Ch7: Text Analytics, Text Mining, and Sentiment Analysis

text mining: A semi-automated process of extracting knowledge from unstructured data sources also, text data mining or knowledge discovery in textual databases.

Structured versus unstructured data

- Structured data: in databases
- Unstructured data: Word documents, PDF files, text excerpts, XML files, and so on

The similar and differ between Data Mining versus Text Mining

- Both seek for novel and useful patterns
- Both are semi-automated processes
- Difference is the nature of the data:
- Structured versus unstructured data
- Structured data: in databases
- Unstructured data: Word documents, PDF files, text excerpts, XML files, and so on
- Text mining – first, impose structure to the data, then mine the structured data.

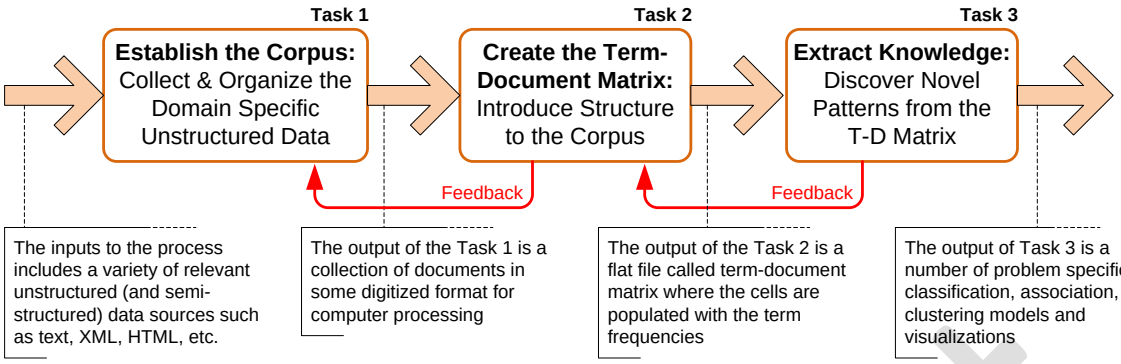
This important…. Natural Language Processing (NLP):

- **Structuring a collection of text**
  - **Old approach: bag-of-words**
  - **New approach: natural language processing**
- **NLP is : a very important concept in text mining**
  - **a subfield of artificial intelligence and computational linguistics**
  - **the studies of "understanding" the natural human language**
- **Syntax versus semantics-based text mining**

Dream of AI community to have algorithms that are capable of automatically reading and obtaining knowledge from text. T

Sentiment Analysis: A technique used to detect favorable and unfavorable opinions toward specific products and services. Fill or t

Soooooo important>>> Text Mining Process? The three-step text mining process



SVD is similar to principle component analysis

Sentiment :belief, view, opinion, conviction

Sentiment analysis : opinion mining, subjectivity analysis, and appraisal extraction

The goal is to answer the question:

"What do people feel about a certain topic?" multi or fill (Sentiment analysis)

Soo important in this chapter…Sentiment Analysis Process?

- Step 1 – Sentiment Detection: Comes right after the retrieval and preparation of the text documents . It is also called detection of objectivity
- Step 2 – N-P Polarity Classification: Given an opinionated piece of text, the goal is to classify the opinion as falling under one of two opposing sentiment polarities
- Step 3 – Target Identification: The goal of this step is to accurately identify the target of the expressed sentiment (e.g., a person, a product, an event, etc.)
- Step 4 – Collection and Aggregation: Once the sentiments of all text data points in the document are identified and calculated, they are to be aggregated

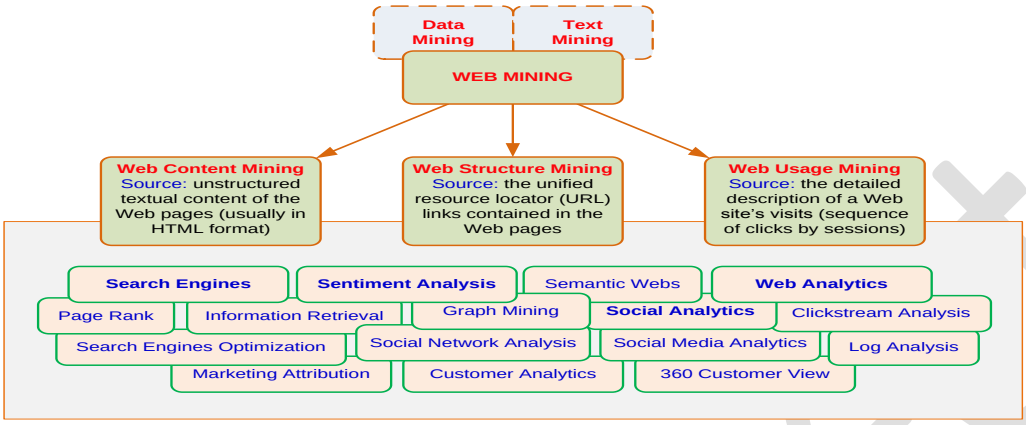
Speech analytics – analysis of voice: Content versus other Voice Features

Two Approaches:

- The Acoustic Approach: Intensity, Pitch, Jitter, Shimmer, etc.
- The Linguistic Approach: Lexical, Disfluencies, Higher semantics.

Ch8: Web Analytics, Web Mining, and Social Analytics:

- Web mining (or Web data mining) is the process of discovering intrinsic relationships from Web data (textual, linkage, or usage)



Search engine is a software program that searches for documents (Internet sites or files) based on the keywords (individual words, multi-word terms, or a complete sentence) that users have provided that have to do with the subject of their inquiry. fill

Search engine are the **workhorses** of the Internet. T

The way of Anatomy of a Search Engine?

1- **Development Cycle**

Steps:

- Step 1 – Pre-Processing the Documents
- Collecting, organizing, and storing
- Step 2 – Parsing the Documents
- Step 3 – Creating the Term-by-Document Matrix

2- **Response Cycle**

- Query Analyzer
- Document Matcher/Ranker

Search Engine Optimization (SEO): It is the intentional activity of affecting the visibility of an e-commerce site or a Web site in a search engine's natural (unpaid or organic) search results.

SEO Part of an Internet marketing strategy. T

What are the Methods for Search Engine Optimization(SEO):

- **Search engine recommended techniques** (**White-Hat SEO**): **Producing results based on good site design, accurate content** (for users, not engines)
- **Search engine disapproved techniques** (**Black-Hat SEO**): **Spamdexing**? (search spam, search engine spam, or search engine poisoning), **Deception** (what is shown is different to human and machine/spider).

Web Usage Mining (Web Analytics):

- Extraction of information from data generated through Web page visits and transactions
- Clickstream data, clickstream analysis.

Web Analytics Metrics: Provides near-real-time data to deliver invaluable information to …

Improve site usability, Manage marketing efforts, Better document ROI.

