Knowledge Representation for the Semantic Web
Lecture 1: Introduction

Daria Stepanova

max planck institut informatik

Max Planck Institute for Informatics
D5: Databases and Information Systems group

WS 2017/18

Overview

Organization

Content

Semantic Web

Knowledge Representation

KRSW

About me

- **Short CV:**
  - **2005-2010** Diploma in applied informatics from St. Petersburg state university
  - **2011-2015** PhD in computational logic from TU Wien
  - **2015-present** Postdoctoral researcher in D5 group of MPI
- **Research interests:**
  - Knowledge representation and reasoning
  - Semantic web
  - Inductive rule learning
- **Appointments:** by email dstepano@mpi-inf.mpg.de

Basic course info

- **Number of credits**: 6 ECTS
- **Lectures**: Thursdays 14:00-16:00 @ 014, E1.3
- **Tutorials**: In January in small groups (every student is expected to attend three 1-hour tutorials)
- **TA**: Mohamed Gad-Elrab[1]
- **Material** will be put on the course web page[2]
- **Assignments:** two theoretical and two practical assignments will have to be completed
- **Final exams:** in a written form

---

[1] http://people.mpi-inf.mpg.de/~gadelrab/

[2] https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/teaching/winter-semester-201718/knowledge-representation-for-the-semantic-web/

Evaluation

- Final number of points sums up from
  - 2 exercise sheets (max. 10 points)
  - 2 projects (max. 20 points)
  - final exam (max. 70 points)

- The grades are computed as follows:
  - ≥ 91 1
  - ≥ 81 2
  - ≥ 71 3
  - ≥ 60 4
  - < 60 5

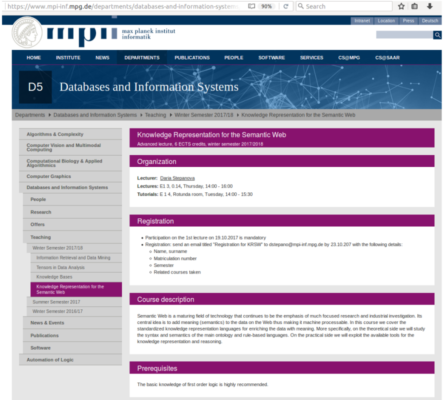
Course agenda

- Motivation
- Description logics (4 lectures)
- Answer set programming (3 lectures)
- Rule learning and other advanced topics

Course agenda

- **Motivation (today)**
  - **What is Semantic Web?**
  - **What is Knowledge Representation?**
  - **How are KR and SW connected?**
- Description logics (4 lectures)
- Answer set programming (3 lectures)
- Rule learning and other advanced topics

Syntactic Web



- Typical web page markup consists of
  - Rendering information (font size and color)
  - Hyper-links to related content
- Semantic content is accessible to humans but not machines

Current syntactic Web

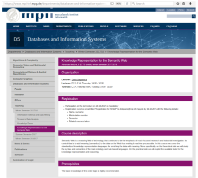
- Immensely successful
- Huge amounts of data
- Syntax standards for transfer of structured data
- Machine-processable, human-readable documents

BUT:

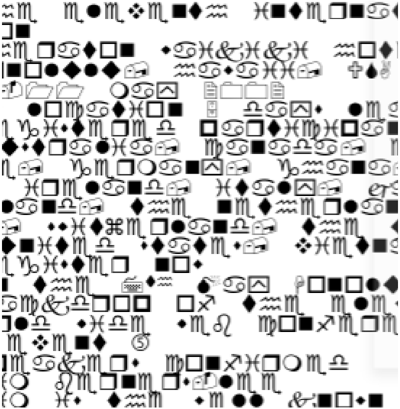
- Content/knowledge cannot be accessed by machines, i.e. machine-processable but not machine-understandable
- Meaning (semantics) of transferred data is not accessible

What can we see?

- KR for SW course is an advanced course of 6 ECTS
- In takes place on Thursdays at 14:00-16:00
- The location is 014 of E 13
- Offered by D5: Databases and Information systems
- Other courses offered by D5 in winter semester 2017/2018 are ...

