never ignore suspicious findings
- Haste can lead to believing everything data owners tell us about the data and, even worse, believing everything our own analysis tells us

Believing in alchemy

- Many data owners believe that data mining is a form of an alchemic process that will magically transform their straw databases into golden knowledge
- Failure may come from wrong suspicion, technology misapplication, or unsuitable information
- Managers should consider:
  - Select the right application
  - In-house development or outsourcing
  - Assessing vendor selection
  - Team building

3. Knowledge Portals

- Topics Covered-

- Basics of portal- evolution and key characteristics
- The business challenges – transforming business, market potential
- Knowledge portal technologies- collaboration, content management, intelligent agent
- Implications for KM- building enterprise portal, sponsorship, bandwith, and portal product selection

Portals: The Basics

Portals are considered to be virtual workplaces that:

- Promote knowledge sharing among different categories of end users
- Provide access to stored structured data
- Organize unstructured data

Portals are tool that could:

Simplify access to data stored in various application systems

Facilitate collaboration among employees

Assist the company in reaching its customers

Knowledge portals

Allow producers and users of knowledge to interact.

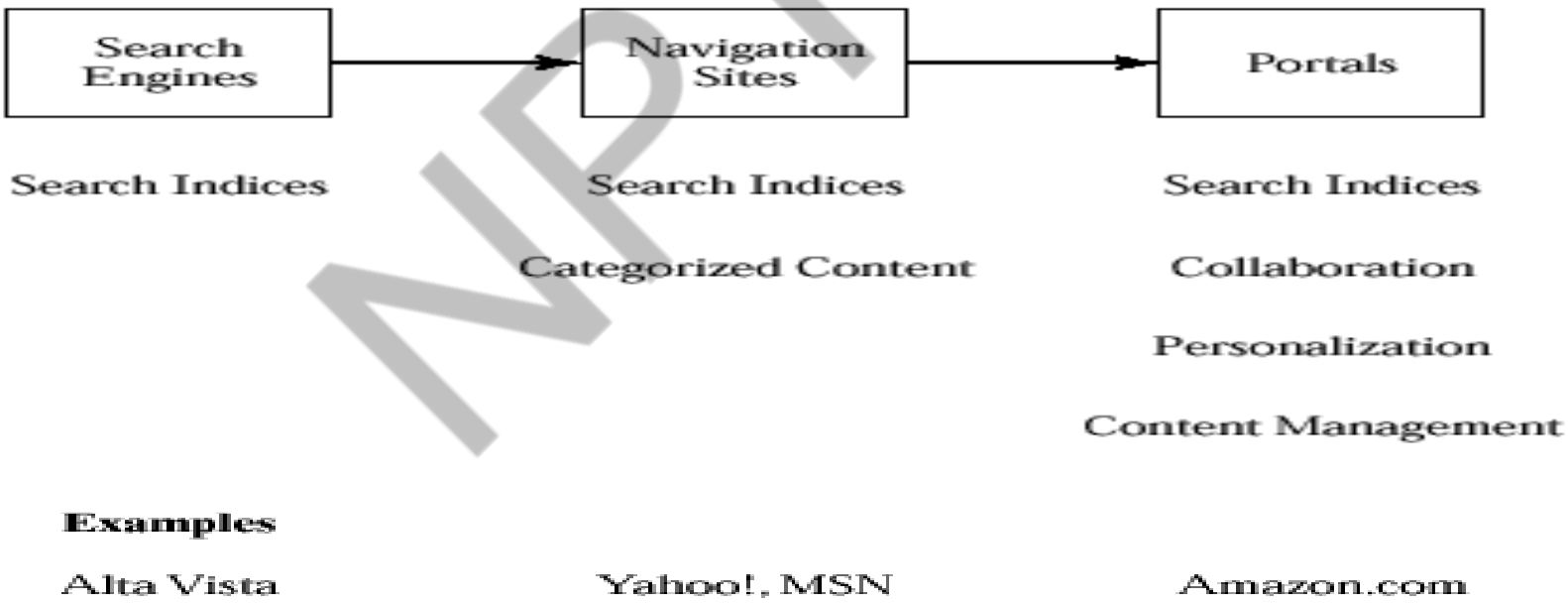
Knowledge portals provide two kinds of interfaces:

The knowledge producer interface

The knowledge consumer interface

EVOLUTION OF PORTALS

- Search engines
- Navigation sites
- portals evolved to include advanced search capabilities and taxonomies
- Evolution of the Portal Concept:



Knowledge Portals Versus Information Portals

Enterprise Information Portals

- Use both “push” and “pull” technologies to transmit information to users through a standardized Web-based interface
- Integrate disparate applications into a single system
- Have the ability to access both external and internal sources of data

Enterprise Knowledge Portals

- Are goal-directed toward knowledge production, knowledge acquisition, knowledge transmission, and knowledge management
- Are focused on enterprise business processes
- Provide, produce, and manage information about the validity of the information they supply
- Include all EIPs functionalities

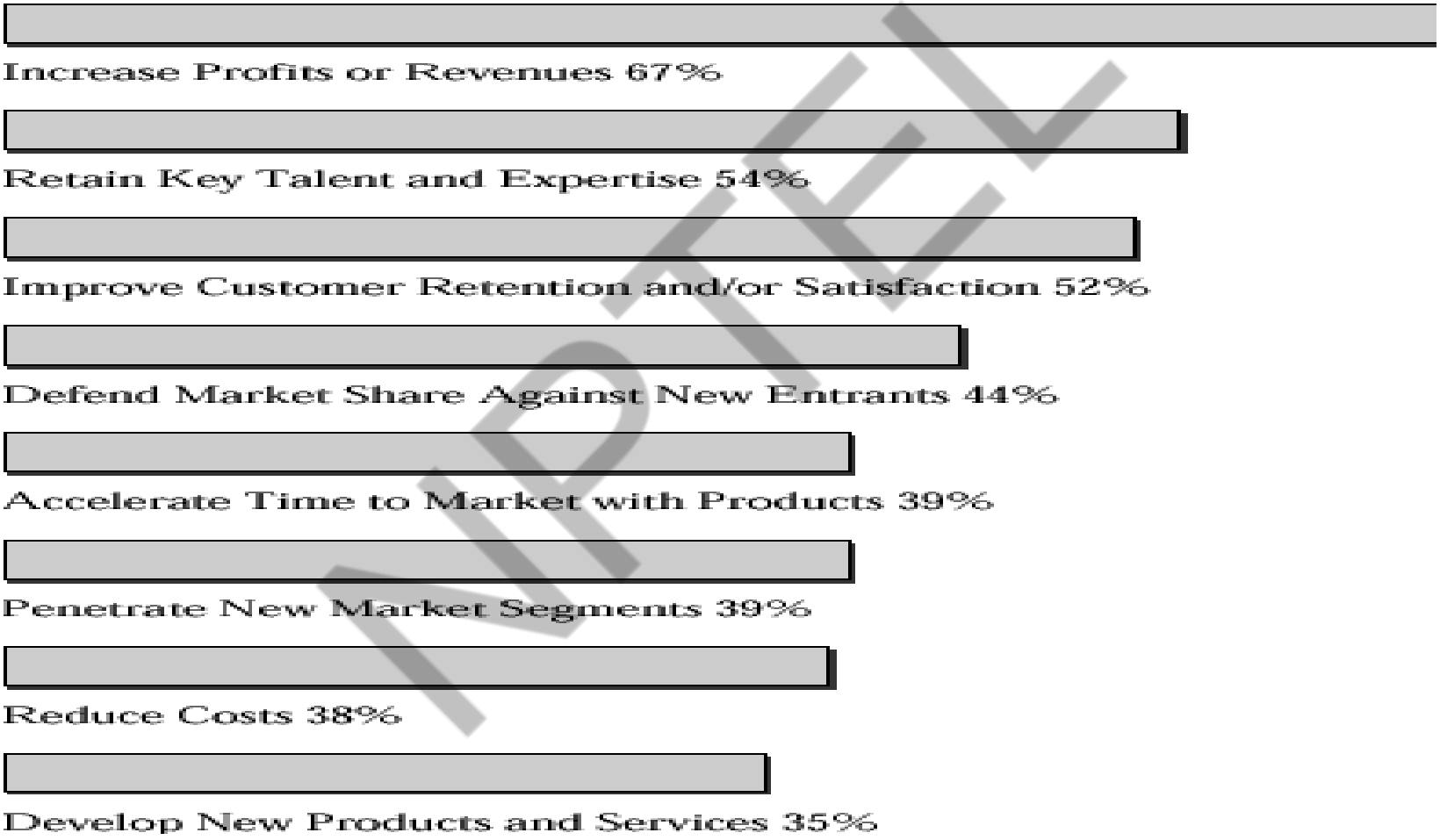
Business Challenges of knowledge portals