Web analytics metric categories: is 4

Web Site Usability	Traffic Source	Visitor Profiles	Conversion Statistics
1. Page views 2. Time on site 3. Downloads 4. Click map 5. Click paths	1. Referral Web sites 2. Search engines 3. Direct 4. Offline campaigns 5. Online campaigns	1. Keywords 2. Content groupings 3. Geography 4. Time of day 5. Landing page	1. New visitors 2. Returning visitors 3. Leads 4. Sales/conversions 5. Abandonment rates

Social Network - social structure composed of individuals linked to each other

Analysis of social dynamics in Interdisciplinary field are:

- Social psychology
- Sociology
- Statistics
- Graph theory

Social Networks help study relationships between individuals, groups, organizations, societies. Fill or t

Social Networks help study on **Self organizing, Emergent, Complex**. Fill

Typical social network types: Communication networks, community networks, criminal networks, innovation networks.

Important….Social Network Analysis Metrics:

Connections	Segmentation	Distribution
Homophily Multiplexity Network closure Propinquity	Cliques and social circles Clustering coefficient Cohesion	Bridge Centrality Density Structural holes Tie strength

Social Media : Enabling technologies of social interactions among people

Social Media : Relies on enabling technologies of **Web 2.0.** T

Takes on many different forms ex: **Internet forums, Web logs, social blogs, microblogging, wikis, social networks, podcasts, pictures, video, and product reviews**

Different **types of social media**: Based on **media research** and **social process.**

What are the Different Types of Social Media?

1. Collaborative projects (e.g., Wikipedia)
2. Blogs and microblogs (e.g., Twitter)
3. Content communities (e.g., YouTube)
4. Social networking sites (e.g., Facebook)
5. Virtual game worlds (e.g., World of Warcraft), and
6. Virtual social worlds (e.g., Second Life)

Web-based social media are different from traditional/industrial media, such as newspapers, television, and film. T

Social Media Analytics: It is the systematic and scientific ways to consume the vast amount of content created by Web-based social media outlets, tools, and techniques for the betterment of an organization's competitiveness.

Measuring the Social Media Impact:

- Descriptive analytics – simple counts/statistics
- Social network analysis
- Advanced analytics – predictive analytics, text mining

Ch9:Model-Based Decision Making: Optimization and Multi-Criteria Systems

Model-Based Decision Making: DSS modeling (optimization & simulation) contribute to organizational success ex: Pillowtex,Fiat.

Major Modeling Issues:

- Problem identification and environmental analysis (information collection)
- Variable identification
- Forecasting/predicting
- Multiple models: An MSS can include several models, each of which represents a different part of the decision-making problem
- Model management – DBMS vs. MBDM

Categories of Models: should to know each one of them what is the process and techniques

TABLE 9.1 Categories of Models

Category	Process and Objective	Representative Techniques
Optimization of problems with few alternatives	Find the best solution from a small number of alternatives	Decision tables, decision trees, analytic hierarchy process
Optimization via algorithm	Find the best solution from a large number of alternatives, using a step-by-step improvement process	Linear and other mathematical programming models, network models
Optimization via an analytic formula	Find the best solution in one step, using a formula	Some inventory models
Simulation	Find a good enough solution or the best among the alternatives checked, using experimentation	Several types of simulation
Heuristics	Find a good enough solution, using rules	Heuristic programming, expert systems
Predictive models	Predict the future for a given scenario	Forecasting models, Markov analysis
Other models	Solve a what-if case, using a formula	Financial modeling, waiting lines

Model Categories:

Static Analysis	Dynamic Analysis
■ Single snapshot of the situation ■ Single interval ■ Steady state	■ Dynamic models ■ Evaluate scenarios that change over time ■ Time dependent ■ Represents trends and patterns over time ■ More realistic: Extends static models

The Structure of a Mathematical Model: should to know the equation of **simple present value** :

- Example — Profit - $P = R - C$

where P= profit, R= revenue, and C= cost

- Example - Simple Present-Value -

$$P = \frac{F}{(1+i)^n} = \frac{100{,}000}{(1+0.1)^5} = 62{,}092$$

where P= present value, F= future cash-flow, i= interest-rate, and n = number of period (years)

Modeling and Decision Making - Under Certainty, Uncertainty, and Risk. Fill

Certainty	Uncertainty	Risk
▪ Assume complete knowledge ▪ All potential outcomes are known ▪ May yield optimal solution	▪ Several outcomes for each decision. ▪ Probability of each outcome is unknown. ▪ Knowledge would lead to less uncertainty.	Probability of each of several outcomes occurring. Level of uncertainty => Risk (expected value).

Decision Modeling with **Spreadsheets**: Fill or multi

- Most popular end-user modeling tool
- Flexible and easy to use
- Powerful functions (add-in functions)
- Programmability (via macros)
- What-if analysis and goal seeking
- Simple database management
- Seamless integration of model and data
- Incorporates both static and dynamic models

$$F = P(1+i)^n$$

$$A = P\left[\frac{i(1+i)^n}{(1+i)^n - 1}\right]$$

you should to know the formula for this or the name of this formula :

- Static model example: (Simple loan calculation of monthly payments)
- Dynamic model example: (Simple loan calculation of monthly payments and effects of prepayment)

- Mathematical Programming : A family of tools designed to help solve managerial problems in which the decision maker must allocate scarce resources among competing activities to optimize a measurable goal

Optimal solution: The best possible solution to a modeled problem

- Linear programming (LP): A mathematical model for the optimal solution of resource allocation problems. All the relationships are linear.

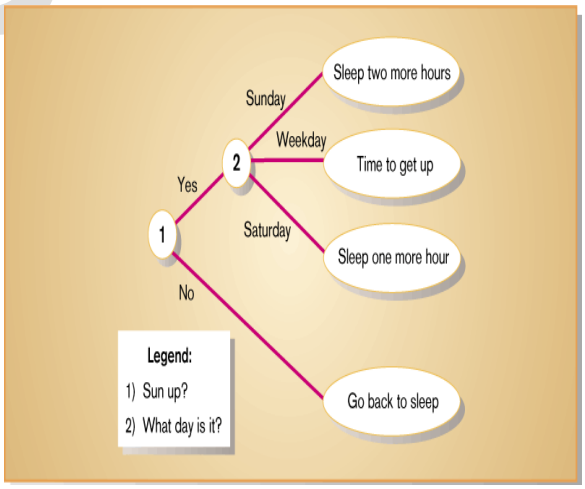
Common Optimization Models:

- Product-mix problems.
- Transportation.
- Assignment
- Investment.
- Network optimization models for planning and scheduling.
- Replacement.

