PhD in Information Technology and Electrical Engineering

Università degli Studi di Napoli Federico II

PhD Student: Iliana Mineva Petrova

XXIX Cycle

Training and Research Activities Report – Second Year

Tutor: Piero Andrea Bonatti

UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II

1. Information

Iliana Mineva Petrova, MS in Computer Science (Information Systems) – Università degli Studi di Napoli Federico II.

XXIX Cycle - Information Technology and Electrical Engineering - Università degli Studi di Napoli Federico II.

Three-year fellowship grant by Università degli studi di Napoli Federico II.

Tutor: Prof. Dr. Piero Andrea Bonatti.

2. Study and Training activities

A. Courses

"Designing and writing scientific manuscripts for publication in English language scholarly journals, and related topics", ad hoc module, held by Prof. Barnett Parker, 3 CFU.

"Semantic web reasoners: struttura, uso e ottimizzazioni", ad hoc module, held by Prof. Piero Bonatti, 5 CFU, final examination expected to take place in march.

B. Seminars

"Answering queries over inconsistent databases", Dr. Gaelle Fontaine, 18 April 2015, 10.30 - 12.30, University of Chile, Santiago, Chile.

"Colloquium on Robotics Six Keynote Talks by International Experts", Prof. O. Khatib, Prof. T. Asfour, Prof. R. Lumia, Prof. G. Indiveri, Prof. K. Kyriakopoulos, Prof. R. Madhavan, 21 April 2015.

"Memory technologies for Android based systems", Prof. Simon Pietro Romano, 10 November 2015

"Real-Time Embedded Control Systems", Prof. Giorgio Buttazzo, 16 November 2015, 1.2 credits.

"Test and Diagnosis of Integrated Circuits", Prof. Alberto Bosio, 17 – 18 November 2015, 2.4 credits.

"Hardware Security and Trust", Prof. Giorgio Di Natale, 19 – 20 November 2015, 2.4 credits.

"Armi autonome: problemi etici e decisioni politiche", by Prof. Giuglielmo Tamburrini, 1 December 2015

C. External courses

11th Summer School on Ontology Engineering and the Semantic Web (SSSW'15)

5 - 11 July, Bertinoro, Italy, 2 CFU

The 11th Reasoning Web Summer School (RW 2015)

31 July - 4 August, Berlin, Germany, 2 CFU

D. Summary of training activities for the second year

Student: Iliana Mineva Petrova ilianamineva.petrova@unina.it

Tutor: Piero Andrea Bonatti pieroandrea.bonatti@unina.it

Cycle XXIX

	Credits year 1								Credits year 2								Credits year 3									
	Estimated	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	Summary	Estimated	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	Summary	Estimated	1 bimonth	2 bimonth	3 bimonth	4 bimonth	5 bimonth	6 bimonth	Summary	Total	Check
Modules	20			3		6	8	**17**	13		3	4				**7**	6							**0**	**24**	30-70
Seminars	5				2 2/5	2 9/10	1 7/10	**7**	5	1 2/5				6 3/5		**8**	0							**0**	**15**	10-30
Research	35	10	10	7	7	2	3	**39**	42	9	7	5	10	4	10	**45**	54							**0**	**84**	80-140
	60	10	10	10	9,4	10,9	12,7	**63**	60	10,4	10	9	10	10,6	10	**60**	60	0	0	0	0	0	0	**0**	**123**	180

