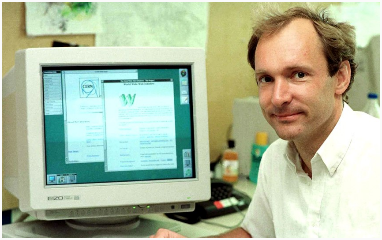
Symbolic Artificial Intelligence
Lecture 2: Knowledge Graphs

Natalia Díaz Rodríguez , PhD

ENSTA Paris, Institute Polytechnique de Paris and INRIA Flowers
flowers.inria.fr http://asr.ensta-paris.fr/ nataliadiaz.github.io
natalia.diaz@ensta-paris.fr
IA301 Logics and Symbolic Artificial Intelligence
https://perso.telecom-paristech.fr/bloch/OptionIA/Logics-SymbolicAI.html

The Semantic Web Vision:

I have a dream for the Web to become capable of analyzing all the data on the Web - the content, links, and transactions between people and computers. A Semantic Web, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The intelligent agents people have touted for ages will finally materialize.



Tim Berners Lee, CERN, 1999[1]

[1] *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*. T. Berners-Lee with Mark Fischetti. Harper San Francisco, 1999.

OWL vs Other Languages[2]

	DTD	XSD	RDF(S)	OWL
Bounded lists ("X is known to have exactly 5 children")				X
Cardinality constraints (Kleene operators)	X	X		X
Class expressions (unionOf, complementOf)				X
Data types		X		X
Enumerations	X	X		X
Equivalence (properties, classes, instances)				X
Formal semantics (model-theoretic & axiomatic)				X
Inheritance			X	X
Inference (transitivity, inverse)				X
Qualified contraints ("all children are of type person"				X
Reification			X	X

[2] **DTD**: *Document Type Definition*, Markup declarations that define a document type for an SGML-family markup language (SGML, XML, HTML). Defines the legal building blocks of an XML document through a list of legal elements and attributes. **XSD**: *XML Schema Definition*: W3C recommendation to formally describe the elements in an XML document and verify each piece of item content in a document [Lagoze]. **Reification**: the ability to treat a statement as a resource, and hence to make assertions about that statement (to reason in FOL [McCarthy'87,79], relates to *provenance*).

Knowledge Graphs

Knowledge Graphs [10, 7]

What is a Knowledge Graph (KB)[3]?:

- a set of interconnected typed entities and their attributes
- has an ontology as schema defining its vocabulary

[3] originating from Pierce's **existential graphs** and Quillian' Semantic Networks [12] (semantic memory -fact, concept, relationship- models)[10].

Why Knowledge Graphs (KG)? [10] IBM Watson: 1, Humans: 0

- 10% of Watson's winning performance in *Jeopardy* TV quiz game came from represented knowledge
- Explainability

eXplainable AI:

XAI: *a suite of machine learning techniques that produces details or reasons to make its functioning clear or easy to understand*.

XAI draws insights from Social Sciences and the psychology of explanation

Objectives:

(1) produce more explainable models maintaining high level performance

(2) enable humans to understand, trust, and manage the emerging generation of artificially intelligent partners [1].

Given an audience, an explainable AI is one that produces details or reasons to make its functioning clear or easy to understand.

Explainability [1]: important since the 1st expert system MYCIN [16]

Model of Inexact Reasoning in Medicine

It is useful to consider the advantages provided by a rule-based system for computer use of judgmental knowledge. It should be emphasized that we see these advantages as being sufficiently strong in certain environments that we have devised an alternative and approximate approach that parallels the results available from using Bayes' Theorem. I do not argue against the use of Bayes' theory in those medical environments in which sufficient data are available to permit adequate use of the theorem.

The advantages of rule-based systems for diagnostic consultations include:

(1) the use of general knowledge (from textbooks or experts) for consideration of a specific patient; even well-indexed books may be difficult for a nonexpert to use when considering a patient whose problem is not quite the same as those of patients discussed in the text;
(2) the use of judgmental knowledge for consideration of very small classes of patients with rare diseases about which good statistical data are not available;
(3) ease of modification; since the rules are not explicitly related to one another and there need be no prestructured decision tree for such a system, rule modifications and the addition of new rules need not require complex considerations regarding interactions with the remainder of the system's knowledge;
(4) facilitated search for potential inconsistencies and contradictions in the knowledge base; criteria stored explicitly in packets such as rules can be searched and compared without major difficulty;
(5) straightforward mechanisms for explaining decisions to a user by identifying and communicating the relevant rules;
(6) an augmented instructional capability; a system user may be educated regarding system knowledge in a selective fashion, i.e., only those portions of the decision process that puzzle him need be examined.