Multiple Goals	Sensitivity Analysis	What-If Analysis	Goal Seeking
Vast majority of **managerial problems has multiple goals** (objectives) to achieve Attaining simultaneous goals. Methods of handling multiple goals: ▪ Utility theory ▪ Goal programming ▪ Expression of goals as constraints, using LP ▪ A points system	It is the process of assessing the impact of change in inputs on outputs. Can be **automatic or trial and error.**	**Assesses solutions based on changes** in variables or assumptions (scenario analysis)	**Backwards approach, starts with the goal** and determines values of inputs needed.

- Having **more than one criterion** makes **decision-making process complicated**. T

Decision Analysis with **Decision Tables** and **Decision Trees**:

Decision Tables	Decision Trees
a tabular representation of the decision situation (alternatives). Example : Investment To summary the ex the table content : ▪ Payoff decision variables (alternatives) ▪ Uncontrollable variables (states of economy) : Solid growth, Stagnation and Inflation ▪ Result variables (projected yield) **TABLE 9.3** Investment Problem Decision Table Model State of Nature (Uncontrollable Variables) Alternative / Solid Growth (%) / Stagnation (%) / Inflation (%) Bonds / 12.0 / 6.0 / 3.0 Stocks / 15.0 / 3.0 / -2.0 CDs / 6.5 / 6.5 / 6.5 **Should to know when can be solid grow or stagnation or inflation:** ▪ If **solid growth** in the economy, bonds yield 12%; stocks 15%; time deposits 6.5% ▪ If **stagnation**, bonds yield 6%; stocks 3%; time deposits 6.5% ▪ If **inflation**, bonds yield 3%; stocks lose 2%; time deposits yield 6.5%	▪ **Graphical representation of relationships** ▪ Multiple criteria approach ▪ Demonstrates complex relationships ▪ Cumbersome, if many alternatives exists Ex: 

Multi-Criteria Decision Making(MCDM): A very popular **technique** for MCDM are: **Popular Tools** and **Web-based Tools.**

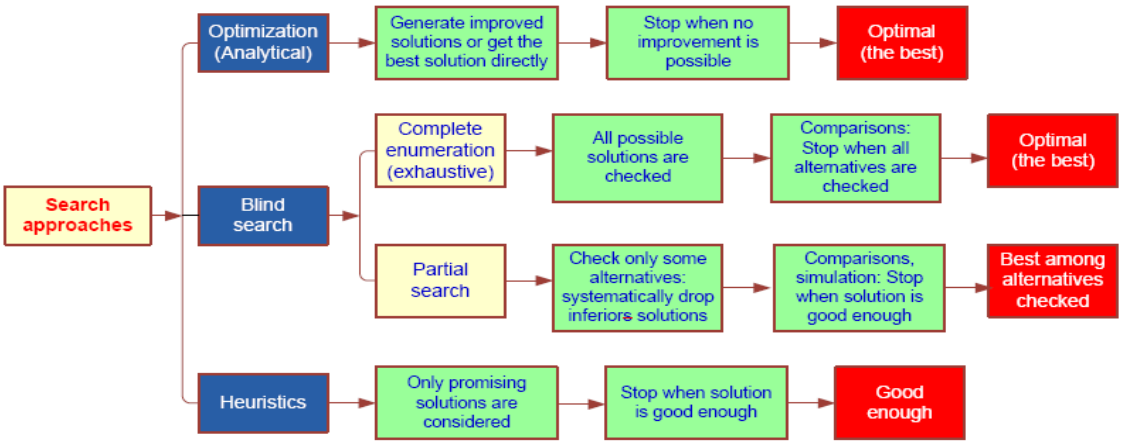
Ch10: Modeling and Analysis: Heuristic Search Methods and Simulation

Search: **choice** phase of decision making

Search is the process of identifying the best possible solution / course of action

Search techniques include

- analytical techniques,
- algorithms,
- blind searching, and
- heuristic searching



Problem-Solving Search Methods- Algorithmic/Heuristic :Gets satisfactory solutions more quickly and less expensively. T or fill

Total number of unique routes (TNUR): TNUR = (1/2) (Number of Cities – 1)! $\frac{1}{2}(n-1)!$

Limitations of Heuristics! Cannot guarantee an optimal solution. T

Modern Heuristic Methods:

- Tabu search: Intelligent search algorithm
- Genetic algorithms: Survival of the fittest
- Simulated annealing: Analogy to Thermodynamics
- Ant colony and other Meta-heuristics

Genetic Algorithms

Genetic Algorithms: **An efficient, domain-independent search heuristic for a broad spectrum of problem domains. T or fill**

Genetic Algorithms **Main theme: Survival of the fittest. T**

Evolutionary Algorithm: Current generation ⇒ Selection ⇒Reproduction: Crossover, Mutation - ⇒ Next generation.

GA Structure:

- Each **candidate** solution is called a **chromosome**
- A chromosome is a string of **genes**

Simulation: is the "appearance" of reality

- It is often used to conduct what-if analysis on the model of the actual system. T
- Simulation should be used only when a numerical optimization is possible. F (not possible)

Major Characteristics of Simulation:

- Imitates reality and captures its richness both in shape and behavior
- Technique for conducting experiments
- Descriptive, not normative tool

Advantages of Simulation	Disadvantages of Simulation
▪ The theory is fairly straightforward ▪ Great deal of time compression ▪ Experiment with different alternatives ▪ The model reflects manager's perspective ▪ Can handle wide variety of problem types ▪ Can include the real complexities of problems ▪ Produces important performance measures ▪ Often it is the only DSS modeling tool for non-structured problems	▪ Cannot guarantee an optimal solution ▪ Slow and costly construction process ▪ Cannot transfer solutions and inferences to solve other problems (problem specific) ▪ So easy to explain/sell to managers, may lead to overlooking analytical solutions ▪ Software may require special skills

The steps of Simulation Methodology are:

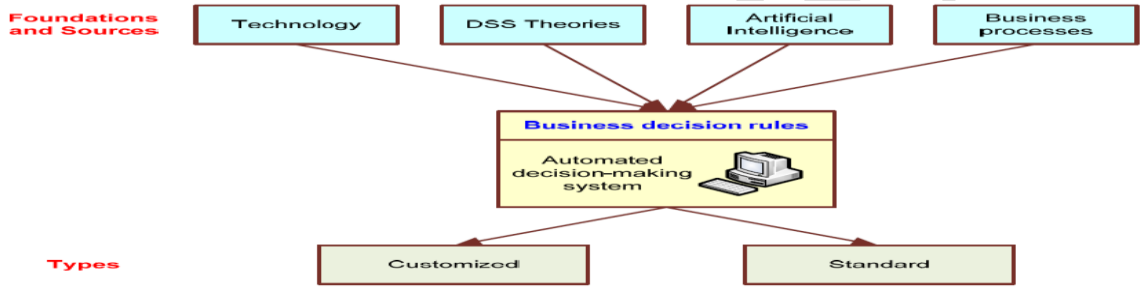
1. Define problem
2. Construct the model
3. Test and validate model
4. Design experiments
5. Conduct experiments
6. . Evaluate results
7. Implement solution