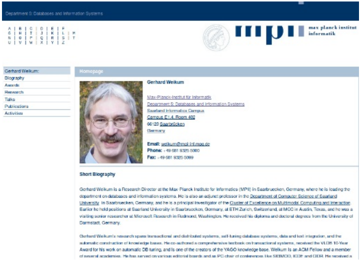


What can machines see?



WWW: humans only!

How can we answer the queries:



- Which papers has Prof. G. Weikum published in 2017?
- Which advanced lectures does the department headed by Prof. G. Weikum offer in WS 2017/2018?

Just google “Prof. G. Weikum”!

- Web page contains enough info to answer queries, **but**
  - this info is implicit
  - we understand it because we know the context
  - machines cannot make sense of it

Why Syntactic Web is not enough?

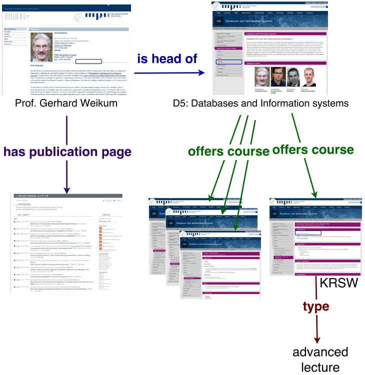
Cannot answer "knowledge queries" such as:

- Which polititians are also scientists?
- What genes are involved in signal transduction and are related to pyramidal neurons?
- What is the price, duration of warrantee, and technical features of phones that cost less than 300 Euro and are not of Apple brand?
- Which papers has Prof. G. Weikum published in 2017?
- Which advanced lectures does the department headed by Prof. G. Weikum offer in WS 2017/2018?

How can we liberate the Web data?

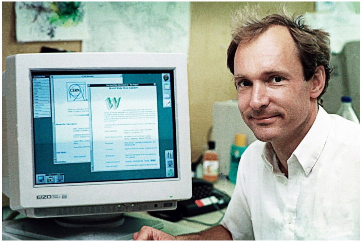
How can we answer the queries:

- Which papers has Prof. G. Weikum published in 2017?
- Which advanced lectures does the department headed by Prof. G. Weikum offer in WS 2017/2018?



- some extra information-metadata must be added to links and data
- this information links data to other data and gives meaning to it
- this information must be machine readable
- everything must be done in a standardized way

Need for semantics!



„The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation"

Tim Berners-Lee, James Hendler, Ora Lassila: [The Semantic Web](), Scientific American, 284(5), pp. 34-43(2001)

Semantic Web is ...

- the Web of Data as an upgdare of the Web of documents
- the Web as a huge decentralized database (knowledge base) of machine-processable data

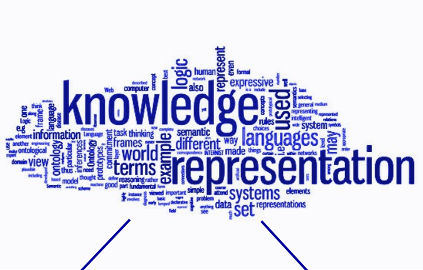
Main challenge:

How to represent knowledge and reason about it?

Knowledge representation

General goal:

develop formalisms for providing high level description of the world that can be effectively used to build intelligent applications



Cognitive approaches,
e.g., network-structures

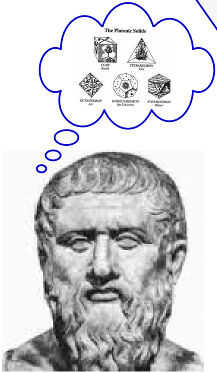
Logic-based approaches,
e.g., fragments of
First Order Logic

History of cognitive KR

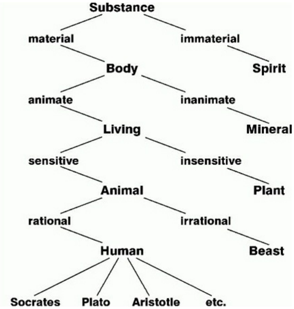
Plato: *“Knowledge is justified true belief”*