One of MYCIN's rules, which I shall use for illustrative purposes throughout this chapter, is the following:

IF: 1) THE STAIN OF THE ORGANISM IS GRAM POSITIVE, AND
2) THE MORPHOLOGY OF THE ORGANISM IS COCCUS, AND
3) THE GROWTH CONFORMATION OF THE ORGANISM IS CHAINS
THEN: THERE IS SUGGESTIVE EVIDENCE (.7) THAT THE IDENTITY OF THE ORGANISM IS STREPTOCOCCUS

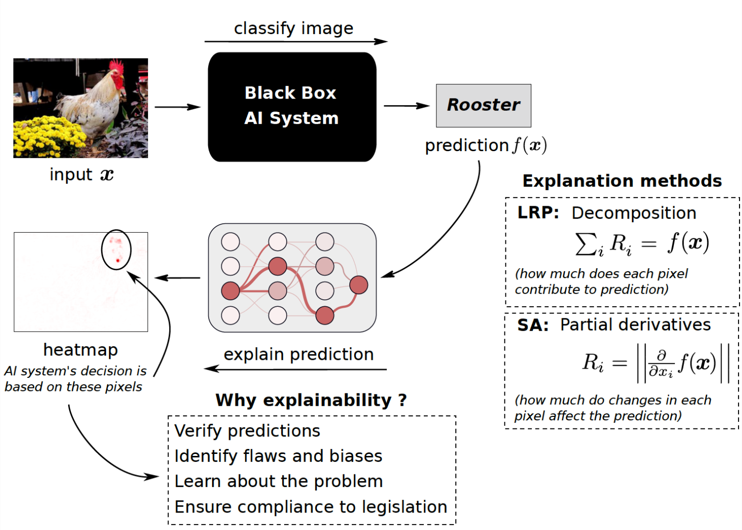
165

Explainability [1, 2] (and comprehensibility): Today

- DARPA XAI Initiative (Explainable AI)
- IJCAI federation of workshops:
  - FAT ML
  - WHI- Workshop on Human Interpretabililty in ML
  - IReDLia-Interpret. & Reasonable Deep Learning and Applications
- ICAPS XAI Planning/NIPS Interpretable ML
- NeSy workshop: `http://www.neural-symbolic.org/`
- GDPR *Right to explanation/be informed* does not exist yet[4]

[4] [20] Art. 13,14, (on notification duties) as it stands, only provides a limited (secret of affairs, etc) right to obtain ex-ante (forecast) explanations about the model

Explaining predictions of an AI system[5]: Why?

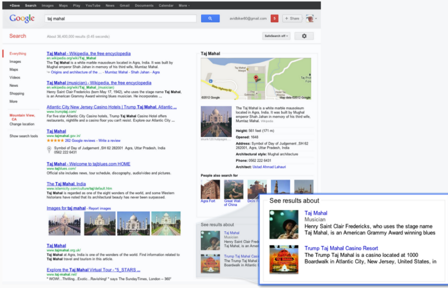


[5] SA: Sensitivity Analysis. LRP: Layer-wise Relevance Propagation [15]

Introducing the Knowledge Graph: *Things, not strings*[6]

Objectives:

- Find the right thing
- Get the best summary
- Go deeper and broader



[6] Google, 2012, `https://www.blog.google/products/search/introducing-knowledge-graph-things-not/`[9, 10]

Knowledge Graphs: Brief history

- Semantic Networks [12]: analyze the meaning of word concepts and the organization of human semantic memory
  - *nodes*: entities, situations;
  - *arcs*: relations: *is-a (instance), part-of, has* (no formal syntax and semantics).