- **Visual interactive modeling (VIM),** also called **Visual Interactive Simulation** or **Visual interactive problem solving. Fill**
- **Visual interactive modeling (VIM) Uses computer graphics to present the impact of different management decisions. T**
- **Agent - an autonomous computer program that observes and acts on an environment and directs its activity toward achieving specific goals.**
- **Agent-based modeling (ABM) is a simulation modeling technique to support complex decision systems where a system is modeled as a set of autonomous decision-making units called agents**
- **Agent-based modeling (ABM) : A bottom-up approach to simulation modeling. T**

Ch11: Automated Decision Systems and Expert Systems:

Automated Decision Systems: Often a rule-based system that provides a solution in a functional area

Automated Decision-Making **Framework**:



Artificial intelligence (AI):

- A subfield of computer science, concerned with symbolic reasoning and problem solving.
- Behavior by a machine that, if performed by a human being, would be considered intelligent
- "…study of how to make computers do things at which, at the moment, people are better
- Theory of how the human mind works

AI Objectives:

- Make machines smarter (primary goal)
- Understand what intelligence is
- Make machines more intelligent & useful

A computer can be considered to be smart **only when** a human interviewer, "conversing" with both an unseen human being and an unseen computer, cannot determine which is which. T

AI provides the scientific foundation for many commercial technologies. T

AI is Often Transparent in Many Commercial Products such as : **Anti-lock Braking Systems (ABS), Automatic Transmissions. T or fill**

Expert Systems (ES): Is a computer program that attempts to **imitate** expert's reasoning processes and knowledge in solving specific problems

Most Popular Applied AI Technology is Enhance Productivity and Augment Work Forces. T

Expert systems do not replace experts, but Make their knowledge and experience more widely available, and thus Permit non-experts to work better. T

Important Concepts in ES:

- **Expert** : A human being who has developed a high level of proficiency in making judgments in a specific domain
- **Expertise**: The set of capabilities that underlines the performance of human experts.

Transferring Expertise: From expert to computer to non experts via **acquisition, representation, inferencing, transfer. Short or fill or t**

Conventional vs. Expert Systems may coming as T or F:

TABLE 11.1 Comparison of Conventional Systems and Expert Systems

Conventional Systems	Expert Systems
Information and its processing are usually combined in one sequential program.	The knowledge base is clearly separated from the processing (inference) mechanism (i.e.,knowledge rules are separated from the control).
The program does not make mistakes (programmers or users do).	The program may make mistakes.
Conventional systems do not (usually) explain why input data are needed or how conclusions are drawn.	Explanation is a part of most ES.
Conventional systems require all input data. They may not function properly with missing data unless planned for.	ES do not require all initial facts. ES can typically arrive at reasonable conclusions with missing facts.
Changes in the program are tedious (except in DSS).	Changes in the rules are easy to make.
The system operates only when it is completed.	The system can operate with only a few rules (as the first prototype).
Execution is done on a step-by-step (algorithmic) basis.	Execution is done by using heuristics and logic.
Large databases can be effectively manipulated.	Large knowledge bases can be effectively manipulated.
Conventional systems represent and use data.	ES represent and use knowledge.
Efficiency is usually a major goal.	
Effectiveness is important only for DSS.	Effectiveness is the major goal.
Conventional systems easily deal with quantitative data.	ES easily deal with qualitative data.
Conventional systems use numeric data representations.	ES use symbolic and numeric knowledge representations.
Conventional systems capture, magnify, and distribute access to numeric data or information.	ES capture, magnify, and distribute access to judgment and knowledge.

Applications of Expert Systems: Classical Applications, reported Applications.

Structure of Expert Systems:

- Development Environment
- Consultation Environment
- Major Components
    - Knowledge acquisition subsystem:
        - Knowledge Engineer
    - Knowledge Base
    - Inference Engine
    - User Interface
    - Blackboard (workplace)
    - Explanation subsystem (justifier)
    - Knowledge-refining system

The primary goal of **KE** is to help experts articulate *how they do what they do,* and to document this knowledge in a reusable form. T

Knowledge Engineering (KE): A set of intensive activities encompassing the acquisition of knowledge from human experts (and other information sources) and converting this knowledge into a repository (commonly called a knowledge base)

The Knowledge Engineering Process: Mcq

- Knowledge acquisition.
- Knowledge Representation
- Knowledge Validation.
- Inferencing (Reasoning)
- Explanation Justification.

is difficult to acquire knowledge from experts for all the following reasons EXCEPT:

- experts often change their behavior when observed.
- testing and refining of knowledge is complex and difficult
- experts may not be able to put into words how they conduct their work
- many business areas have no identifiable experts.

Validation versus Verification:

- **Validation** is the part of evaluation that deals with the performance of the system
- **Verification** is building the system right or substantiating that the system is correctly implemented to its specifications

Expert knowledge must **be represented in a computer-understandable** format and organized properly in the knowledge base. T

The most common/popular way to **represent human knowledge**: Production rules. Fill

Forms of Production Rules: لازم تعرفو كيف كل وحده تم استنتاجها لانو يمكن تكون صح او خطا .. مع انو اعتقد منو مهم

- IF premise, THEN conclusion
- Conclusion, IF premise
- Inclusion of ELSE

Knowledge rules (declarative rules), state all the facts and relationships about a problem. fill

Knowledge rules are stored in the knowledge base. fill

Inference rules (procedural rules), advise on how to solve a problem, given that certain facts are known. Fill

Know this is important information to know it in this chapter and I think the long Q coming in these as following :

Inference is the process of chaining multiple rules together based on available data

Forward chaining: should to know important

- A data-driven search in a rule-based system.
- If the premise clauses match the situation, then the process attempts to assert the conclusion.

Backward chaining: important

- A goal-driven search in a rule-based system.
- It begins with the action clause of a rule and works backward through a chain of rules in an attempt to find a verifiable set of condition clauses.