350 BC

Personal



External



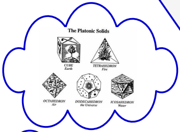
Tree of Knowledge

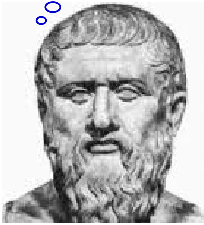
History of cognitive KR

Plato: *“Knowledge is justified true belief”*

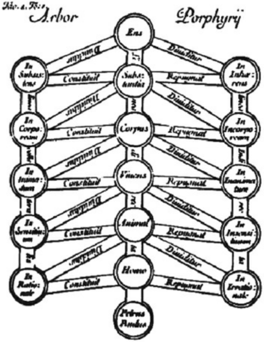
350 BC

Personal



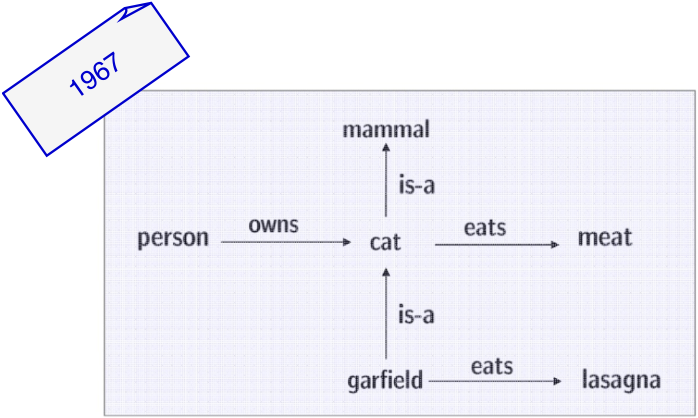


External



History of cognitive KR

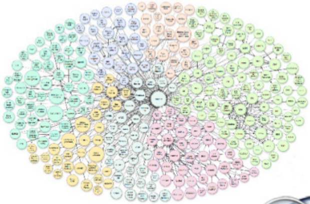
Semantic Networks introduced in [Quillan, 1967]



Modern days: Knowledge graphs

2017

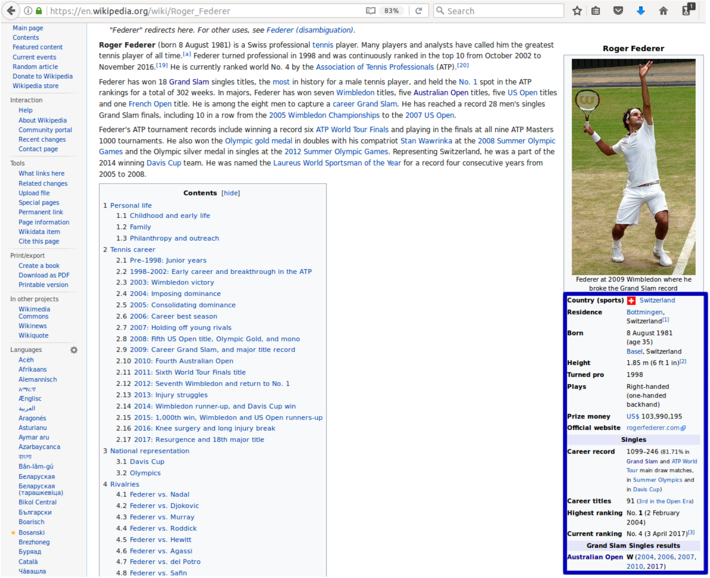
yago
select knowledge



WIKIDATA

NELL

Knowledge graphs



https://en.wikipedia.org/wiki/Roger_Federer

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

In other projects
Wikimedia Commons
Wikinews
Wikiquote

Languages
Acèh
Afrikaans
Alemannisch
Ænglisc
العربية
Aragonés
Asturianu
Aymar aru
Azərbaycanca
বাংলা
Bân-lâm-gú
Беларуская
Беларуская (тарашкевіца)
Bikol Central
Български
Boarisch
Bosanski
Brezhoneg
Буряад
Català
Чӑвашла

"Federer" redirects here. For other uses, see Federer (disambiguation).

Roger Federer (born 8 August 1981) is a Swiss professional tennis player. Many players and analysts have called him the greatest tennis player of all time.[a] Federer turned professional in 1998 and was continuously ranked in the top 10 from October 2002 to November 2016.[19] He is currently ranked world No. 4 by the Association of Tennis Professionals (ATP).[20]

Federer has won 18 Grand Slam singles titles, the most in history for a male tennis player, and held the No. 1 spot in the ATP rankings for a total of 302 weeks. In majors, Federer has won seven Wimbledon titles, five Australian Open titles, five US Open titles and one French Open title. He is among the eight men to capture a career Grand Slam. He has reached a record 28 men's singles Grand Slam finals, including 10 in a row from the 2005 Wimbledon Championships to the 2007 US Open.

