TDDD43 Advanced Data Models and Databases

Examples of exam question types

### Information Retrieval

* Describe the components of IR models.
* Explain and exemplify the basic principles of boolean/vector/probabilistic model for IR (different components).
* Describe the advantages and drawbacks of the boolean/vector/probabilistic model.
* Explain TF-IDF.

### Semi structured data, XML, RDF

* Describe general properies of semi-structured data.
* Model a given scenario or set of data in the OEM model.
* Explain a Lorel query.
* Describe the general idea and define the properties of a data guide.
* Define minimal and strong data guides.
* Given a description of data, construct minimal and strong data guides for that data.
* Describe the general properties and concepts of XML.
* Model a given scenario (text or ER) using XML.
* Model a given scenario (text or ER) using RDF.
* Use and understand DTD, XML schema, RDF schema
* Write an XPath query.
* Write an XQuery query.
* Explain a SPARQL query; show the result of a SPARQL query on given data.
* Explain what Linked (Open) Data is.

### Semantic web and ontologies

* Describe the problems in the current web and the vision of the semantic web.
* Show how semantic annotations based on ontologies would help solve the problems of the current web.
* Explain what ontologies are used for.
* Describe and explain the OBO Foundry principles.
* Explain and give examples of the components of ontologies.
* Describe from a knowledge representation point of view the different kinds of ontologies.

### Description logics and OWL

* Describe the notions of T-box, A-box, knowledge base, subsumption, satisfiability.
* Model a given a scenario using description logics.
* Give the semantics for a given description logic construct.
* Describe the difference between open-world assumption and closed-world assumption.
* Describe the different reasoning services.
* Given 2 concepts, prove that one concept subsumes the other (or not) using a tableau algorithm.
* Know some reasons for intractibility, undecidability of description logics.
* Know the difference between the variants of OWL.

### NoSQL databases

* Explain the basic concepts of NoSQL.
* Explain the difference of vertical scalability and horizontal scalability.
* Explain the CAP theorem.
* Understand the principles of consistent hashing.
* Understand the principles of vector clocks.
* Explain the basic principles and steps of MapReduce.
* Give an example of applications that benefit from the MapReduce model.

### Ontology alignment and debugging

* Describe the problem of ontology alignment.
* Describe a framework for ontology alignment and explain the different components.
* Explain and give examples for different strategies for preprocessing/matching/combining/filtering.
* Explain how to evaluate ontology alignment strategies / systems.
* Understand performance of different approaches.
* Know challenges for ontology alignment.
* Explain and exemplify different kinds of defects in ontologies.
* Give examples of the influence of defects in ontologies for semantically-enabled applications.
* Describe a framework for ontology debugging and explain the different components.
* Explain/compute MIPS and MUPS.
* Explain the formulation of GTAP.
* Expain the intuitions behind the preferences (semantically maximal, subset minimal, minmax, maxmin, skyline optimal) for GTAP solutions.

### NOT IN HT 2015 COURSE: Data source integration

* Describe the problems occuring with data source integration.
* Describe the different steps in data source integration.
* Describe the method for integration based on link driven federations.
* Understand a query in the SRS query language.
* Know advantages and disadvantages of link driven federations.
* Describe the method for integration based on mediation.
* Describe mappings and query processing for GAV and LAV.
* Given two data sources (schema, data guide, data, ...), integrate them using GAV or LAV.
* Define the notion of capabilities.