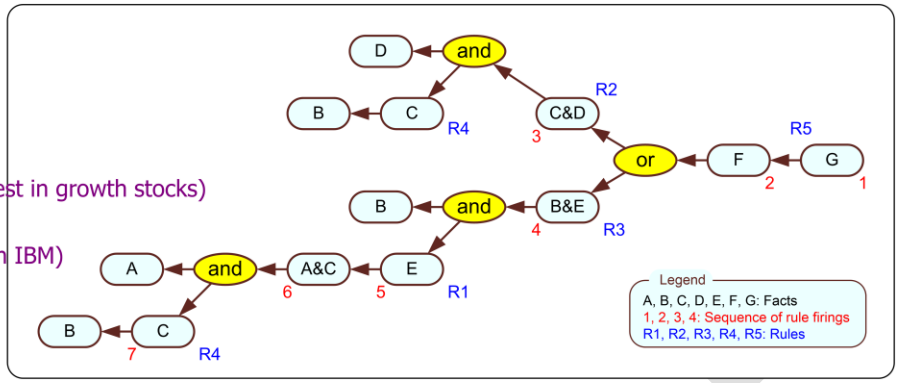
Firing a rule: When all of the rule's hypotheses (the "if parts") are satisfied, a rule said to be FIRED. T

Inferencing – Backward Chaining (أعتقد واحد منهم مع الرسمه حايجي بالونق انسر و الله اعلم)

Goal-driven: Start from a potential conclusion (hypothesis), then seek evidence that supports (or contradicts with) it

Knowledge Base

Rule 1: A & C -> E
Rule 2: D & C -> F
Rule 3: B & E -> F (invest in growth stocks)
Rule 4: B -> C
Rule 5: F -> G (invest in IBM)

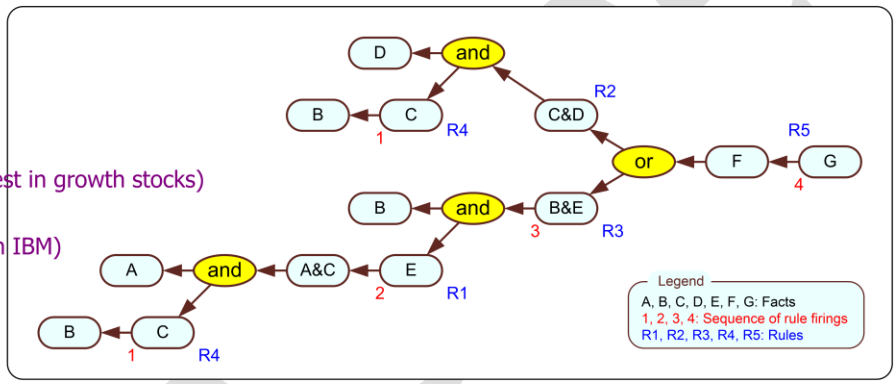


Inferencing – Forward Chaining

Data-driven: Start from available information as it becomes available, then try to draw conclusions

Knowledge Base

Rule 1: A & C -> E
Rule 2: D & C -> F
Rule 3: B & E -> F (invest in growth stocks)
Rule 4: B -> C
Rule 5: F -> G (invest in IBM)



- **If the expert first collect data then infer from it => Forward Chaining**
- **If the expert starts with a hypothetical solution and then attempts to find facts to prove it => Backward Chaining**

Certainty Factors (CF) express belief in an event based on evidence (or the expert's assessment).T or fill

- **CFs are NOT probabilities. T**
- **CFs need not sum to 100. T**
- **Combining Several Certainty Factors in TWO Rule where parts are combined using AND and OR logical operators. F (One Rule)**

الان دي المسالة مره مهمه جدا عن طريقة جمع اكثر من قاعدة حاذكر المثال الي بالسلايد و الحل تبعو و تكمله للمثال في حالة لو جانا في الاختبار :

Combining two or more rules?

R1: IF the inflation rate is less than 5 percent,

THEN stock market prices go up (CF = 0.7)

R2: IF unemployment level is less than 7 percent,

THEN stock market prices go up (CF = 0.6)

Inflation rate = 4 percent and the unemployment level = 6.5 percent

Combined Effect(دي القاعدة الي حانعوض فيها عشان نحل ع حسب المعطيات فوق الرول ون و تو)

- CF(R1,R2) = CF(R1) + CF(R2)[1 - CF(R1)]; or
- CF(R1,R2) = CF(R1) + CF(R2) - CF(R1) × CF(R2)

The solution is:

Given CF(R1) = 0.7 AND CF(R2) = 0.6, then:

CF(R1,R2) = 0.7 + 0.6(1 - 0.7) = 0.7 + 0.6(0.3) = 0.88

- Expert System tells us that there is an 88 percent chance that stock prices will increase
- For a third rule to be added

CF(R1,R2,R3) = CF(R1,R2) + CF(R3) [1 - CF(R1,R2)]

R3: IF bond price increases THEN stock prices go up (CF = 0.85)

Assuming all rules are true in their IF part, the chance that stock prices will go up is

CF(R1,R2,R3) = 0.88 + 0.85 (1 - 0.88) = 0.982

>>>

This is another example from m Rules

R1: IF blood test result is yes

THEN the disease is malaria (CF 0.8)

R2: IF living in malaria zone

THEN the disease is malaria (CF 0.5)

R3: IF bit by a flying bug

THEN the disease is malaria (CF 0.3)

Questions:

- What is the CF for having malaria (as its calculated by ES), if

Q1:The first two rules are considered to be true ?

Q2:All three rules are considered to be true?

The solution is:

- **Answer 1:**

1. $CF(R1, R2) = CF(R1) + CF(R2) * (1 – CF(R1)$

$= 0.8 + 0.5 * (1 - 0.8) = 0.8 – 0.1 = 0.9$

2. $CF(R1, R2, R3) = CF(R1, R2) + CF(R3) * (1 - CF(R1, R2))$

$= 0.9 + 0.3 * (1 - 0.9) = 0.9 – 0.03 = 0.93$

- **Answer 2:**

1. $CF(R1, R2) = CF(R1) + CF(R2) – (CF(R1) * CF(R2))$

$= 0.8 + 0.5 – (0.8 * 0.5) = 1.3 – 0.4 = 0.9$

2. $CF(R1, R2, R3) = CF(R1, R2) + CF(R3) – (CF(R1, R2) * CF(R3))$

$= 0.9 + 0.3 – (0.9 * 0.3) = 1.2 – 0.27 = 0.93$

>>.

Explanation as a Metaknowledge:

- Explanation : Human experts justify and explain their actions
- Explanation: an attempt by an ES to clarify reasoning, recommendations, other actions (asking a question)
- **Two Basic Explanations**: **Why** Explanations - **How** Explanations.
- **Explanation** is essential in ES and Used for training and evaluation. T

Name and describe **three problem** areas suitable for expert systems.

- Interpretation: Inferring situation descriptions from observations.
- Prediction: Inferring likely consequences of given situations.
- Diagnosis: Inferring system malfunctions from observations.
- Design: Configuring objects under constraints.
- Planning: Developing plans to achieve goals.
- Monitoring: Comparing observations to plans and flagging exceptions.
- Debugging: Prescribing remedies for malfunctions.
- Repair: Executing a plan to administer a prescribed remedy.
- Instruction: Diagnosing, debugging, and correcting student performance.
- Control: Interpreting, predicting, repairing, and monitoring system behaviors

The development of expert systems is often described as a tedious process. What activities does it typically include?