Federer's ATP tournament records include winning a record six ATP World Tour Finals and playing in the finals at all nine ATP Masters 1000 tournaments. He also won the Olympic gold medal in doubles with his compatriot Stan Wawrinka at the 2008 Summer Olympic Games and the Olympic silver medal in singles at the 2012 Summer Olympic Games. Representing Switzerland, he was a part of the 2014 winning Davis Cup team. He was named the Laureus World Sportsman of the Year for a record four consecutive years from 2005 to 2008.

Contents [hide]

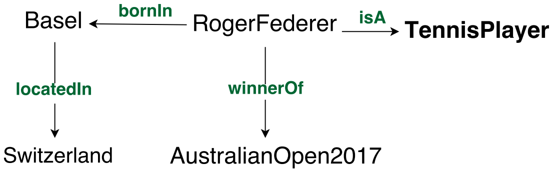
1 Personal life
1.1 Childhood and early life
1.2 Family
1.3 Philanthropy and outreach
2 Tennis career
2.1 Pre–1998: Junior years
2.2 1998–2002: Early career and breakthrough in the ATP
2.3 2003: Wimbledon victory
2.4 2004: Imposing dominance
2.5 2005: Consolidating dominance
2.6 2006: Career best season
2.7 2007: Holding off young rivals
2.8 2008: Fifth US Open title, Olympic Gold, and mono
2.9 2009: Career Grand Slam, and major title record
2.10 2010: Fourth Australian Open
2.11 2011: Sixth World Tour Finals title
2.12 2012: Seventh Wimbledon and return to No. 1
2.13 2013: Injury struggles
2.14 2014: Wimbledon runner-up, and Davis Cup win
2.15 2015: 1,000th win, Wimbledon and US Open runners-up
2.16 2016: Knee surgery and long injury break
2.17 2017: Resurgence and 18th major title
3 National representation
3.1 Davis Cup
3.2 Olympics
4 Rivalries
4.1 Federer vs. Nadal
4.2 Federer vs. Djokovic
4.3 Federer vs. Murray
4.4 Federer vs. Roddick
4.5 Federer vs. Hewitt
4.6 Federer vs. Agassi
4.7 Federer vs. del Potro
4.8 Federer vs. Safin

Roger Federer

Federer at 2009 Wimbledon where he broke the Grand Slam record

Country (sports)	Switzerland
Residence	Bottmingen, Switzerland[1]
Born	8 August 1981 (age 35) Basel, Switzerland
Height	1.85 m (6 ft 1 in)[2]
Turned pro	1998
Plays	Right-handed (one-handed backhand)
Prize money	US$ 103,990,195
Official website	rogerfederer.com
Singles	
Career record	1099–246 (81.71% in Grand Slam and ATP World Tour main draw matches, in Summer Olympics and in Davis Cup)
Career titles	91 (3rd in the Open Era)
Highest ranking	No. 1 (2 February 2004)
Current ranking	No. 4 (3 April 2017)[3]
Grand Slam Singles results	
Australian Open	W (2004, 2006, 2007, 2010, 2017)

Knowledge graphs

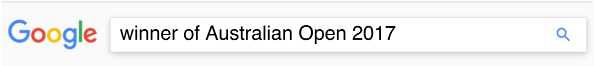


KGs are huge collections of positive unary and binary facts

tennisPlayer(rogerFederer)
bornIn(rogerFederer, basel)

Country (sports)	Switzerland
Residence	Bottmingen, Switzerland[1]
Born	8 August 1981 (age 35) Basel, Switzerland
Height	1.85 m (6 ft 1 in)[2]
Turned pro	1998
Plays	Right-handed (one-handed backhand)
Prize money	US$ 103,990,195
Official website	rogerfederer.com
Singles	
Career record	1099–246 (81.71% in Grand Slam and ATP World Tour main draw matches, in Summer Olympics and in Davis Cup)
Career titles	91 (3rd in the Open Era)
Highest ranking	No. **1** (2 February 2004)
Current ranking	No. 4 (3 April 2017)[3]
Grand Slam Singles results	
Australian Open	**W** (2004, 2006, 2007, 2010, 2017)