  Ex: `Bird` ← *`is-instance`* `- Penguin -` *`eats`* → `Fish`
- Frames [8]: represent knowledge as collections of separate, simple fragments:

  1 (entity and class) slot = 1 record-like fragment defining relationships, constraints intersections, unions, negations, FOL. **Ex**:

```
Bird
    subclass-of:  Animal
    member-slot:  has-part     value-class:  Wing
Penguin
    subclass-of:  Bird
    colour:  black and white
```

- No standard frame language until 2004 (OWL)

Knowledge Graphs: Brief history (II)

- KL-ONE [4]: Most well known KR frame system
  - 1st supporting DL.
  - 1st using deductive classifier for computing subsumption relations
  - 1st where class hierarchies are *inferred* (vs *asserted* in previous frame systems).
- Semantic Web *stack* re-cap:
  - RDF: the modern W3C recommendation graph-based *standard data model* for semantic networks to describe entities[7].
  - OWL: W3C *standard language* to define rich and complex vocabularies for RDF graph data annotation. Allows concept descriptions and datatypes.
  - Linked Data: Framework to publish, share and link (via RDF and OWL mappings) data across applications and domains[8].
  - SPARQL: the *standard RDF query language* (*the SQL for RDF/OWL graphs*, supports conjunctive & navigational queries)[9].

---

[7] RDF, as semantic networks, does not allow users to define concepts (this is addressed by OWL).

[8] RDF graphs can be linked together via schema-level (e.g., *rdfs:subClassOf*) and entity-level (e.g. *owl:sameAs*) mappings

[9] Other pattern matching languages look for small subgraphs of interests (e.g. look for a clique of 3 individuals that are friends with each other) or navigational queries (when conditions are between nodes that are not necessarily adjacent), RPQ (Regular Path Queries, use RE)

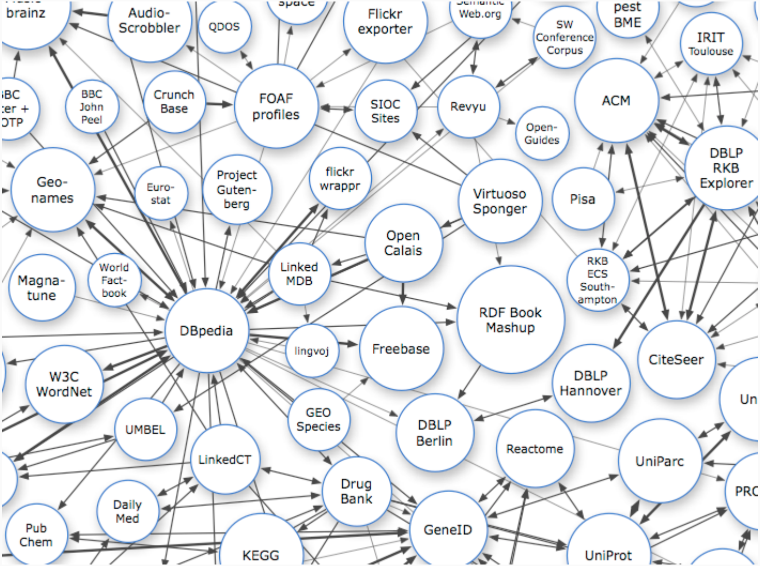
Knowledge Graphs: Brief history (II)

Today's largest KGs:

- Linked Open Data (LOD)
- *NELL*
- Google KG
- Microsoft *Satori*
- *Watson*
- Facebook Graph
- *YAGO*
- *DBpedia*
- *BBC*'s

Let's put these onto *Knowledge Engineering* context!

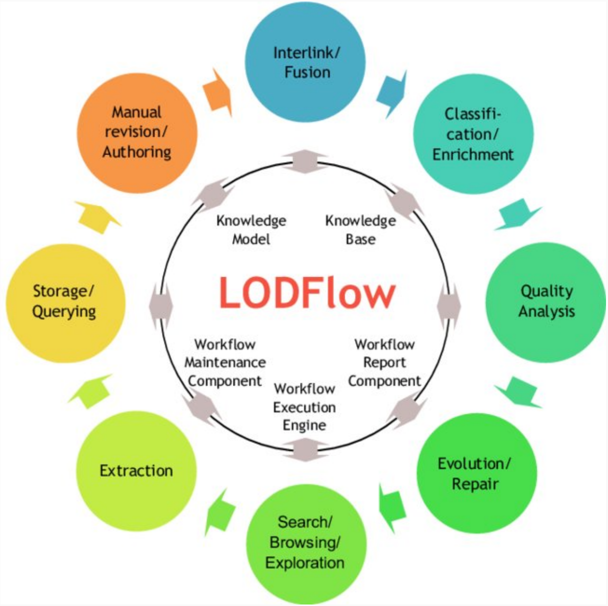
Largest KGs: Linked Open Data (LOD)