Year	Lecture/Activity	Type	Credits
1	First EATCS Young Researchers School	International Summer School	3
1	Fractional Programming for Energy Efficie	Seminar	2/5
1	Fractional Programming for Energy Efficie	Seminar	2/5
1	Nano-carbon based components and mat	Seminar	4/5
1	Developmental Robotics: From Babies to	Seminar	2/5
1	Quantum teleportation	Seminar	2/5
1	EuroProgettazione	Ad hoc module	3
1	IBM-Bluemix Corso tecnico sulla nuova p	Ad hoc module	3
1	Reliability and Availability Modeling in Pra	Seminar	3/5
1	Capacity Planning for Infrastracture-as-a-	Seminar	2/5
1	Heterogeneities in temporal networks em	Seminar	1/5
1	Methods and tools for smart device integr	Seminar	2/5
1	UML Profiles for the specification of non f	Seminar	1/2
1	Site Reliability Engineering at Google	Seminar	2/5
1	Verification of Mobile Agents in Partially K	Seminar	2/5
1	Machine Learning e applicazioni mod A: M	MS Module	6
1	Three core issues for the Internet: things,	Ad hoc Module	2
1	Applications for software development: ty	Seminar	2/5
1	Risk management meets model checking	Seminar	2/5
1	Workshop "Efficient service distribution in	Seminar	9/10
2	Answering queries over inconsistent data	Seminar	2/5
2	Colloquium on Robotics Six Keynote Talk	Seminar	1
2	Designing and writing scientific manuscrip	Ad hoc Module	3
2	11th Summer School on Ontology Engine	International Summer School	2
2	The 11th Reasoning Web Summer Schoc	International Summer School	2
2	Memory technologies for Android based s	Seminar	2/5
2	Real-Time Embedded Control Systems	Seminar	1 1/5
2	Test and Diagnosis of Integrated Circuits	Seminar	2 2/5
2	Test and Diagnosis of Integrated Circuits	Seminar	2 2/5
2	Test and Diagnosis of Integrated Circuits	Seminar	1/5

3. Research activity

Design, implementation and optimization of non-standard reasoning services

Nowadays, within the field of knowledge representation there is the increasingly growing need to extend the ontology languages and reasoning engines with non-standard characteristics. Here we focus on two concrete application contexts - Nonmonotonic Description Logics (DLs) and Confidentiality Models for Ontologies. Several authors advocated nonmonotonic inference as useful means to address real-world problems in important fields such as biomedical domains and access control policies. On the other hand, OWL and Linked Open Data have been increasingly used to encode sensitive knowledge on individuals (e.g. health, companies and public organizations. As Semantic Web techniques make it possible to extract implicit information, any access control method that does not deal with inference fails to ensure privacy.

In both cases the available classical tools are just not sufficient to offer the required non-standard reasoning services. A considerable additional effort reveals mandatory in order to provide a practical and scalable support of these kind of applications. As a consequence my research activity is currently focused on the design, implementation and optimization of the above mentioned frameworks. We adopt a suitable, easy to grasp, representation of the new constructs introduced by the specific context (non monotonic axioms in the first and background metarules in the second case). Due to the very large size of knowledge bases (KBs) that both applications need to deal with, general purpose optimization techniques such as incremental reasoning and *suitably adapted* module extraction proves to improve performance drastically. Effective conjunctive query answering techniques are as well required for the efficient implementation of the second framework. When it comes to a standard reasoning tasks these applications rely on publicly available classical reasoners.

- **Nonmonotonic Description Logics (DLs)**

  As far as the first context is concerned, my working group contribute to the practical support of nonmonotonic inferences in DLs by introducing a new semantics (cf. [1]) expressly designed to model priority-based overriding. This formalism makes it possible to design knowledge bases by describing prototypical instances whose general properties can be refined later, by adding suitable exceptions. The new family of nonmonotonic description logics is called DL^N. Automated reasoning in the new logic is carried out by means of polynomial reduction to classical DLs.

  The first experimental scalability tests on a semi-naive implementation (cf. [1], relying only on the optimization techniques of the underlying classical reasoner), called NMReasoner,

revealed promising. The result graphs showed that nonmonotonic constructs cause an approximately linear increase in reasoning time.
More recently we studied optimization techniques ([3, 4]) based on module extraction and optimistic computation techniques for reasoning in DL^N :

I. **Module Extraction Techniques**

Module extraction is an optimization technique for querying large ontologies. In practice many of the axioms in a large KB are expected to be irrelevant to a given query. Module extractors can be used to focus reasoning on relevant axioms only. More precisely we have introduced an iterated module extraction procedure for DL^N reasoning. It is worth noting that the approach is not trivial (module extractors are unsound for most nonmonotonic logics, including default logics and circumscription) and requires an articulated correctness proof. Furthermore, a module extraction algorithm for ABoxes has been proposed and has been proved to be correct under the assumption of consistency of the knowledge bases (this hypothesis, in practice, is compatible with some of the main intended uses of module extraction). The commonly employed bottom-module extractors are not very effective in the presence of non empty ABoxes. This phenomenon is even amplified in DL^N, where reasoning requires repeated incremental classifications of the knowledge base. The new module extraction algorithm is proved to discards significantly more axioms in the presence of non empty ABoxes. Finally we developed a parallel module extraction procedure to further improve the performance. It is designed to take advantage of the multiprocessor architectures by evaluating the relevance of axioms in parallel.