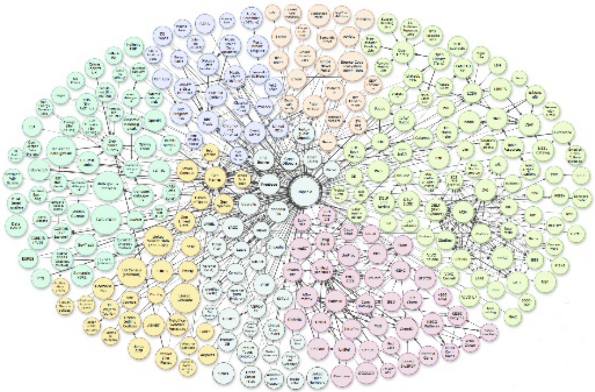
Semantic Web search today



Semantic Web search today

Google $\exists X \;\; \mathrm{winnerOf}(X, \mathrm{AustralianOpen2017})$

$\Downarrow$



Roger Federer

Tennis player

rogerfederer.com

Roger Federer is a Swiss professional tennis player who is currently ranked world No. 10 by the Association of Tennis Professionals. Many players and analysts have called him the greatest tennis player of all time. Wikipedia

Born: August 8, 1981 (age 35 years), Basel, Switzerland

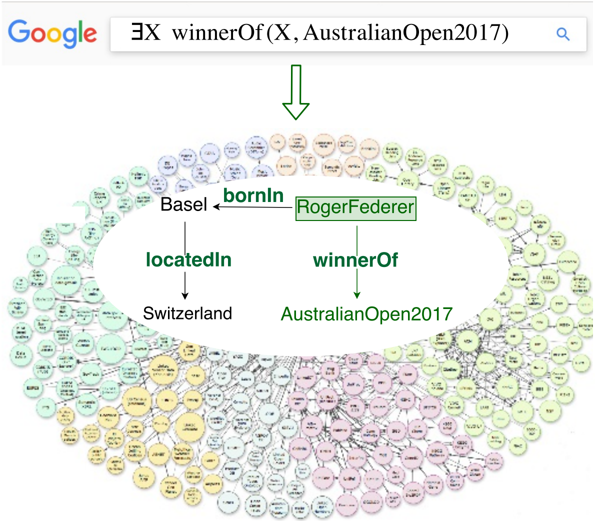
Height: 1.85 m

Weight: 85 kg

Spouse: Mirka Federer (m. 2009)

Children: Lenny Federer, Myla Rose Federer, Charlene Riva Federer, Leo Federer

Semantic Web search today



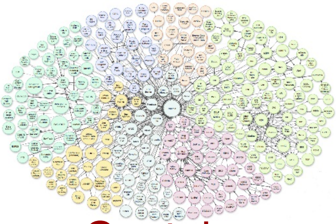
Problem: Inconsistency

Knowledge
Graph

Information
Extraction

Web
pages

Difficult

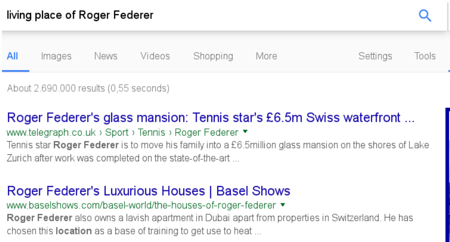


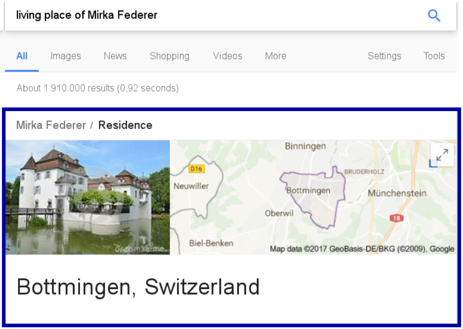
Noisy

Contains
mistakes

Problem: Incompleteness

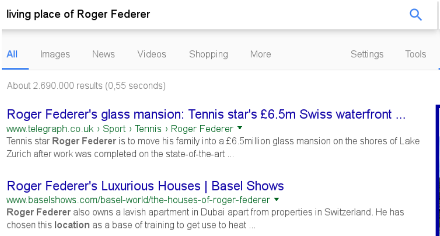
Google KG misses Roger's living place, but contains his wife's Mirka's..