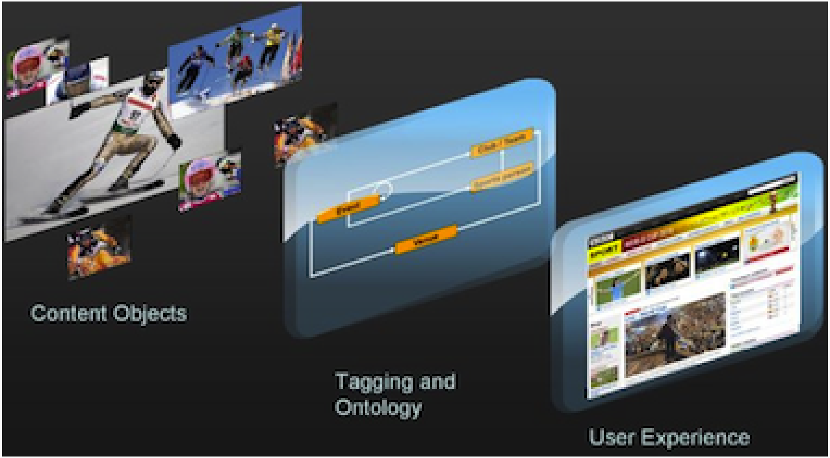
Largest KGs: Linked Open Data (LOD)

- Aim: avoid data silos
- *"Datasets that don't have this LOD ontology logic or interconnection capability (such as DBpedia) are **data feudalism**–data that's limited in its scope, it lacks contextual relevance. We have data manors with well-manicured lawns, but elsewhere lots of impoverished, underdescribed, underconnected data that machines can't help us much with. That's why information overload is so pervasive.*
  $\rightarrow$ *LOD logic allows **data globalism**"*. ["What is LOD?" Quora answer]

Largest KGs: Linked Open Data (LOD) Lifecycle [Auer]

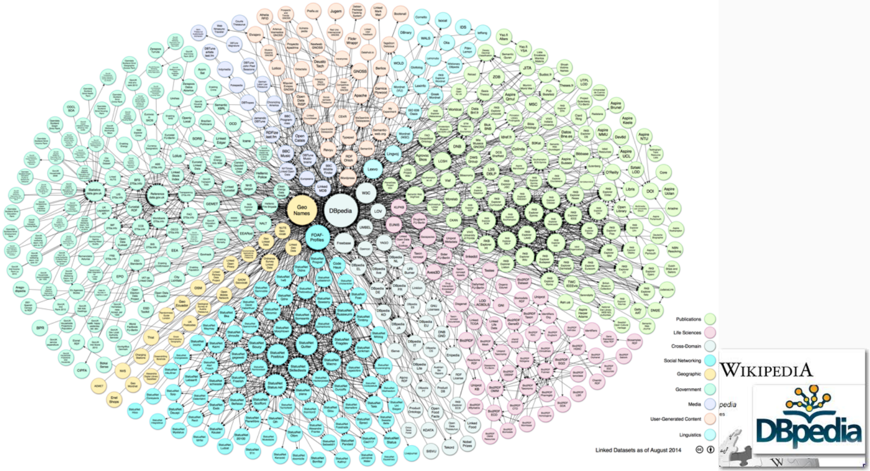


KB examples: BBC User Experience [Source: Ontotext]