- To optimize the performance of operational processes in order to reduce costs and enhance quality
- Companies need to commercialize their products at the lowest price possible

Portals and Business Transformation

- The explosion of key business information captured in electronic documents
- The speed by which the quantity and kinds of content is growing
- Challenges:
  - Shorter time to market
  - Knowledge worker turnover
  - More demanding customers and investors

Why Organizations Launch KM Programs



Source: Awad and Ghaziri: Knowledge Management, 2007)

The Benefits of Knowledge Portals

Productivity

Locating Documents

Collaboration

Better Decisions

Quality of Data

Sharing Knowledge

Identifying Experts

E-mail Traffic

Bandwidth Use

Time in Meetings

Phone Calls

Response Times

Redundant Efforts

Operating Costs

Time to market

Knowledge Portals components and technologies

- **Components:**

- Content management
- Business intelligence
- Data warehouses and data marts
- Data management

- **Technologies:**

-Gathering
- Categorization
- Distribution
- Collaboration
- Publish
- Personalization
- Search/navigate

Portal Features and Benefits

Common Features	Business Benefits
Search	Quick access to hidden information to **facilitate business processes**
Categorization	Ability to organize information assets by business process, group, or job category thus **promoting access to relevant information**
Query, Reporting, and Analysis	**Better decision support** as well as information dissemination and sharing
Integration of Information and Applications	Ability to access through a single interface, all applications and information required for **increased job throughput**
Publish and Subscribe	Maturation of business processes by collaborating with others, sharing information, and **improving business performance**
Personalization	Arranging the interface to meet an individual's needs and desires for **increased job productivity**

Source: Awad and Ghaziri: Knowledge Management, 2007)

Types of collaborations

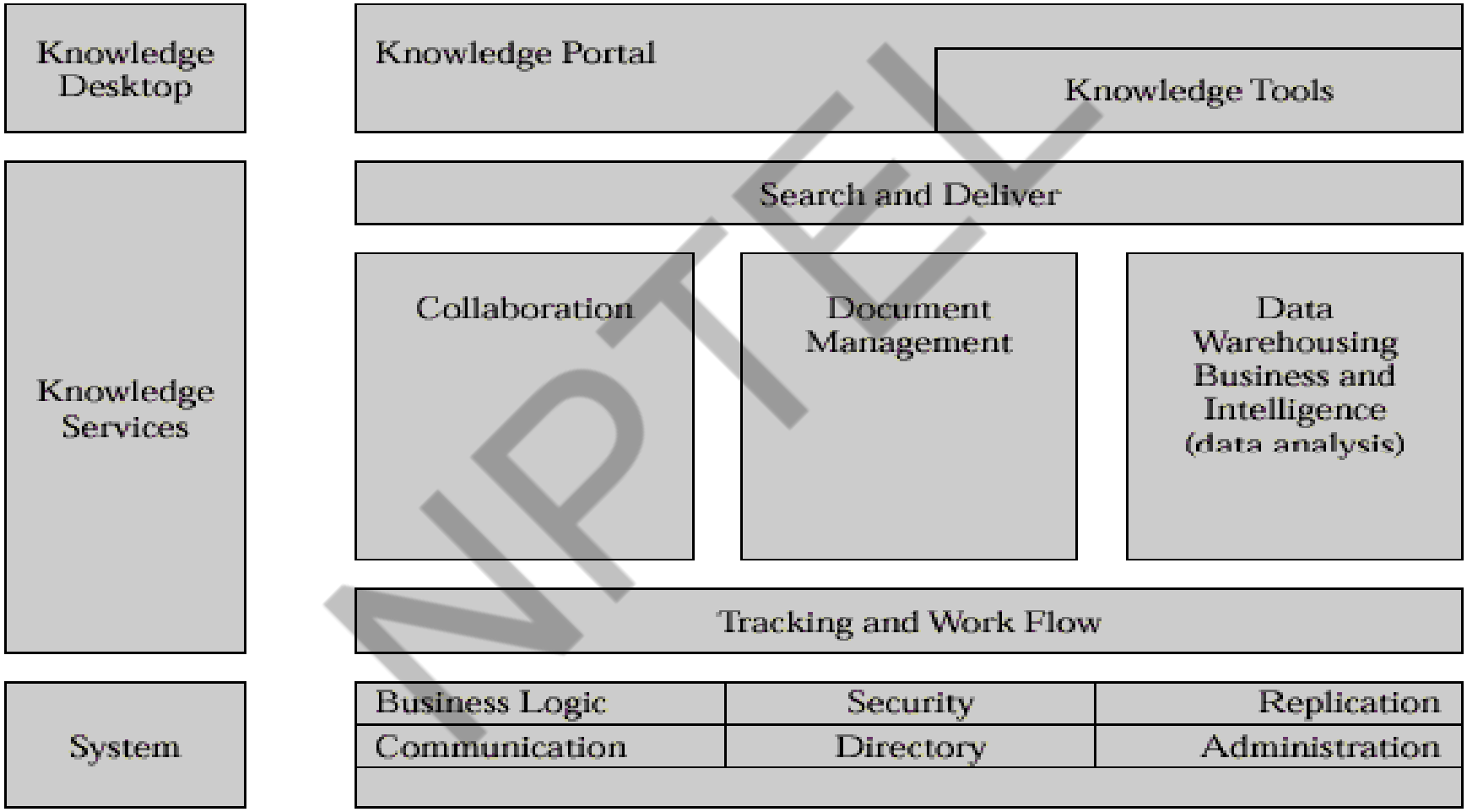
- Asynchronous collaboration is human-to-human interactions via computer sub-systems having no time or space constraints. Queries, responses, or access occur anytime and anyplace.
- Synchronous collaboration is computer-based, human-to-human interaction that occurs immediately (within 5 seconds). It can use audio, video, or data technologies.