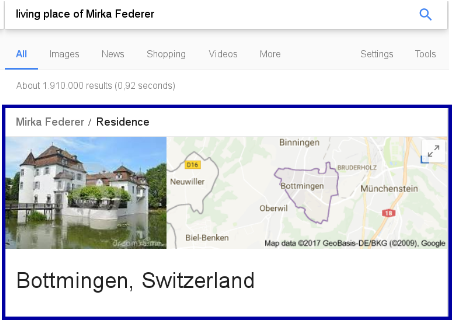




Need for logical reasoning on top of KGs

Google KG misses Roger’s **living place**, but contains his **wife’s** Mirka’s..





Need for logical reasoning on top of KGs

Google KG misses Roger's **living place**, but contains his **wife's** Mirka's..

Need for reasoning!

KG: Mirka lives in Bottmingen

KG: Roger is married to Mirka

Axiom: Married people live together

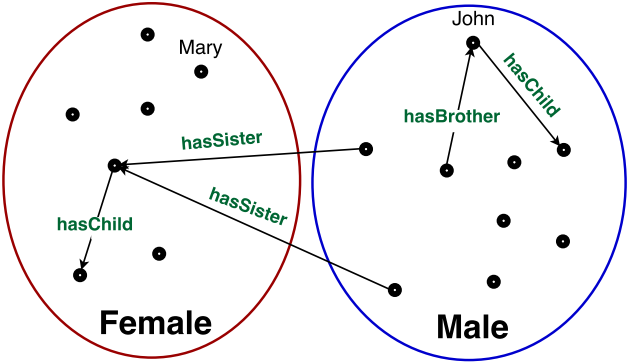
Derivation: Roger lives in Bottmingen

History of logic-based KR

- **1950's:** First Order Logic (FOL) for KR (undecidable)
  (e.g. [McCarthy, 1959])
- **1970's:** Network-shaped structures for KR (no formal semantics)
  (e.g. semantic networks [Quillan, 1967], frames [Minsky, 1985])
- **1979:** Encoding of network-shaped structures into FOL [Hayes, 1979]
- **1980's:** Description Logics (DL) for KR
  - Decidable fragments of FOL
  - Theories encoded in DLs are called ontologies
  - Many DLs with different expressiveness and computational features
  - Particularly suited for conceptual reasoning

Description logic ontologies

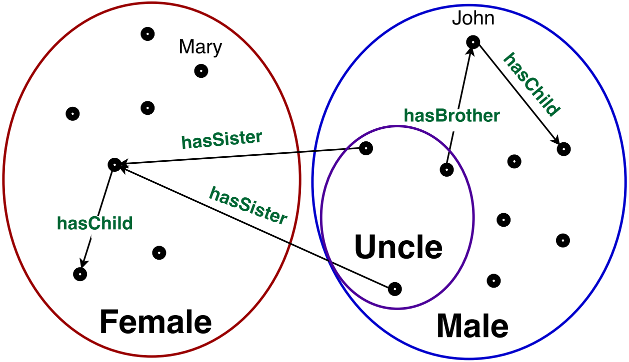
Open World Assumption (OWA): what is not derived is unknown



Inclusions: $Female \sqsubseteq \neg Male, hasSister \sqsubseteq hasSibling, hasBrother \sqsubseteq hasSibling$

Description logic ontologies

Open World Assumption (OWA): what is not derived is unknown



Inclusions: $Female \sqsubseteq \neg Male, hasSister \sqsubseteq hasSibling, hasBrother \sqsubseteq hasSibling$

Complex axioms: $Uncle \equiv Male \sqcap \exists hasSibling.\exists hasChild$

What can not be said in DLs?

- Exceptions from theories (due to monotonicity)

What can not be said in DLs?

- Exceptions from theories (due to monotonicity)

$WithBeard \sqsubseteq Male$ — *People with beards are male*

$Female \sqsubseteq \neg Male$ — *Female are not male*

$WithBeard(c)$ — *C has a beard*

What can not be said in DLs?

