Iliana M. Petrova
Tutor: Prof. Piero A. Bonatti
XXIX Cycle - III year presentation

Design, implementation and optimisation of non-standard reasoning services

Università degli Studi di Napoli Federico II

Background

- MSc in Computer Science from Università degli Studi di Napoli Federico II
- I have been granted a fellowship sponsored by ESF Regional Operational Programme 2014 - 2020.
- Part of the KRSP group.
- Collaborate with Prof. Dr. Piero A. Bonatti, Dr. Luigi Sauro and Prof. Dr. Marco Faella

Credits summery

- Third year credits

	Year 1		Year 2		Year 3								Total	Check
						1	2	3	4	5	6			
	Estimated	Summary	Estimated	Summary	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary		
Modules	20	17	13	7	6			6				6	30	30-70
Seminars	5	7	5	8	0		4/5	3/5				1,4	16,4	10-30
Research	35	39	42	45	54	10	9	4	10	10	10	53	137	80-140
	60	63	60	60	60	10	9,8	10,6	10	10	10	60,4	183	180

- Experience abroad
  - March to August 2016 – visiting student under the supervision of Dr. Yevgeny Kazakov at the Institute of Artificial Intelligence, University of Ulm, Germany

The Semantic Web

- It has been 15 years since the first publications proposed the use of **ontologies** as a basis for defining resource content on the Web creating opportunities in using **Description Logics** in a practical setting.

- Description Logics are equipped with a notion of ***entailment*** that can be computed by ***reasoning engines***
  - Thereby "squeezing" potentially unbounded amounts of implicit knowledge from the given ontologies

Ontology

- Specifies knowledge in a concrete domain in terms of symbols that represent concepts and their relations.
- Simple ontology and inference example

The cerebral cortex is part of the brain.

$$CerebralCortex \sqsubseteq \exists isPartOf.Brain$$

Epilepsy is a disease that affects the cerebral cortex.

$$Epilepsy \sqsubseteq Disease \sqcap \exists affects.CerebralCortex$$

A brain disease is a disease that affects a part of the brain.

$$BrainDisease \equiv Disease \sqcap \exists affects.\exists isPartOf.Brain$$

Epilepsy is a brain disease.

$$Epilepsy \sqsubseteq BrainDisease$$

- A **TBox** describes general knowledge about a domain: finite set of inclusions.
- An **ABox** contains facts about specific individuals: finite set of assertions:

$$Disease(d) \qquad isPartOf(a,b)$$

The Semantic Web cont'd

- Increasingly growing need in the knowledge representation field to extend ontology languages and reasoning engines *with non-standard characteristics*.
- Focus on two concrete applications
  - Non monotonic semantics for description logics
  - Confidentiality model for ontologies
- Goal: Brings technology closer to practical execution of non-standard reasoning services on very large knowledge bases.

Nonmonotonic Description Logics (DLs): Motivation

Why should we adopt a nonmonotonic semantics?

- Biomedical ontologies are rich of exceptions to be represented.

A well-known example of prototypical property in biomedical domain:

"Mammalian red blood cells are an ***exceptional*** *class of eukariotic cells:* $MamRedBldCell \sqsubseteq EukCell$

the latter have a nucleus, $EukCell \sqsubseteq_n \exists has\_nucleus$

while the former, in their mature stage, do not have a nucleus." $MamRedBldCell \sqsubseteq \neg\exists has\_nucleus$

Infer that:

✓ normal eukariotic cells have a nucleus

✓ normal mammalian red blood cells do not have a nucleus

Nonmonotonic Description Logics (DLs) : Motivation

Why should we adopt a nonmonotonic semantics?

- Semantic web policy formulation (formalization of standard default policies and authorizations inheritance with exceptions

(1) In general, users cannot access confidential files.

(2) Staff can read confidential files.

(3) Blacklisted users are not granted any access.

This directive cannot be overridden.

$$Staff \sqsubseteq User$$

$$Blklst \sqsubseteq User$$

$$\exists subj.User \sqcap \exists target.Confidential \sqsubseteq_n \neg \exists privilege \quad (1)$$

$$\exists subj.Staff \sqcap \exists target.Confidential \sqsubseteq_n \exists privilege.Read \quad (2)$$

$$\exists subj.Blklst \sqsubseteq \neg \exists privilege \quad (3)$$

Infer that:

- ✓ Normally, read operations on confidential files are permitted if the request comes from staff member
- ✓ Blacklisted users cannot do anything

DLN: new family of nonmonotonic DLs

- Extension of the classical Description Logics designed to address real-world problems and concrete knowledge engineering needs.

