1. Natural language processing (NLP), a subfield of artificial intelligence and computational linguistics, is an important component of text mining. What is the definition of NLP?

NLP is a discipline that studies the problem of "understanding" the natural human language, with the view of converting depictions of human language into more formal representations in the form of numeric and symbolic data that are easier for computer programs to manipulate.

1. What is the difference between white hat and black hat SEO activities?

An SEO technique is considered white hat if it conforms to the search engines' guidelines and involves no deception. Because search engine guidelines are not written as a series of rules or commandments, this is an important distinction to note. White-hat SEO is not just about following guidelines, but about ensuring that the content a search engine indexes and subsequently ranks is the same content a user will see.

Black-hat SEO attempts to improve rankings in ways that are disapproved by the search engines, or involve deception or trying to trick search engine algorithms from their intended purpose.

1. In what ways does the Web pose great challenges for effective and efficient knowledge discovery through data mining?

•The Web is too big for effective data mining. The Web is so large and growing so rapidly that it is difficult to even quantify its size. Because of the sheer size of the Web, it is not feasible to set up a data warehouse to replicate, store, and integrate all of the data on the Web, making data collection and integration a challenge.

•The Web is too complex. The complexity of a Web page is far greater than a page in a traditional text document collection. Web pages lack a unified structure. They contain far more authoring style and content variation than any set of books, articles, or other traditional text-based document.

•The Web is too dynamic. The Web is a highly dynamic information source. Not only does the Web grow rapidly, but its content is constantly being updated. Blogs, news stories, stock market results, weather reports, sports scores, prices, company advertisements, and numerous other types of information are updated regularly on the Web.

•The Web is not specific to a domain. The Web serves a broad diversity of communities and connects billions of workstations. Web users have very different backgrounds, interests, and usage purposes. Most users may not have good knowledge of the structure of the information network and may not be aware of the heavy cost of a particular search that they perform.

•The Web has everything. Only a small portion of the information on the Web is truly relevant or useful to someone (or some task). Finding the portion of the Web that is truly relevant to a person and the task being performed is a prominent issue in Web-related research.

1. List and describe four categories of models. Give examples in each category.

•Optimization of problems with few alternatives: Find the best solution from a small number of alternatives; e.g., 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; e.g., linear and other mathematical programming models, network models

•Optimization via an analytic formula: Find the best solution in one step, using a formula; e.g., some inventory models

•Simulation: Find a good enough solution or the best among the alternatives checked, using experimentation; e.g., Monte Carlo simulation

•Heuristics: Find a good enough solution, using rules; e.g., heuristic programming, expert systems

•Predictive models: Predict the future for a given scenario; e.g., forecasting models, Markov analysis

1. List four well-known search methods used in the choice phase of problem solving.

• Analytical techniques

• Algorithms

• Blind searching

• Heuristic searching

1. Give a brief definition of genetic algorithms.

Genetic algorithms are sets of computational procedures that conceptually follow the steps of the biological process of evolution.

.

1. Give a simple definition of simulation in MSS. ?Under what circumstances is simulation normally used? List five major types of simulation?

A simulation is the appearance of reality.

Simulation is normally used only when a problem is too complex to be treated using numerical optimization techniques.

• probabilistic simulation

• time-dependent and time-independent simulation

• visual simulation

• system dynamics modeling

• agent-based modeling

.

1. List five disciplines of artificial intelligence.

• Human Behavior

• Pattern Recognition

• Statistics

• Information Systems

• Engineering

• Computer Science

• Mathematics

1. What are three components that may be included in an expert system in addition to the three major components found in virtually all expert systems?

• Knowledge acquisition subsystem

• Blackboard (workplace)

• Explanation subsystem (justifier)

• Knowledge-refining system

1. Conventional System Vs Expert System 



1. 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.

1. 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?
2. 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.
3. Capture knowledge. New knowledge must be identified as valuable and be represented in a reasonable way.
4. 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.
5. Store knowledge. Useful knowledge must be stored in a reasonable format in a knowledge repository so that others in the organization can access it.
6. Manage knowledge. Like a library, a repository must be kept current. It must be reviewed to verify that it is relevant and accurate.
7. Disseminate knowledge. Knowledge must be made available in a useful format to anyone in the organization who needs it, anywhere and anytime.
8. Briefly describe three benefits (process gains) derived from working in groups.
* It provides learning. Groups are better than individuals at understanding problems.
* People readily take ownership of problems and their solutions. They take responsibility.
* Group members have their egos embedded in the decision, so they are committed to the solution.
* Groups are better than individuals at catching errors.
* A group has more information (i.e., knowledge) than any one member. Group members can combine their knowledge to create new knowledge. More and more creative alternatives for problem solving can be generated, and better solutions can be derived (e.g., through stimulation).
* A group may produce synergy during problem solving. The effectiveness and/or quality of group work can be greater than the sum of what is produced by independent individuals.
* Working in a group may stimulate the creativity of the participants and the process.
* A group may have better and more precise communication working together.
* Risk propensity is balanced. Groups moderate high-risk takers and encourage conservatives.
1. List and describe the three main "V"s that characterize Big Data.
* Volume: This is obviously the most common trait of Big Data. Many factors contributed to the exponential increase in data volume, such as transaction-based data stored through the years, text data constantly streaming in from social media, increasing amounts of sensor data being collected, automatically generated RFID and GPS data, and so forth.
* Variety: Data today comes in all types of formats–ranging from traditional databases to hierarchical data stores created by the end users and OLAP systems, to text documents, e-mail, XML, meter-collected, sensor-captured data, to video, audio, and stock ticker data. By some estimates, 80 to 85 percent of all organizations' data is in some sort of unstructured or semistructured format.
* Velocity: This refers to both how fast data is being produced and how fast the data must be processed (i.e., captured, stored, and analyzed) to meet the need or demand. RFID tags, automated sensors, GPS devices, and smart meters are driving an increasing need to deal with torrents of data in near—real time.
1. When considering Big Data projects and architecture, list and describe five challenges designers should be mindful of in order to make the journey to analytics competency less stressful.
* Data volume: The ability to capture, store, and process the huge volume of data at an acceptable speed so that the latest information is available to decision makers when they need it.
* Data integration: The ability to combine data that is not similar in structure or source and to do so quickly and at reasonable cost.
* Processing capabilities: The ability to process the data quickly, as it is captured. The traditional way of collecting and then processing the data may not work. In many situations data needs to be analyzed as soon as it is captured to leverage the most value.
* Data governance: The ability to keep up with the security, privacy, ownership, and quality issues of Big Data. As the volume, variety (format and source), and velocity of data change, so should the capabilities of governance practices.
* Skills availability: Big Data is being harnessed with new tools and is being looked at in different ways. There is a shortage of data scientists with the skills to do the job.
* Solution cost: Since Big Data has opened up a world of possible business improvements, there is a great deal of experimentation and discovery taking place to determine the patterns that matter and the insights that turn to value. To ensure a positive ROI on a Big Data project, therefore, it is crucial to reduce the cost of the solutions used to find that value.
1. Hadoop Vs MapReduce.
* Hadoop is an open source framework for storing and analyzing massive amounts of distributed, unstructured data
* MapReduce distributes the processing of very large multi-structured data files across a large cluster of ordinary machines/processors to achieving high performance with “simple” computers.
1. Give examples of Synchronous and Asynchronous Tools for Indirect Support of Decision Making.

– Synchronous products

• Web conferencing

• Instant messaging (IM)

• Voice over IP (VoIP)

– Asynchronous products

• E-mail

• Wikilogs

• Online workspaces