- Exceptions from theories (due to monotonicity)

$\mathit{WithBeard} \sqsubseteq \mathit{Male}$	*People with beards are male*
$\mathit{Female} \sqsubseteq \neg\mathit{Male}$	*Female are not male*
$\mathit{WithBeard}(c)$	*C has a beard*
$\mathit{Male}(c)$	*C is male*

What can not be said in DLs?

- Exceptions from theories (due to monotonicity)

$WithBeard \sqsubseteq Male$	*People with beards are male*
$Female \sqsubseteq \neg Male$	*Female are not male*
$WithBeard(c)$	*C has a beard*
$Female(c)$	*C is female*
$Male(c)$	*C is male*
$\neg Male(c)$	*C is not male*

Monotonicity: the more we add, the more we get!

History of logic-based KR

- **1970's:** Logic programming
  (e.g. Prolog)
- **1980's:** Nonmonotonic logics
  (e.g. circumscription [McCarthy, 1980], default logic [Reiter, 1980])
- **1988:** Nonmonotonic rules under answer set semantics (ASP) [Gelfond and Lifschitz, 1988]
  - Logic programs with model-based semantics
  - Disjunctive datalog with default negation *not*

Not is not ¬!

Default negation *not*

At a rail road crossing cross the road if **no train is known** to approach

$walk \leftarrow at(X), crossing(X), \textbf{not}\ train\_approaches(X)$

Classical negation ¬

At a rail road crossing cross the road if **no train** approaches

$walk \leftarrow at(X), crossing(X), \neg train\_approaches(X)$

Nonmonotonic rules

Closed World Assumption (CWA): what is not derived is false

Rule: $\underbrace{a_1 \vee \ldots \vee a_k}_{\text{head}} \leftarrow \underbrace{b_1, \ldots, b_m, \mathit{not}\, b_{m+1}, \ldots, \mathit{not}\, b_n}_{\text{body}}$

Informal semantics: If $b_1, \ldots, b_m$ are true and none of $b_{m+1}, \ldots, b_n$ is known, then at least one among $a_1, \ldots, a_k$ must be true

Default negation: unless a child is adopted one of his parents must be female

$$female(Y) \vee female(Z) \leftarrow hasParent(X, Y), hasParent(X, Z), Y \neq Z, not\ adopted(X)$$

Constraint: ensure that a person cannot be parent of himself.

$$\perp \leftarrow parent(X, X).$$

Answer set programs

Evaluation of ASP programs is model-based

Answer set program (ASP) is a set of nonmonotonic rules

(1) $hasParent(john, pat)$ (2) $hasParent(john, alex)$ (3) $male(alex)$

$$(4)\ female(Y) \leftarrow hasParent(X, Y), hasParent(X, Z), Y \neq Z, male(Z), not\ \ adopted(X)$$

Answer set programs

Evaluation of ASP programs is model-based

1. Grounding: substitute all variables with constants in all possible ways

Answer set program (ASP) is a set of nonmonotonic rules

(1) $hasParent(john, pat)$ (2) $hasParent(john, alex)$ (3) $male(alex)$

$$(4)\ female(Y) \leftarrow hasParent(X, Y), hasParent(X, Z), Y \neq Z, male(Z), not\ \ adopted(X)$$

Answer set programs

Evaluation of ASP programs is model-based

1. Grounding: substitute all variables with constants in all possible ways

Answer set program (ASP) is a set of nonmonotonic rules

(1) $hasParent(john, pat)$ (2) $hasParent(john, alex)$ (3) $male(alex)$

(4) $female(pat) \leftarrow hasParent(john, pat), hasParent(john, alex),$
$male(alex), not\ \ adopted(john)$

Answer set programs

Evaluation of ASP programs is model-based

1. Grounding: substitute all variables with constants in all possible ways
2. Solving: compute a minimal model (answer set) I satisfying all rules

Answer set program (ASP) is a set of nonmonotonic rules

$$(1)\ hasParent(john, pat) \quad (2)\ hasParent(john, alex) \quad (3)\ male(alex)$$

$$(4)\ female(pat) \leftarrow hasParent(john, pat), hasParent(john, alex), male(alex), not\ \ adopted(john)$$