Another distinction:

Push technology places information in a place where it is difficult to avoid seeing it.

Pull technologies require you to take specific actions to retrieve information

Layers of The Portal Architecture



Source: Awad and Ghaziri: Knowledge Management, 2007)

Requirements for Successful Collaboration Tools

- Comfortable e-mail systems
- A Web browser
- Simple search functionalities
- Collaboration services with a multipurpose database
- Web services
- Indexing services for full-text search of documents
- Well-organized central storage locations

Synchronous and Asynchronous collaboration

Synchronous collaboration

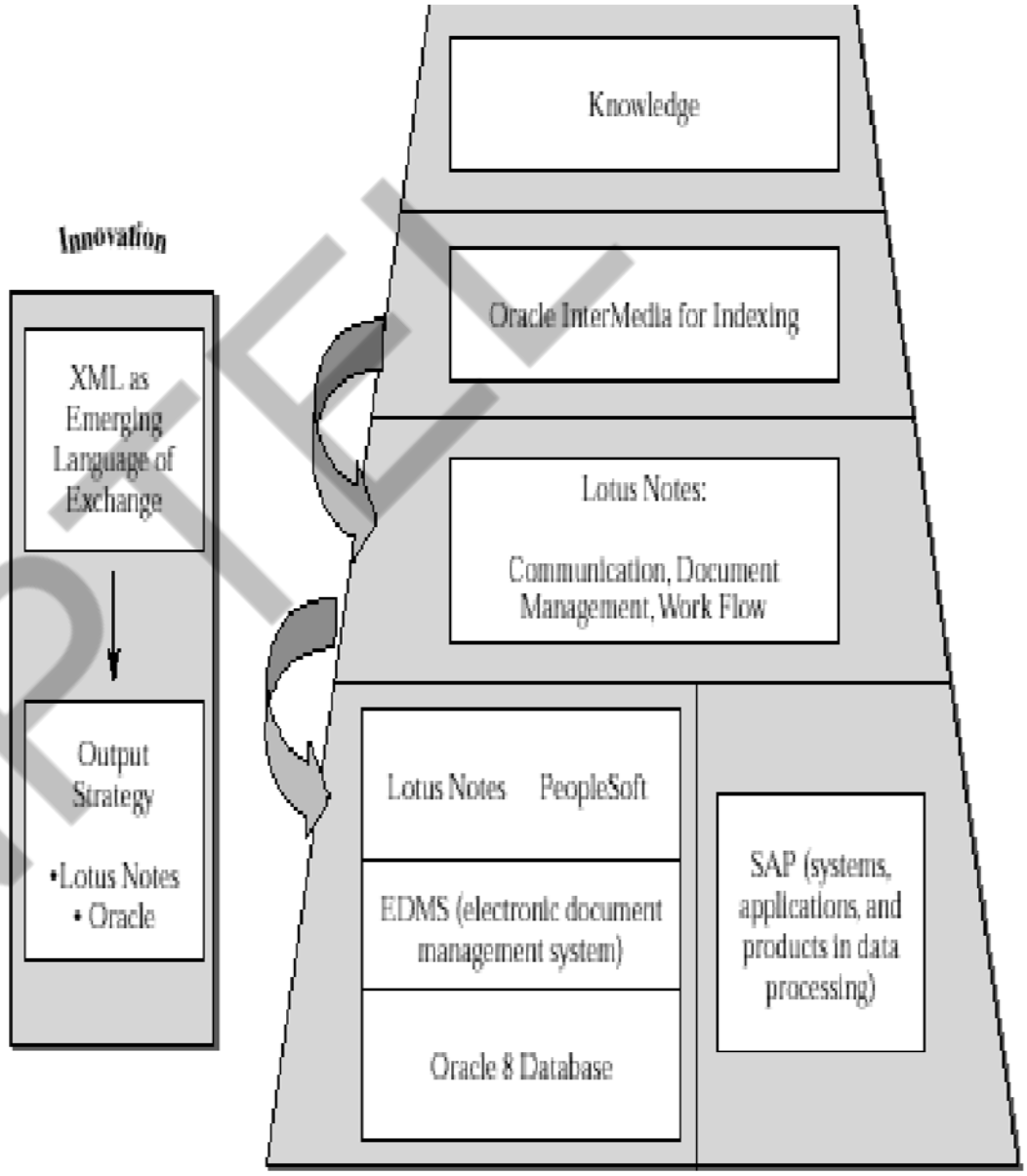
- **Teleconferencing**
  - *Advantages:* personal, immediate feedback
  - *Disadvantages:* expensive, often does not work well across time zones
- **Computer Video/ Teleconferencing**
  - Computer-based teleconferencing and video-conferencing is a rapidly evolving technology that has tremendous potential for distributed organizations.
- **Online Chat Forums**
  - Allow multiple users to communicate simultaneously by typing messages on a computer screen.