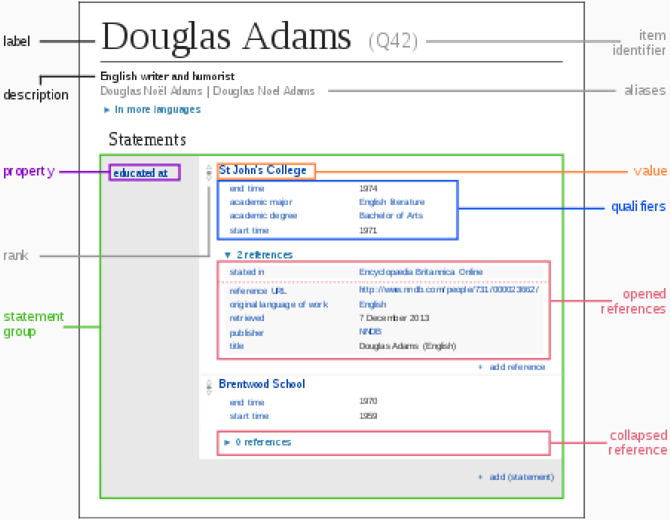


Largest KGs examples: DBPedia Project

Aim: extract structured content from the information created in the Wikipedia and make it available on the WWW

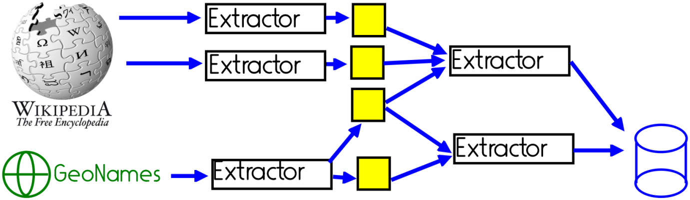


KB examples: and more general: Wikidata



Knowledge base examples

Example: YAGO



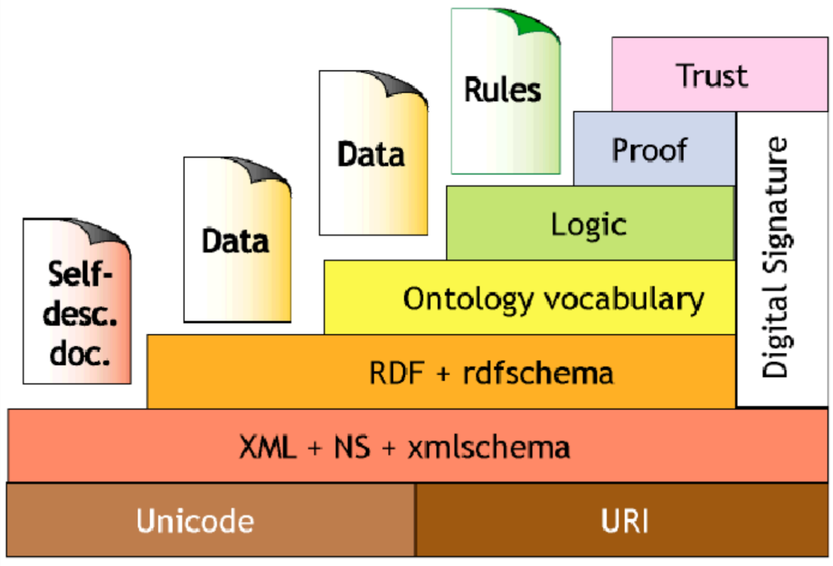
YAGO is a knowledge base that was automatically constructed from Wikipedia and other sources:

- 10m entities, 100m facts
- 95% accuracy
- 1700+ citations on WWW 2007 paper
- 10 languages
- used by IBM Watson, Bloomberg, DBpedia,...

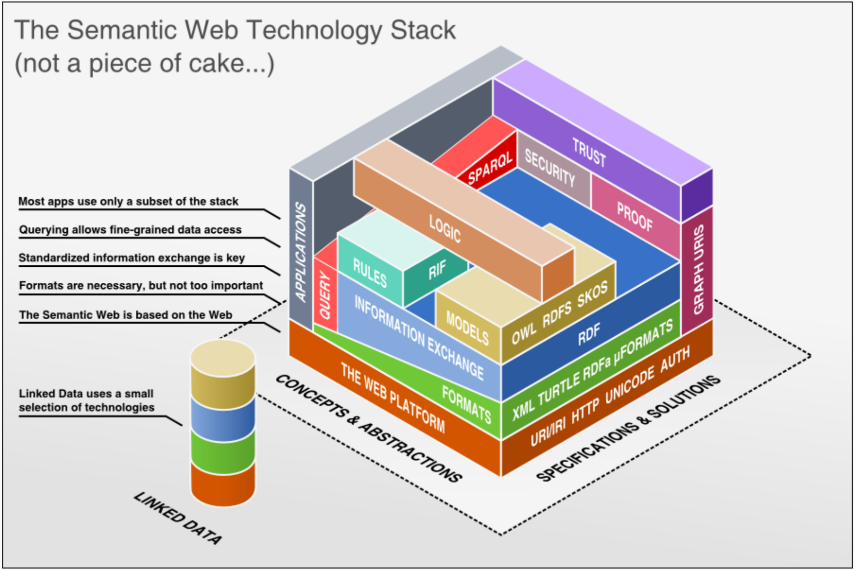
http://yago-knowledge.org

Every good AI has a good cake

From Tim Berners-Lee Semantic Web (2001)...