- First formalism to provide practical support of nonmonotonic inferences by modelling priority-based overriding.

- Key idea: Design knowledge bases by describing prototypical instances whose general properties can be refined later, by adding suitable exceptions.
  - A nonmonotonic axiom (defeasible inclusion) is an expression

$$C \sqsubseteq_n D$$

    - intended meaning "*by default, all instances that satisfy* C *satisfy also* D*, unless stated otherwise*".
    - higher priority axioms may ***ovveride*** it (contradict the implication).

- Automated reasoning in the new logic is carried out by means of polynomial reduction to classical description logics.

Confidentiality model for ontologies: Motivation

Why knowledge confidentiality matters?

- ➢ Private knowledge of companies and public organizations encoded with OWL and Linked Open Data
- ➢ Sensitive linked open government data (e.g. health)
- ➢ Medical records annotated with SNOMED

IHTSDO delivering SNOMED CT® the global clinical terminology

HL7

LOINC® Logical Observation Identifiers Names and Codes

- Semantic Web techniques help in linking different knowledge sources and extract implicit information -- increasing security and privacy risk

Confidentiality model for ontologies: Motivation

- A very simple attack employing background knowledge

One secret: $s = OncologyPatient(John)$

Ontology is:

$$SSN(John, 12345),\ SSN(usr1, 12345), OncologyPatient(usr1)$$

$KB \not\models OncologyPatient(John) \Rightarrow$ the ontology can be published

However, it is common knowledge that SSN is a key

So the user can infer

$$John = usr1$$

- Other types of attacks exploit metaknowledge

Confidentiality model for ontologies

- A confidentiality model based on a fully generic formalization of the user's background knowledge, and the definition of a method for computing secure knowledge views.

- To compute views in practice we adopt a safe, generic method for approximating background knowledge, together with a specific rule-based language for expressing metaknowledge

  – The metaknowledge is encoded by *metarules* :

$$\alpha_1, \ldots, \alpha_n \Rightarrow \beta_1 \mid \ldots \mid \beta_m \qquad (n \geq 0, m \geq 0)$$

where all α_i and β_i are DL axioms

Research activity

- Need to deal with very large knowledge bases
- Identifing optimisations is mandatory in order to provide a practical and scalable support of these kind of applications
- Goal: Brings technology closer to practical execution of non-standard reasoning services on large knowledge bases.

Module Extraction Techniques

- Optimization technique for querying large ontologies: many of the axioms in a large KB are expected to be irrelevant to the given query. Module *extractors can be used to focus reasoning on relevant axioms only.*

- **Parallel module extraction:** The performance can be improved by evaluating the axioms' relevance in parallel, taking advantage of the multiprocessor architectures.
  - ✓ speed-up: up to 50% for large ontologies.

Iterated Module Extraction for DLN Reasoning

- The approach is not trivial: off the shelf *module extractors are unsound for most nonmonotonic logics*

- Module extraction is idempotent for classical DLs, but not for the DLN family of nonmonotonic logics

- Basic idea: iterating module extraction makes it possible to progressively discarding more axioms that turn out to be irrelevant to the given query

- The novel iterated module extraction algorithm requires an articulated correctness proof.

  ✓ Speed-up : make DLN reasoning *at least one order of magnitude faster* (and *up to ~780 times faster* in some case)

Module Extraction for ABoxes

- The existing *module extractors* are not very effective in the presence of non empty ABoxes.
  - phenomenon is amplified in DLN, where reasoning requires repeated incremental classifications of the knowledge base.
- A new module extraction algorithm that *discards significantly more axioms in the presence of non empty Aboxes*
- Proved to be *correct under the assumption that the knowledge base is consistent* (hypothesis compatible with some of the main intended uses of module extraction).
- The conditional module extractor for nonempty ABoxes is very effective when the ABox assertions are loosely interconnected, with speedups up to 75%.

DLN Optimistic computation

- DLN's reasoning method (and also Secure view generation) requires multiple consistency checks over a sequence of knowledge bases that share a possibly large common part
  - Need of incremental reasoning mechanisms that help by avoiding recomputations of the same consequences.
- Assertion of new axioms is processed very efficiently, while computational cost of axiom deletion is generally not negligible
- The optimistic reasoning method is expected to reduce the number of deletions.
- The speedup factor is about two.

Evaluation of DLN Reasoning Engine

- Optimisations incorporated in **NMReasoner**, a nonmonotonic reasoner that implements the DLN framework

- Currently no "*real*" KBs encoded in a nonmonotonic DL exist, as standard DL technology does not support nonmonotonic reasoning.