Asynchronous collaboration

- **Electronic Mailing Lists**
  - *Advantages:* cheap
  - *Disadvantages:* limited communication medium
- **Web-Based Discussion Forums**
  - *Advantages:* same as electronic mailing lists except requires slightly faster Internet connection
  - *Disadvantages:* cultural resistance
- **Lotus Notes**
  - *Advantages:* comprehensive collaborative solution employing state-of-the-art technologies for communication, document management, and work flow
  - *Disadvantages:* expensive to deploy when compared with other collaboration technologies

The World Bank Case: KM Architecture at the World Bank

- The World Bank spent a fortune on classifying knowledge.
- The bank employs XML–enabled Oracle data engine to drive a document management system linked to Lotus Notes groupware.
- Codification technologies needs to be evaluated in terms of a return on investment



Source: Awad and Ghaziri: Knowledge Management, 2007)

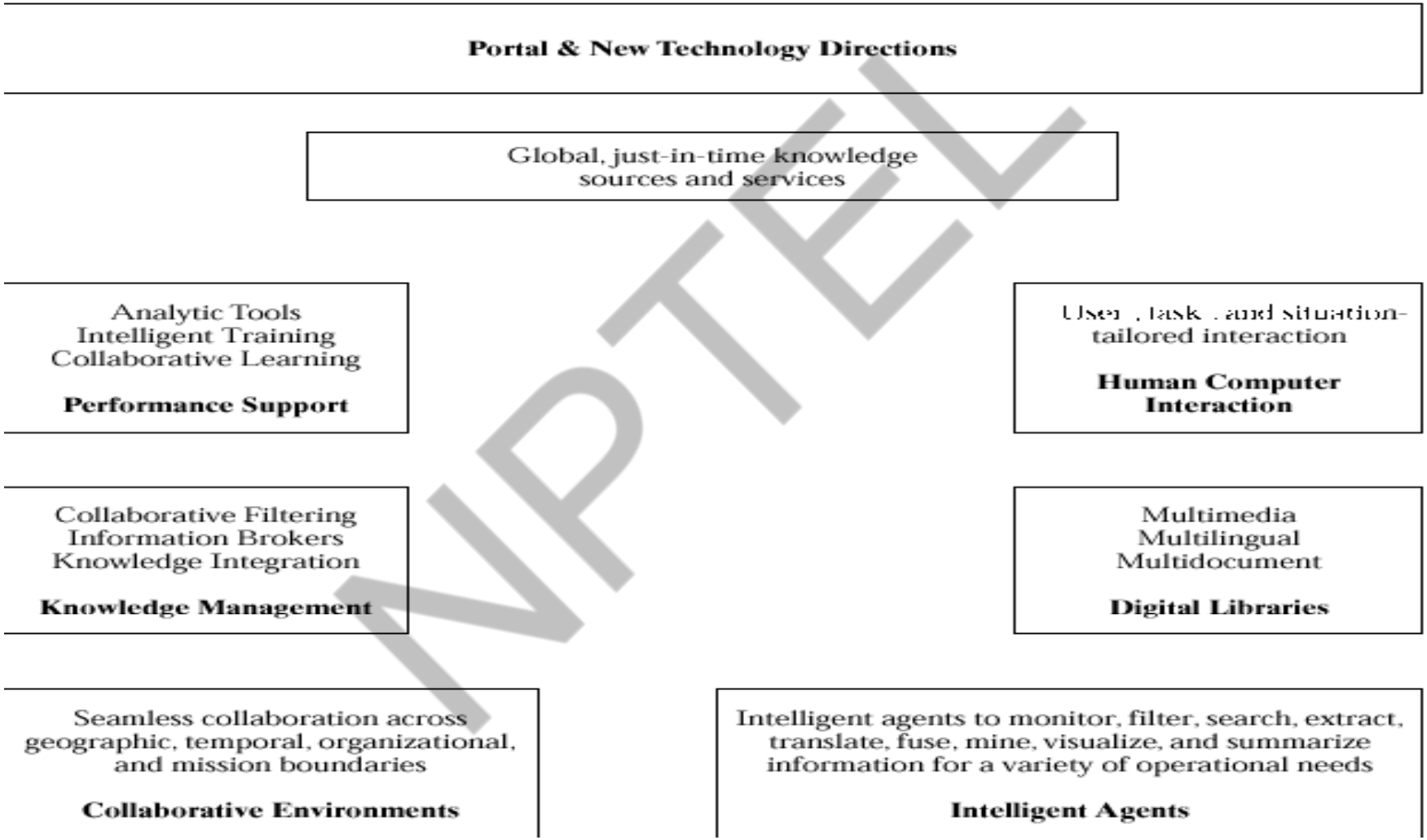
Intelligent Agents

- Intelligent agents are tools that can be applied in numerous ways in the context of EKPs.
- Intelligent agents are still in their infancy.
- Agents are software entities that are able to execute a wide range of functional tasks

Intelligent Agents Services

- Customized customer assistance with online services
- Customer profiling based on business experiences
- Integrating profiles of customers into a group of marketing activities
- Predicting customer requirements
- Negotiating prices and payment schedules
- Executing financial transactions on the customer's behalf

New Trends in Portals Technologies



Source: Awad and Ghaziri: Knowledge Management, 2007)

Critical Issues for Knowledge-Sharing Programs