$I = \{hasParent(john, pat), hasParent(john, alex), male(alex), female(pat)\}$

CWA: $adopted(john)$ can not be derived, thus it is false

Answer set programs

Evaluation of ASP programs is model-based

1. Grounding: substitute all variables with constants in all possible ways
2. Solving: compute a minimal model (answer set) I satisfying all rules

Answer set program (ASP) is a set of nonmonotonic rules

$$(1)\ hasParent(john, pat) \quad (2)\ hasParent(john, alex) \quad (3)\ male(alex)$$

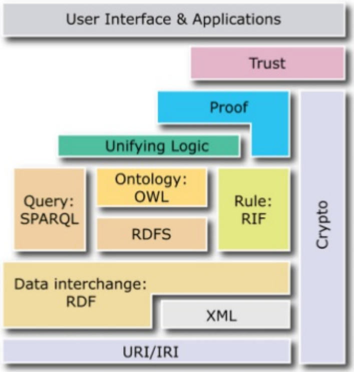
$$(4)\ female(pat) \leftarrow hasParent(john, pat), hasParent(john, alex),$$
$$male(alex), not\ \ adopted(john)$$

$$(5)\ adopted(john)$$

$$I = \{hasParent(john, pat), hasParent(john, alex), male(alex), \overset{adopted(john)}{\cancel{female(pat)}}\}$$

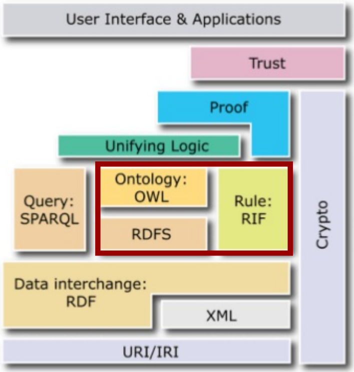
Nonmonotonicity: adding facts might lead to loss of consequences!

Knowledge representation standards in SW context



1994	First public presentation of the Semantic Web idea
1998	Start of standardization of data model (RDF) and a first ontology languages (RDFS) at W3C
2000	Start of large research projects about ontologies in the US and Europe (DAML & Ontoknowledge)
2002	Start of standardization of a new ontology language (OWL) based on research results
2004	Finalization of the standard for data (RDF) and ontology (OWL)
2008	Standardization of a query language (SPARQL)
2009	Extension of OWL to OWL 2.0
2010	Standard Rule Interchange Format (RIF)

Knowledge representation standards in SW context



1994	First public presentation of the Semantic Web idea
1998	Start of standardization of data model (RDF) and a first ontology languages (RDFS) at W3C
2000	Start of large research projects about ontologies in the US and Europe (DAML & Ontoknowledge)
2002	Start of standardization of a new ontology language (OWL) based on research results
2004	Finalization of the standard for data (RDF) and ontology (OWL)
2008	Standardization of a query language (SPARQL)
2009	Extension of OWL to OWL 2.0
2010	Standard Rule Interchange Format (RIF)

Course agenda

1. **Description Logic ontologies (DL)**
   - Theoretical background
   - Ontology Web Language (OWL)
   - Tools and applications
2. **Answer Set Programming rules (ASP)**
   - Theoretical background
   - Answer set programming semantics
   - Tools and applications
3. **Learning rules from data and other topics**
   - Relational association rule learning
   - Learning rules with exceptions under incompleteness
   - Other advanced topics

References I

Michael Gelfond and Vladimir Lifschitz.
The stable model semantics for logic programming.
In *Proceedings of the 5th International Conference and Symposium on Logoc Programming, ICLP 1988*, pages 1070–1080. The MIT Press, 1988.

P. J. Hayes.
The logic of frames.
In *Frame Conceptions and Text Understanding*, pages 46–61. 1979.

John McCarthy.
Programs with common sense.
In *TeddingtonConference*, pages 75–91, 1959.

John McCarthy.
Circumscription - A form of non-monotonic reasoning.
Artif. Intell., 13(1-2):27–39, 1980.

Marvin Minsky.
A framework for representing knowledge.
In *Readings in Knowledge Representation*, pages 245–262. Kaufmann, 1985.

M. Ross Quillan.
Word concepts: A theory and simulation of some basic capabilities.
Behavioral Science, pages 410–430, 1967.

Raymond Reiter.
A logic for default reasoning.
Artif. Intell., 13(1-2):81–132, 1980.