Semantic Technologies for Nuclear Knowledge Modelling and Applications

D. Beraha

3rd International Conference on Nuclear Knowledge Management
7.-11.11.2016, Vienna, Austria
Why Semantics?

• Machines “understanding” content are changing the notions of Knowledge and Knowledge Management profoundly, e.g.
  • Knowledge at everybody’s fingertips
    • Knowledge graphs in search engine and Knowledge discovery
  • Watson wins Jeopardy [https://en.wikipedia.org/wiki/Watson_(computer)]
  • Google Deep Mind beats Go champion [https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol]
  • Industry 4.0 [https://en.wikipedia.org/wiki/Industry_4.0], France’s “Industrie du Futur”
  • ...
• Semantic technologies comprise a wide range of tools and methodologies based on Standards set by the World Wide Web Consortium (W3C) that allow to process information according to its meaning in context
Semantic technologies in the nuclear domain

• Not established to much extent. Obstacles:
  • High requirements on the reliability of information
  • Limitations in distribution of information (classified information, corporate confidential information, operation and research data)
  • Wide diversity of (sometimes quite aged) information technologies
  • Conservative approaches compared to other business areas, resistance to implementation of new information technologies

→ Knowledge and information repositories are mostly stand-alone, with limited connectivity, confined within the organization ("silos")
→ However, access to distributed and heterogeneous sources is highly desirable for e.g. improved information retrieval and discovery or ("smart") applications
Establishing a common language

- Many groups of different stakeholders collaborate during the lifecycle of a facility.
- Using terms defined in a Knowledge Organization System (KOS – vocabularies, taxonomies, thesauri, ontologies) reduces uncertainties and prevents errors of misunderstandings.
  - Terminology is important since used in licensing documents, contracts, specifications, design documents, ...
- KOS’s should be published according to W3C Standards as Linked Data.
  -Readable by people and by machines on the Web.
  -Linking between KOS’s and other data sources (public, e.g. Wikipedia, or restricted) provides more information (definitions, examples, descriptions).
  -Reusable by third party applications.
Thesauri and glossaries exist (e.g. INIS, safety glossaries published by several organizations, thematical ontologies), but
- Developed independently, so might have inconsistent, even contradicting definitions
- Application: International Nuclear Terminology Repository Platform
Vision: “Nuclear Ontology”

Connected by relations (sameAs, related to, subclassOf, broader, narrower …)

May overlap
Text Analytics, Data Mining and Knowledge Discovery

- Aim: extract new insights from the huge amount of data continuously produced by many information sources
- Discover patterns in mostly unstructured data
- Techniques include e.g.
  - Natural language processing
  - Machine learning, neural networks
  - Statistical analysis
  - Fuzzy systems
- Application: Extracting Lessons Learning from Operational Experience (OPEX)
  - Extracting concepts and finding associations between them, e.g. temporal relations, relations to components, similar events ...
**Integration of heterogeneous knowledge sources**

- Nuclear domain very well documented, many repositories exist
  - in single organizations distributed among organizational units
  - on national and international level, in different languages
- Documents are usually related to other information
- Semantic technologies are the means to break the barriers of non-interoperable, heterogeneous data sources (data bases, file servers, repositories on PC’s …), enabling queries across all sources
- Applications: Distributed data base for Competency Networks of Networks; International Nuclear Management Academy (INMA)
  - training courses, university degrees, PhD topics, opportunities, educational material, organizations
  - Queried from a central access point
• **Semantic search and auto-tagging of documents**
  • Semantic Search: context, substance, intent of search
  • Based on controlled vocabularies with synonyms, acronyms, variant spellings, relations
  • Importance of annotation with metadata
    • Available KOS’s (e.g. Dublin core)
    • Categories and subject keywords
      • Manual assignement infeasible with large repositories → **autotagging**
      • Term extraction supported by KOS’s
  • Applications: Search in large document repositories
    • Advances in autotagging methods provide good results in tagging and searching
• Knowledge portals and Wikis
  • A KOS is an excellent starting point for constructing a K-Portal
    • A taxonomy can be directly translated in pages and subpages
    • KOS relations provide the linking of pages
    • External links by Linked Data methods provide information from other sources
    • (Auto-)tagging of documents and semantic search
  • Software today includes capabilities for handling semantic Standards and KOS‘s
    • Extensions for extraction, metadata management, taxonomies etc. are available
  • Applications:
    • Several knowledge portals (e.g. fast reactors, nuclear accidents, decommissioning and environmental remediation) at IAEA
    • NKM Wiki, available on the NKM homepage
Many more applications are emerging, e.g.

- **Plant Information Models**
  - A representation of the plant throughout its lifecycle by linking facility information, relationships, rules and knowledge frameworks, and virtual reality techniques

- **Learning Objects Repositories for Educational Networks**
  - Storing, retrieving and linking of learning objects described by semantic metadata

- **Competency networks**
  - An integration of job descriptions, competencies required for given tasks, taxonomies describing required skills and knowledge may fill gaps between job requirements and education

- **Crisis Management and Emergency Preparedness**
  - Ensuring interoperability by domain vocabularies used by several information systems and groups of people is the key to share and operate on information from several data sources in crisis management
Conclusions

• Semantic technologies are changing the way we work with knowledge on personal, organizational and inter-organizational levels
  • Semantic technologies are starting to be deployed within the nuclear domain
• Activities in the NKM section
  • A semantic technology platform is available with
    • Tools for developing and managing KOS’s
    • Interlinking KOS’s and publishing according to Linked Data principles
    • Auto-tagging of documents and semantic search within Knowledge Portals
  • Exploratory projects and proofs of concept under way
  • A document on Exploring Semantic Technologies and their Application to Nuclear Knowledge Management to be published soon
• Member States are invited to collaborate in developing semantic applications for nuclear knowledge management