PDD ASSIGNMENT

Done By,
Harikrishnan.V
CS-4A
Roll No: 27

SEMANTIC WEB

Semantic web is a web with a meaning. It an extension to the world wide web & is a universal / global database. Semantic web is a vision of information that is understandable by computers, so that they can perform more of the tedious works involved in finding, sharing & combining information on the web. It provides a common standard (RDF) for websites to publish the relevant information in a more readily machine processable & integratable form.

Semantic web involves publishing the data in a language, resource description framework specially for data, so that it can be manipulated & combined.

HTML describes documents & links between them. RDF describes arbitrary things such as people, meeting & airplane parts.

Components of Semantic Web

(i) Identifiers:- Uniform Resource Identifier (URI)

(ii) Documents:- Extensible Markup Language (XML)

(iii) Statements:- Resource Description Framework (RDF)

Schemas and Ontologies

(i) Logic

(ii) Proof

(iii) Trust

eg: A computer might be constructed to list the prices of flat screen HD TVs larger than 40 inches at shops in the nearest town that are open until 8 pm on Tuesday evening.

RESOURCE DESCRIPTION FRAMEWORK - RDF

RDF is the most fundamental building block of semantic web. It is a format for defining information on the web. It is a markup language for describing information & resources on the web.

RDF provides a model for data & a syntax so that independent parties can exchange & use it.

It is designed to be read & understood by computers.