- Scalability tests have been carried out on synthetic test cases automatically generated in a principled way.
  - The test suites obtained by modifying large biomedical ontologies (*GALEN, Gene Ontology, Fly Anatomy*) and proved to be nontrivial w.r.t. a number of structural parameters.

Optimizing Secure ontology view construction (I)

SOVGen is a tractable implementation of the framework.

- Module extractors used on the background knowledge bases (such as SNOMED) in order to make reasoning focus on relevant knowledge only.
  - *Modules extracted are on average two or three orders of magnitude smaller* which drastically improves performance.

- ❑ Given the amount of background knowledge (consider that SNOMED-CT describes about 300K concepts) the use of *module extraction techniques improves the computation time of two-three orders of magnitude* at a cost of about 30 sec of overhead.

Optimizing Secure ontology view construction (II)

- Evaluation of metarules requires techniques for effective query evaluation.
  - Two different implementations *Apache Jena – ARQ, OWL-BGP* available
  - ❖ *Metarule Evaluation Engine (MEE)*: ad hoc module designed to take advantage of the specific nature of Horn metarules which extensively exploit incremental reasoning, short-circuit evaluation and memoization techniques.

- The experimental results obtained by using MEE (resp. OWL-BGP) to evaluate metarules *show MEE is 1-2 orders of magnitude faster*
  - for Jena-ARQ all the test cases exceeded 1 hour time-out

Evaluation of the secure ontology view construction

- Scalability experimental evaluations carried out on synthetic test cases specifically designed to simulate the employment of the confidentiality model in a e-health scenario as part of the SmartHealt 2.0 Project.

- Each test case represents the encoding of sensitive data in a CDA-compliant electronic health record. Clinical Document Architecture (CDA) is an iternational standard for information exchange, based on the Health Level 7 Reference Information Model.

- According to the CDA standards a disease is represented by a ICD-9CM code, pharmaceutical products and procedures by SNOMED CT codes, while diagnostic tests and laboratory data by LOINC codes.

Summing up ..

✓ Nonmonotonic Description Logics (DLs)

- ✧ All optimizations are sound and complete.
- ✧ For the first time response times compatible with real-time reasoning are obtained with nonmonotonic KBs of this size (more than 20K concept names and over 30K inclusions).

✓ Confidentiality model for ontologies

- ✧ Managed to construct secure views in presence of background knowledge bases with more than 300K axioms within a minute - a one-time cost before the secure view is published so no overhead is placed on user queries.
- ✧ Module extraction techniques and a suitable, ad-hoc metarule evaluation engine - which intensively exploit incremental reasoning - largely outperform general conjunctive query evaluation engines.

Summing up ..

- Work in progress
  - Experimental evaluation of an inference engine that allows incremental retraction in both application domains.
  - Further extension of the synthetic test suites as part of the preparation of a journal paper describing the advanced optimisations for DLN

- What's to come
  - Study different parallelization strategies, based on suitable reorderings of the operations performed during the translation of DLN in the corrisponding classical DL.
  - Further possible optimizations for DLN , such as caching the translations used for previous queries..
  - Extend SOVGen for general metarules…

Products

- **International journal papers**
  "A new semantics for overriding in description logics", P.A. Bonatti, M. Faella, I. M. Petrova and L. Sauro. Artificial Intelligence Journal, 222:1–48, 2015. Available online. Ranked Q1 on ISI and Scopus.

- **International conference papers**
  "Optimizing the computation of overriding", P.A. Bonatti, I.M. Petrova, L. Sauro: In M. Arenas et all. (eds.): Proceedings of the 14th International Semantic Web Conference (ISWC-14). LNCS 9366, 356-372. Springer 2015. Ranked A+ on the conference classification by GRIN-GII.

  "Optimized Construction of Secure Knowledge-Base Views", P.A. Bonatti, I.M. Petrova, L. Sauro: In Proceedings of the 28th International Workshop on Description Logics. CEUR Workshop Proceedings 1350, CEUR-WS.org

- **Conference papers**
  "A mechanism for ontology confidentiality", P. A. Bonatti, I. M. Petrova and L. Sauro, Proceedings of the 29th Italian Conference on Computational Logic, volume 1195 of CEUR-WS.org

- **Tecnical reports**
  "Optimizing the computation of overriding", P. A. Bonatti, I. M. Petrova, L. Sauro, CoRR abs/1507.04630 (2015)

Thank you
for your attention!

Any questions?