Answer:

- Identifying proper experts
- Acquiring knowledge
- Selecting the building tools
- Coding the system
- Evaluating the system

Ch12: Knowledge Management and Collaborative Systems

- **Knowledge management** : The active management of the expertise in an organization. It involves collecting, categorizing, and disseminating knowledge
- Intellectual capital: The invaluable knowledge of an organization's employees
- **Knowledge** is information that is contextual, relevant, and actionable understanding, awareness, or familiarity acquired through education or experience.

In **a knowledge management system**, "knowledge **is information in action**" . T

Knowledge-based economy: The economic shift **from natural resources to intellectual assets. T**

Define knowledge management (KM), and briefly explain the process through which it is implemented within an organization.

Answer:

- Knowledge management is the systematic and active management of ideas, information, and knowledge residing in an organization's employees.
- Knowledge management is a process that helps organizations identify, select, organize, disseminate, and transfer important information and expertise that are part of the organization's memory and that typically reside within the organization in an unstructured manner.

What are the four characteristics of knowledge that make it unlike other organizational assets?

Answer:

- Extraordinary leverage and increasing returns. Knowledge is not subject to diminishing returns. When it is used, it is not decreased (or depleted); rather, it is increased (or improved). Its consumers can add to it, thus increasing its value.
- Fragmentation, leakage, and the need to refresh. As knowledge grows, it branches and fragments. Knowledge is dynamic; it is information in action. Thus, an organization must continually refresh its knowledge base to maintain it as a source of competitive advantage.

- Uncertain value. It is difficult to estimate the impact of an investment in knowledge. There are too many intangible aspects that cannot be easily quantified.
- Value of sharing. It is difficult to estimate the value of sharing one's knowledge or even who will benefit most from it.

Tacit (embedded) knowledge: Knowledge that is usually in the domain of subjective, cognitive, and experiential learning.

Tacit knowledge is described as formal, structured knowledge that is well documented. t

Explicit (leaky) knowledge : Knowledge that deals with objective, rational, and technical material.

The term "leaky knowledge" MOST accurately refers to :

A. tacit knowledge.
B. individual knowledge.
C. explicit knowledge.
D. social knowledge.

- **Learning organization** : An organization capable of learning from its past experience, implying the existence of an organizational memory and a means to save, represent, and share it through its personnel
- **Organizational memory** : Repository of what the organization knows
- **Organizational culture** : The aggregate attitudes in an organization concerning a certain issue.

What are and the differ between them (Approaches to Knowledge Management):

Process approach	Practice approach	Hybrid approaches
■ Process approach to knowledge management attempts to codify organizational knowledge through formalized controls, processes and technologies ■ Focuses on explicit knowledge and IT	■ Practice approach focuses on building the social environments or communities of practice necessary to facilitate the sharing of tacit understanding Focuses on tacit knowledge and socialization	■ Hybrid approaches to knowledge management

Knowledge repository is the actual storage location of knowledge in a knowledge management system.

A functioning **knowledge management system (KMS)** follows six steps in a cycle. The reason for the cycle is that knowledge is dynamically refined over time. What are the six steps in the KMS cycle?

Answer:

1. Create knowledge. Knowledge is created as people determine new ways of doing things or develop know-how. Sometimes external knowledge is brought in. Some of these new ways may become best practices.
2. Capture knowledge. New knowledge must be identified as valuable and be represented in a reasonable way.
3. Refine knowledge. New knowledge must be placed in context so that it is actionable. This is where human insights (i.e., tacit qualities) must be captured along with explicit facts.
4. Store knowledge. Useful knowledge must be stored in a reasonable format in a knowledge repository so that others in the organization can access it.
5. Manage knowledge. Like a library, a repository must be kept current. It must be reviewed to verify that it is relevant and accurate.
6. Disseminate knowledge. Knowledge must be made available in a useful format to anyone in the organization who needs it, anywhere and anytime.

Groupwork → the work done by two or more people together

Group Decision Support Systems(GDSS): the major characteristics of a GDSS:

- It is an interactive computer-based system that facilitates the solution of semi-structured or unstructured problems by a group of decision makers
- Goal – support group decision making
- A specially designed IS to enhance collaborative decision processes
- It encourages generation of ideas, freedom of expression, and resolution of conflicts

GDSS – Pros and Cons:

Gains:	Loses:
Parallelism Anonymity Triggering Synergy Structure Record keeping	Free-riding Flaming

Why Very Few Organizations Use Decision Rooms? May coming as Mcq

- High Cost
- Need for a Trained Facilitator
- Requires Specific Software Support for Different Cooperative Tasks
- Infrequent Use
- Different Place / Different Time Needs
- May Need More Than One

Ch13: Big Data Analytics

"Big Data" = massive volumes of data

The Vs that define Big Data:

- Volume
- Variety
- Velocity
- Veracity
- Variability
- Value

A High-level Conceptual Architecture for Big Data Solutions (by **Aster Data / Teradata**). Fill

You can't process the amount of data that you want to because of the limitations of your current platform. T

You need to (or want to) integrate data as quickly as possible to be current on your analysis. T

List and describe four of the most critical success factors for Big Data analytics.

- **A clear business need** (alignment with the vision and the strategy): Business investments ought to be made for the good of the business, not for the sake of mere technology advancements.
- **Strong, committed sponsorship** (executive champion):It is a well-known fact that if you don't have strong, committed executive sponsorship, it is difficult (if not impossible) to succeed.
- **Alignment between the business and IT strategy**: It is essential to make sure that the analytics work is always supporting the business strategy.
- **A strong data infrastructure**: Data warehouses have provided the data infrastructure for analytics

List the five Challenges of Big Data Analytics?

- Data volume
  - The ability to capture, store, and process the huge volume of data in a timely manner
- Data integration
  - The ability to combine data quickly/cost effectively
- Processing capabilities
  - The ability to process the data quickly, as it is captured (i.e., stream analytics)
- Data governance (… security, privacy, access)
- Skill availability (… data scientist)
- Solution cost (ROI)

Define MapReduce: MapReduce is a programming model and an associated implementation for processing and generating large data sets. Programs written in this functional style are automatically parallelized and executed on a large cluster of commodity machines. This allows programmers without any experience with parallel and distributed systems to easily utilize the resources of a large distributed system."

Define Hadoop is an open source framework for storing and analyzing massive amounts of distributed, unstructured data

In the world of Big Data, __MapReduce _____ aids organizations in processing and analyzing large volumes of multi-structured data. Examples include indexing and search, graph analysis, etc. fill

A job _tracker _____ is a node in a Hadoop cluster that initiates and coordinates MapReduce jobs, or the processing of the data. fill

Big Data vendor landscape is developing very rapidly.T

This may coming as t or F:”

- Hadoop consists of multiple products. T
- Hadoop is open source but available from vendors, too. T
- Hadoop is an ecosystem, and a single product. F and not a single product
- HDFS is a file system, and a DBMS. F is not DBMS
- Hadoop and MapReduce are related but the same. F not the same

- MapReduce provides control for analytics, not analytics. T
- Hadoop is about data diversity, not just data volume.T
- Hadoop complements a DW; it's rarely a replacement.T
- Hadoop enables many types of analytics, not just Web analytics.T "

What are the differ between Stream Analytics Versus Perpetual Analytics?