Every good AI has a good cake [B. Nowack]



Every good AI has a good cake

Yann Lecun's Cake Theory at NIPS 2016

- **"Pure" Reinforcement Learning** (cherry)
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- **Supervised Learning** (icing)
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Unsupervised/Predictive Learning** (cake)
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

- (Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)

Ontology Engineering Methodologies

Building KGs: bottlenecks [9]

Main challenges in ontology design:

- Authoring tools GUI: unable to handle KGs complexity
- Reasoners and debuggers: unable to deal with such complexity efficiently

- *TDKGC* (Test Driven KG Construction): expresses requirements in form of query-answer pairs $T =< q, a >$ and competency questions [9]
- *OOPS!* (OntOlogy Pitfall Scanner): structural ontology evaluation [11] wrt. number of pitfalls[10]
- Defining inconsistency-tolerant semantics [14]:
  - Able to derive meaningful conclusions from inconsistent ontologies (as a formal basis for an automated treatment of inconsistency)
  - *Repair*: a max. subset of the ABox that is consistent with the TBox

[10] *OOPS! Catalogue* includes creating unconnected ontology elements, missing annotations, domain or range in properties, using different naming criteria in the ontology, or recursive definitions. See Pitfall Rate evaluation examples in [6] and `http://oeg-lia3.dia.fi.upm.es/oops/catalogue.jsp`.

- NeON Methodology [19, 17]
- OMQA (Ontology Mediated Question Answering) [3]
- CQOA (Competency Questions Ontology Authoring) [13]:
  What kind of questions the ontology could answer?
  Given an application scenario where a KG is required, how suitable is a given graph for the purposes of this scenario?[11].

[11] CQs: Question expressions an ontology must be able to answer (functional reqs.) [10]

Ontology Design Methods: CQOA (Competency Questions Ontology Authoring)[12]

ID	Pattern	Example	PA	RT	M	DE
1	Which [CE1] [OPE] [CE2]?	Which pizzas contain pork?	2	obj.		
2	How much does [CE] [DP]?	How much does Margherita Pizza weigh?	2	data.		
3	What type of [CE] is [I]?	What type of software (API, Desktop application etc.) is it?	1			
4	Is the [CE1] [CE2]?	Is the software open source development?	2			
5	What [CE] has the [NM] [DP]?	What pizza has the lowest price?	2	data.	num.	
6	What is the [NM] [CE1] to [OPE] [CE2]?	What is the best/fastest/most robust software to read/edit this data?	3	both	num.	
7	Where do I [OPE] [CE]?	Where do I get updates?	2	obj.		spa.
8	Which are [CE]?	Which are gluten free bases?	1			
9	When did/was [CE] [PE]?	When was the 1.0 version released?	2	data.		tem.
10	What [CE1] do I need to [OPE] [CE2]?	What hardware do I need to run this software?	3	obj.		
11	Which [CE1] [OPE] [QM] [CE2]?	Which pizza has the most toppings?	2	obj.	quan.	
12	Do [CE1] have [QM] values of [DP]?	Do pizzas have different values of size?	2	data.	quan.	

[12] [13] CQ Archetypes (*PQ*: Predicate Arity, *RT*= Relation Type, *M*= Modifier, *DE*=Domain-independent Element; *obj.* & *data.* = object & data property relation resp., *num.* = numeric modifier, *quan.* = quantitative modifier, *term.* = temporal element, *spa.* = spatial element; *CE* = class expression, *OPE* = object property expression, *DP* = datatype property, *I* = individual, *NM* = numeric modifier, *PE*= property expression, *QM* = quantity modifier)

Ontology Design Methodologies: Diagnosis based approaches [9]

- Inconsistency or unsatisfiability ontology *defect* detection tools
- Correctness and scalability
- Diagnosis tools: ECCO[13], ORE (Ontology Repair and Enrichment)[14], inference inspector and Protégé.
- More Ontology Engineering Methodologies (see Ch. 9 [5], [18])

[13] A *diff* tool for OWL 2 `https://github.com/rsgoncalves/ecco`
[14] Allows validation of OWL KBs `aksw.org/Projects/ORE.html`

That’s a wrap

References

References i

[1] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 2019.

