Semantic Web

(Semantic Network)

Semantic Web: Intended to become the next generation of the Internet, where Internet components understand the meanings of words and relationships. The Semantic Web (sometimes referred to as "Web 3.0") is a kind of Semantic Network (a type of "Cloud Computing"). Semantic Networks use nodes and arcs to represent structured knowledge, where the nodes are concepts or entities and the arcs are relationships and associations.

Problem Domain: In the partitioning of a problem space into components to be considered and not to be considered, the problem domain is those components to be considered. The problem domain also considers any connections to external entities that might influence it.

Ontology: Theory about what types of things exist within a problem domain. An ontology has a taxonomy, a thesaurus and a set of inference rules.

Taxonomy: Structured ways to organize domain knowledge (e.g. life ⇒ fauna ⇒ mammal ⇒ canine ⇒ dog ⇒ spaniel ⇒ cocker ; Dewey Decimal system) so it can more easily be understood and shared.

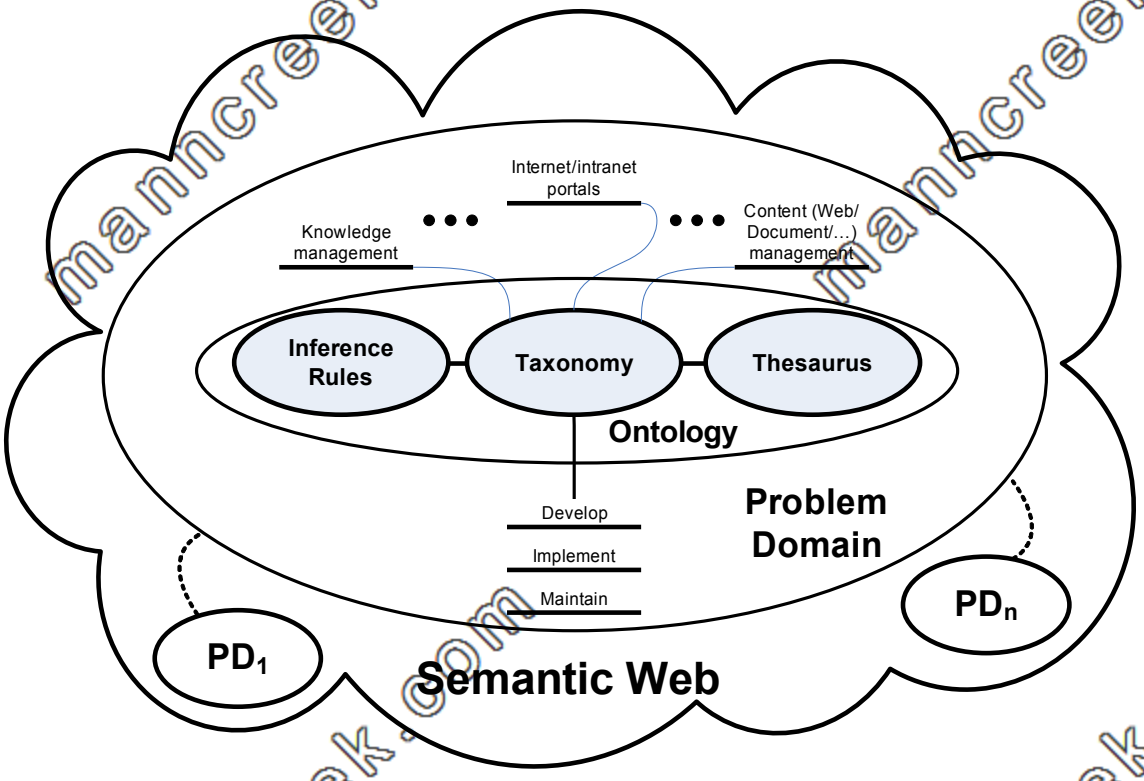
Inference Rules: Premises known or assumed to be true from which logical domain conclusions are derived. The act of reasoning from factual domain knowledge or evidence.

Thesaurus: A detailed vocabulary. Information on preferred and related domain terms, and synonyms.

Taxonomy Development (also see next page): Choose some combination of software (automatic) classification and human classification (usually statistical, but may include natural language processing based on thesaurus / dictionary); choose existing off-the-shelf or new in-house taxonomy; balance taxonomy width and depth to maximize simplicity and understanding (juggle number/distribution of nodes); use skills of (corporate) librarians.

Taxonomy Implementation: Prototype with major go/no-go decision points; involve stake holders at all levels; increment and iterate, including repeated verification and validation.

Taxonomy Maintenance: Include user feedback; update topic maps constantly based on usage and changing terminology; use automation as much as possible but allow for human override.



Semantic Web

(Semantic Network)

Steps in taxonomy building:

- Determine purpose
- Determine scope (type of questions to be answered)
- Choose taxonomy type
  - Hierarchical (each item appears once)
  - Poly-hierarchical (items cross-listed in multiple categories)
  - Faceted (each facet has its own hierarchy)
- Determine source base
- Determine width and depth based on user experience
- Make business case (possible use of prototype / simulation)
  - Value added
  - Cost
  - ROI
  - Failure risk
- Choose tool framework
  - Categorization (taxonomy development)
  - Classification (apply categorization to sources(s); static/dynamic)
    - E.g. linguistic (dealing with sentence structure), statistical (word counts)
    - Typically, combination of manual (top down) and automated (bottom up)
  - Search (e.g. with stemming, spell correction, refinement, …)
- Iterative and incremental implementation processes
  - Stakeholder consensus on requirements
  - Use case building
  - Requirements analysis
  - Design solution (including access, content)
  - Build solution
  - Test implementation (usability, security, performance, satisfaction, productivity, …)
  - Maintenance (monitor, manage change, measure success/use, stakeholder feedback)

Web Ontology Language (OWL) is a markup language (written in XML) for publishing and sharing data using ontologies on the Internet. OWL has three sub-languages (OWL Full(OWL DL(OWL Lite))). OWL is a vocabulary extension of the Resource Description Framework (RDF)

Resource Description Framework (RDF): In general, RDF is a way to define Web resource metadata. OWL, RDF and other components (ref. following) can make up a Semantic Web project.

Ontology Inference Layer (OIL) is an XML derivative for specifying entity relationships.

Web Services Description Language (WSDL), Universal Description, Discovery and Integration language **(UDDI)**: Web-services related languages.