Stream Analytics	Perpetual Analytics
involves applying transaction- level logic to real-time observations. When transactional volumes are high and the time-to-decision is too short, favoring nonpersistence and small window sizes, this translates into using streaming analytics.	evaluates every incoming observation against all prior observations, where there is no window size. when the mission is critical and transaction volumes can be managed in real time, then perpetual analytics is a better answer.

What is the Hadoop Distributed File System (HDFS) designed to handle?

A) unstructured and semistructured relational data

B) structured and semistructured relational data

C) structured and semistructured non-relational data

D) unstructured and semistructured non-relational data

Hadoop Technical Components:

- Hadoop Distributed File System (HDFS)
- Name Node (primary facilitator)
- Secondary Node (backup to Name Node)
- Job Tracker
- Slave Nodes (the grunts of any Hadoop cluster)
- Additionally, Hadoop ecosystem is made-up of a number of complementary sub-projects

How does Hadoop work?

A) It integrates Big Data into a whole so large data elements can be processed as a whole on one computer.

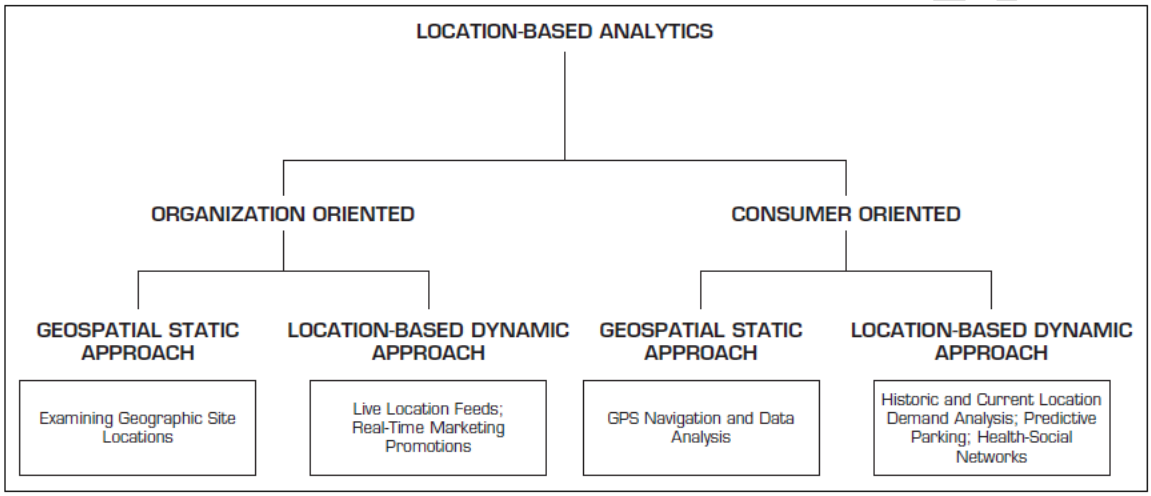
B) It integrates Big Data into a whole so large data elements can be processed as a whole on multiple computers.

C) It breaks up Big Data into multiple parts so each part can be processed and analyzed at the same time on multiple computers.

D) It breaks up Big Data into multiple parts so each part can be processed and analyzed at the same time on one computer.

Ch14: Business Analytics: Emerging Trends and Future Impacts

Location-Based Analytics is Integrate "where" into customer view.



Geographic Information System (GIS): Used to capture, store, analyze, and manage the data linked to a location.

Location Intelligence (LI)? Interactive maps that further drill down to details about any location

Real-Time Location Intelligence:

- Targeting right customer based on their behavior over geographic locations Example Radii app
- Augmented reality: Cachetown - augmented reality-based game.

A critical emerging trend in analytics is the incorporation of location data. ___ Geospatial _____ data is the static location data used by these location-based analytic applications. Fill

What are **recommender systems**, how are they developed, and how is the data used to build a recommendation system obtained? May coming long Q

The term **recommender systems** refers to a Web-based information filtering system that takes the inputs from users and then aggregates the inputs to provide recommendations for other users in their product or service selection choices.

Two basic **approaches** that are employed in the development of recommendation systems are **collaborative filtering** and **content filtering**.

Collaborative filtering:

- Based on previous users' purchase/view/rating data
- Collectively deriving **user ⇔ item** profiling
- Use this knowledge for item recommendations
- Techniques include user-item rating matrix, kNN, correlation, …
- Disadvantage – requires huge amount of historic data

Content filtering:

- Based on specifications/characteristics of items (not just ratings)
- First, characteristics of an item are profiled, and then the content-based individual user profiles are built
- Recommendations are made if there are similarities found in the item characteristics
- Techniques include decision trees, ANN, Bayesian classifiers

The data necessary to build a recommendation system are collected by Web-based systems where each user is specifically asked to rate an item on a rating scale, rank the items from most favorite to least favorite, and/or ask the user to list the attributes of the items that the user likes.

Relate Web 2.0 to knowledge management? هذه سؤال واجب و باخر تلخيصي جزء خاص عنو فاعتقد بما انو كم مره ذكر بالكتاب,, فالله اعلم حا يكون احد الاسئلة الشورت

Answer: Web 2.0 is a collective term for interactive applications that allow communication and collaboration online. These applications include mash ups, social networks, media-sharing sites, RSS, blogs, and wikis. They make it easy and natural to share knowledge, thereby giving knowledge management a strong boost. In fact, they have made the term knowledge management almost redundant because nontechnical people now can readily share their knowledge. The ultimate value of Web 2.0 in knowledge management is its ability to foster greater responsiveness, better knowledge capture and sharing, and more effective collective intelligence.

>>>>>>>

A social network is a place where people create their own space, or homepage, on which they write blogs.

Cloud Computing: A style of computing in which dynamically scalable and often virtualized resources are provided over the Internet.

Users need not have knowledge of, experience in, or control over the technology infrastructures in the cloud that supports them. T

- Web-based email → cloud computing application
- Web-based general application = cloud application

Service-oriented thinking is one of the fastest growing paradigms in today's economy. Which of the following is NOT a characteristic of service-oriented DSS?