[2] A. Bennetot, J.-L. Laurent, R. Chatila, and N. Díaz-Rodríguez. Towards explainable neural-symbolic visual reasoning. In *NeSy Workshop IJCAI 2019, Macau, China*, 2019.

[3] M. Bienvenu. Ontology-mediated query answering: harnessing knowledge to get more from data. In *IJCAI: International Joint Conference on Artificial Intelligence*, 2016.

[4] R. J. Brachman and J. G. Schmolze. An overview of the kl-one knowledge representation system. *Cognitive Science*, 9(2):171 – 216, 1985.

[5] J. Davies, R. Studer, and P. Warren. *Semantic Web technologies: trends and research in ontology-based systems.* John Wiley & Sons, 2006.

[6] N. Díaz-Rodríguez. *Semantic and fuzzy modelling of human behaviour recognition in smart spaces. A case study on ambiental assisted living*. PhD thesis, 2016.

[7] J. M. Gomez-Perez, J. Z. Pan, G. Vetere, and H. Wu. Enterprise knowledge graph: An introduction. In *Exploiting Linked Data and Knowledge Graphs in Large Organisations*, pages 1–14. Springer, 2017.

[8] M. Minsky. A framework for representing knowledge. 1974.

[9] J. Z. Pan, D. Calvanese, T. Eiter, I. Horrocks, M. Kifer, F. Lin, and Y. Zhao. *Reasoning Web: Logical Foundation of Knowledge Graph Construction and Query Answering: 12th International Summer School 2016, Aberdeen, UK, September 5-9, 2016, Tutorial Lectures*, volume 9885. Springer, 2017.

[10] J. Z. Pan, G. Vetere, J. M. Gomez-Perez, and H. Wu. *Exploiting linked data and knowledge graphs in large organisations*. Springer, 2017.

[11] M. Poveda-Villalón, M. Suárez-Figueroa, and A. Gómez-Pérez. Validating ontologies with OOPS! In A. Teije, J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Acquin, A. Nikolov, N. Aussenac-Gilles, and N. Hernandez, editors, *Knowledge Engineering and Knowledge Management*, volume 7603 of *Lecture Notes in Computer Science*, pages 267–281. Springer Berlin Heidelberg, 2012.

[12] M. R. Quillian. Word concepts: a theory and simulation of some basic semantic capabilities. *Behavioral science*, 12 5:410–30, 1967.

[13] Y. Ren, A. Parvizi, C. Mellish, J. Z. Pan, K. van Deemter, and R. Stevens. Towards competency question-driven ontology authoring. In V. Presutti, C. d'Amato, F. Gandon, M. d'Aquin, S. Staab, and A. Tordai, editors, *The Semantic Web: Trends and Challenges*, pages 752–767, Cham, 2014. Springer International Publishing.

[14] R. Rosati. On the complexity of dealing with inconsistency in description logic ontologies. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, IJCAI'11, pages 1057–1062. AAAI Press, 2011.

[15] W. Samek, T. Wiegand, and K.-R. Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*, 2017.

[16] E. H. Shortliffe. Mycin: Computer-based medical consultations, 1976.

[17] M. C. Suárez-Figueroa. *NeOn Methodology for building ontology networks: specification, scheduling and reuse*. PhD thesis, Informatica, 2010.

[18] M. C. Suárez-Figueroa and A. Gómez-Pérez. Neon methodology for building ontology networks: a scenario-based methodology. In *Proceedings of the International Conference on Software, Services & Semantic Technologies*. Sofia, 2009.

[19] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López. The neon methodology for ontology engineering. In *Ontology engineering in a networked world*, pages 9–34. Springer, 2012.

[20] S. Wachter, B. Mittelstadt, and L. Floridi. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7(2):76–99, 2017.