- Responsiveness to user need
- Content structure. In large systems
- Content quality requirements
- Integration with existing systems
- Scalability
- Hardware–software compatibility.
- Synchronization of technology with the capabilities of users.

Portal Vendors

Vendor	*KM Portal Product*	*Feature Summary*	*Best Uses*
Lotus/IBM	Lotus Raven 1.0 (in beta)	• Intelligent taxonomy • QuickPlace collaboration tool • Assigns value to data based on how often it is used • Portal replication • Facilitates content management	• Self-creating and refining taxonomies • Personnel resources linked to data sources • Advanced collaboration • Easy portal repurposing • Rapid application development with associated KM packages
Open Text	MyLivelink Portal 1.0 with Livelink 8.5.1 KM software	• Integrated work flow • Quick integration of features • Quick portal deployment	• Integrated KM • Document management and work flow • Custom collaboration spaces (personal, project, or enterprise)

Source: Awad and Ghaziri: Knowledge Management, 2007)

Portal Vendors

Vendor	KM Portal	Feature Summary	Best Uses
Plumtree	Plumtree Corporate Portal 4.0	• Automatic population • E-mail, voice, and wireless notification • Integration with LDAP directories • E-room tools	• Easy and extensive content and application integration • Scalability • Advanced security • Trainable taxonomies • Various data access • Customization and extensibility
Woolamai	WebMeta Engine 1.0	• Quick integration • Flexible portal interface • Knowledge taxonomy adapts to data views • Data-mining functionality • Web site statistics	• Usability • Tracking site statistics • Content streaming to wireless devices

Source: Awad and Ghaziri: Knowledge Management, 2007)