A) reusability

B) extensibility

C) substitutability

D) originality

Service-Oriented Thinking Component-based service orientation fosters are : Reusability, Substitutability, Extensibility, Scalability, Customizability, Reliability, Low Cost of Ownership, Economy of Scale,…

Service-oriented thinking is one of the **fastest-growing paradigms** today. T

Major Components of Service-Oriented DSS/BI		
Data-as-a-Service (DaaS)	Information-as-a-Service (IaaS)	Analytics-as-a-Service (AaaS)
■ Accessing data "where it lives" ■ Enriching data quality with centralization ■ Better MDM, CDI	■ "Information on Demand" ■ Goal is to make information available quickly to people, processes, and applications across the business (agility)	■ "Agile Analytics" ■ AaaS in the cloud has economies of scale, better scalability, and higher cost savings ■ Data/Text Mining + Big Data → Cloud Computing

__ Information-as-a-Service _____ (IaaS) promises to eliminate independent silos of data that exist in systems and infrastructure and enable sharing real-time information for emerging apps, to hide complexity, and to increase availability with virtualization. Fill

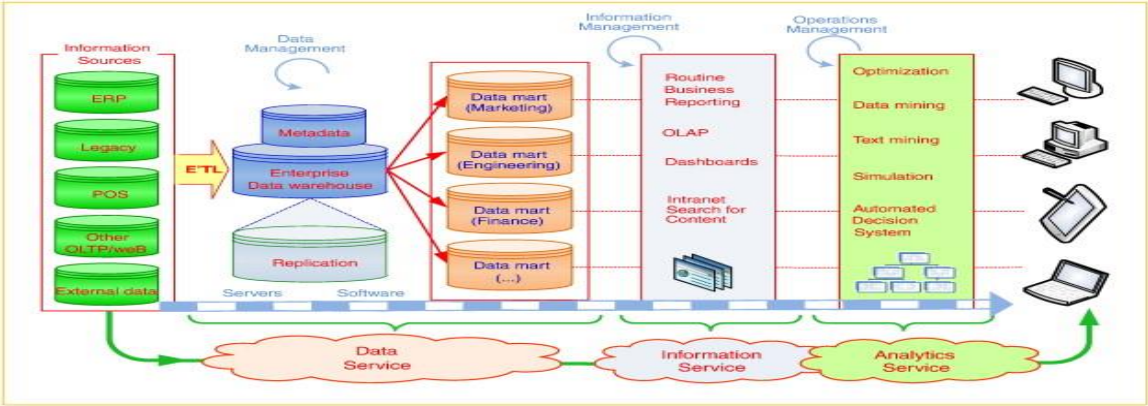
Which of the following is true of data-as-a-Service (DaaS) platforms?

A) Knowing where the data resides is critical to the functioning of the platform.

B) Business processes can access local data only.

C) Data quality happens on each individual platform.

D) There are standardized processes for accessing data wherever it is located.



Which component of service-oriented DSS can be defined as data that describes the meaning and structure of business data, as well as how it is created, accessed, and used?

A) metadata management

B) operations and administration

C) application programming interface

D) analytics

Which component of service-oriented DSS includes such examples as optimization, data mining, text mining, simulation, automated decision systems?

A) analytics

B) operations and administration

C) metadata management

D) application programming interface

Which component of service-oriented DSS can be described as a subset of a data warehouse that supports specific decision and analytical needs and provides business units more flexibility, control, and responsibility?

A) information services with library and administrator

B) extract, transform, load

C) information delivery portals

D) data marts

Impacts of Analytics in Organizations:

- Analytics departments: Chief Analytics Officer, Chief Knowledge Officer
- Restructuring Business Processes and Virtual Teams: Reengineering and BPR
- Job Satisfaction
- Job Stress and Anxiety
- Impact on Managers' Activities/Performance

Soo…there is some Q for it like :

organization-wide, major restructuring is needed, the process is referred to as ___Reengineering _____. fill

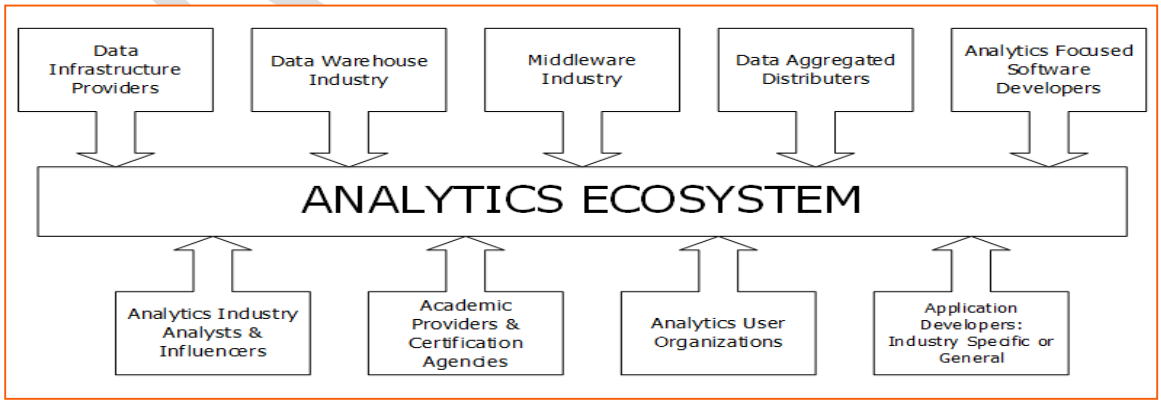
A major structural change that can occur when analytics are introduced into an organization is the creation of new organizational ___units_____. Fill

Privacy: "the right to be left alone and the right to be free from unreasonable personal intrusions"

List 5 for Ethics in Decision Making and Support are:

- Electronic surveillance
- Software piracy
- Invasion of individuals' privacy
- Use of proprietary databases
- Use of knowledge and expertise

Define or what is "The Analytics Ecosystem":



- Analytics Industry Clusters
- Data Infrastructure Data Warehouse Providers
- Middleware/BI Platform Industry
- Data Aggregators/Distributors
- Analytics-Focused Software Developers
- Application Developers or System Integrators
- Analytics User Organizations
- Analytics Industry Analysts and Influencers
- Academic Providers and Certification Agencies

Analytics Ecosystem - Titles of Analytics Program Graduates. T

Short answers maybe… what are Representative Characteristics of Web 2.0?

- Allows tapping into the collective intelligence of users
- Data is made available in new or never-intended ways
- Relies on user-generated/user-controlled content/data
- Lightweight programming tools for wider access
- The virtual elimination of software-upgrade cycles
- Users can access applications entirely through a browser
- An architecture of participation and digital democracy
- A major emphasis is on social networks and computing
- Strong support for information sharing and collaboration
- Fosters rapid and continuous creation of new business models

ملاحظه: ماكتبت كل المعلومات لانها معلومات عامه , من وجهة نظري الاشياء الاساسية حطيتها و في البعض منها مو شرط احط السلايد و لكن حطيت السؤال الخاص بيها و كمان حاولت اكتب كل المهم الي ممكن يجي سؤال اسيه . لا تنسووو الواجبات لازم لازم تتطلعو عليها.

☺موفقين جميعا و دعواتكم لنا