Each RDF statement has 3 parts :-

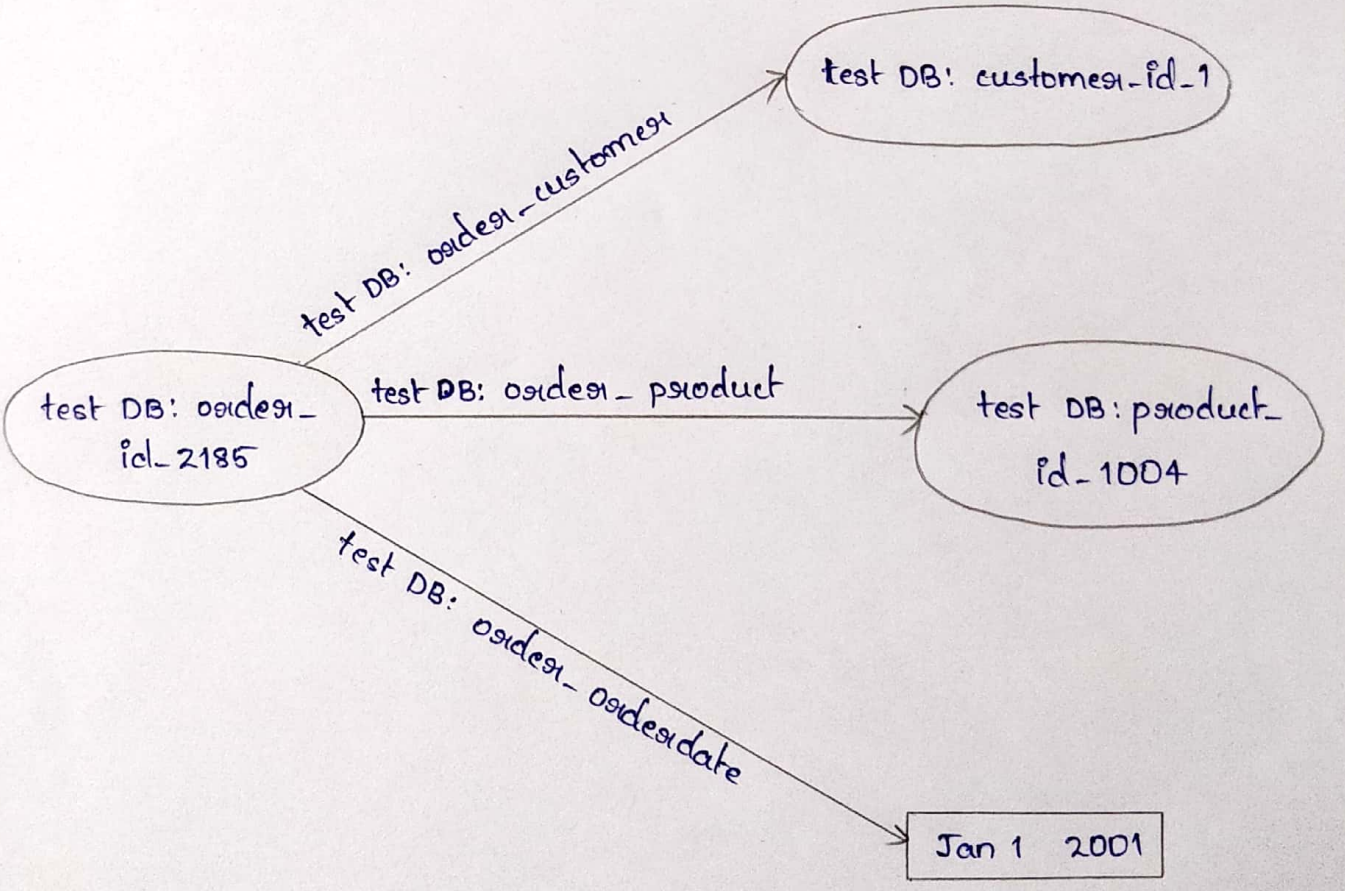
(i) A subject

(ii) A predicate

(iii) An object

RDF is a language for expressing directed labelled graph using URIs as nodes and arc identifiers.

eg:



GEOGRAPHIC INFORMATION SYSTEM - GIS

GIS is used to collect, model, store & analyze information describing physical properties of the geographical world.

Functions of a GIS

(i) Data can be positioned by its known spatial coordinates.
(ii) Data can be organized.
(iii) Data can be stored & retrieved.
(iv) Data can be modified & displayed.

GIS can be represented as raster, vector & real world

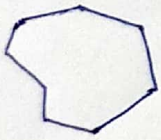
(i) Raster :- Stores images as rows & columns of numbers with a digital value for each cell.

(ii) Vector :- Allows user to specify specific spatial locations & assume that geographic space is continuous, not broken into discrete grid squares. It stores features as set of (x,y) coordinate pairs

We typically represent object in space as 3 distinct spatial elements.

• → point


→ Lines (arcs) [set of connected points]


→ Polygon [set of connected lines]

In raster data model the cell value is the attribute
eg: Brightness
For vector data, attribute records are linked to point, line & polygon features

GIS Applications

(i) Road network analysis
(ii) Route planning
(iii) Site selector for helicopter landing

Features

(i) Multilayer mapping
(ii) Data visualization features:-
   → Plotting category
   → Colouring administrative territories based on dynamic scaling
   → Plotting graph
   → Attaching photo images.
(iii) Navigation features:-
   → Free zoom in / zoom out.
   → Mini map preview.
(iv) Distance measurement
(v) Export to pdf, word & print
(vi) Map retrieval via web services

BIOLOGICAL DATABASE

Bio-informatics is all that concerns with biological databases & software tools written to manipulate biological databases.

Biological databases are the collection of scientific data information generated by individuals regarding a particular biological aspect, grouped & well documented.

Information in the database can be searched, compared, retrieved & analyzed.

Characteristics of Biological Data

(i) Biological data is highly complex when compared with most other domains or applications.

eg: MITOMAP is a database documenting the human mitochondrial genome. This genome is a small, circular piece of information about 16569 nucleotide bases.

(ii) The amount & range of variability in a data is high.

(iii) Schemas in biological databases change at a rapid pace.

eg: Presently system such as Genbank rerelease the entire database with new schema once or twice a year.

(iv) Representation of same data by different biologists will be different.

(v) Most users of biological data do not require write access to database; read-only access is adequate.

(vi) Most biologists are not likely to have any knowledge of internal structure of the database or about schema design.

(vii) The context of data gives added meaning for its use in biological applications.

(viii) Defining & representing complex queries is extremely important to biologists.

(ix) Users of biological information often require access to old value of data.

BIG DATA

Big data is a term for data sets that are so large or complex that traditional data processing application softwares are inadequate to deal with them.

It can be defined as volumes of data available in varying degrees of complexity, generated at different velocities & varying degrees of ambiguity that cannot be processed using traditional technologies, processing methods, algorithms.

Characteristics

(i) The extreme volume of data.
(ii) The wide variety of data types.
(iii) Velocity at which the data must be processed.

Applications

(i) Banking
(ii) Government
(iii) Retail.
(iv) Health care