II. **Optimistic computation (OPT)**

The optimistic computation algorithm for query answering that we introduce is expected to exploit incremental reasoners at their best. Incremental reasoning is crucial as DLN's reasoning method iterates consistency tests on a set of KBs with large intersections. While the assertion of new axioms is processed very efficiently, the computational cost of axiom deletion is generally not negligible. The *optimistic reasoning method is thus expected to reduce the number of deletions*.

Currently no "*real*" KBs encoded in a nonmonotonic DLs exist, as standard DL technology does not support nonmonotonic reasoning. Our scalability tests have been carried out on synthetic test cases automatically generated in a principled way. The test suites have been obtained by modifying large biomedical ontologies (*GALEN, Gene Ontology, Fly Anatomy*) and proved to be nontrivial w.r.t. a number of structural parameters.

All optimizations have been integrated in NMReasoner and validated experimentally on large synthetic KBs with more than 30K axioms. The experimental results show that the iterated module-based method make DLN reasoning at least one order of magnitude faster (and up to ~780 times faster in some case); the OPT speedup factor is about two. The conditional module extractor for nonempty ABoxes is very effective when the ABox assertions are loosely interconnected, with speedups up to 75%.

To briefly summarise, all optimizations introduced are sound and complete. To the best of our knowledge, 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

The main contribution to the area of knowledge base confidentiality of my work group consists in the introduction of a knowledge base 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 that generalizes some previous approaches. In order to compute secure views in practice, we adopt a safe, generic method for approximating background knowledge, together with a specific rule-based language for expressing metaknowledge.

We further provide a tractable implementation of the new framework (cf. [2]), a secure view generator called SOVGen. The generator exploits module extractors on the background knowledge bases (such as SNOMED) in order to make reasoning focus on relevant knowledge only. Experimental analysis show that the extracted modules result on average two or three orders of magnitude smaller which drastically improves performance.
Evaluation of metarules requires techniques for effective query evaluation. We provide three different implementations based on *Apache Jena-ARQ*, *OWL-BGP* and ad hoc module, called Metarule Evaluation Engine (MEE), designed to take advantage of the specific nature of Horn metarules. MEE extensively exploit incremental reasoning, short-circuit evaluation and memoization techniques.

.Synthetic test cases have been 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[1]-compliant electronic health record. 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.

[1] Clinical Document Architecture (CDA) is an international standard for of electronic health information exchange, based on the Health Level 7 Reference Information Model.

Scalability experimental evaluations carried out on the synthetic test cases show that the use of *module extraction techniques improves the computation time of two-three orders of magnitude* at an acceptable cost of about 30 sec of overhead given the amount of background knowledge (consider that SNOMED-CT describes about 300K concepts). 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).

Summing up, the experimental analysis confirm that module extraction techniques and a suitable, ad-hoc metarule evaluation engine – which intensively exploit incremental reasoning - largely outperform general conjunctive query evaluation engines.

Future work will include identification of different parallelization strategies, based on suitable reorderings of the operations performed during the translation of DLN and further possible optimizations for DLN , such as caching the translations used for previous queries. We further plan to extend SOVGen for general metarules. Last but not least we hypothesize the realization of an inference engine that allows incremental retraction.

References

[1] P.A. Bonatti, M. Faella, I. M. Petrova and L. Sauro: A new semantics for overriding in description logics. Artificial Intelligence Journal, 222:1–48, 2015

Cooperations

Currently I am working in collaboration with with Prof. Dr. Piero A. Bonatti and Dr. Luigi Sauro, members of the Department of Electrical Engineering and Information Technologies.

4. Products

A. Conference papers

[2] “Optimized Construction of Secure Knowledge-Base Views”, P. A. Bonatti, I. M. Petrova and L. Sauro, Proceedings of the 28th International Workshop on Description Logics. CEUR Workshop Proceedings 1350, CEUR-WS.org

[3] “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

B. International journal papers

C. Technical reports

[4] “Optimizing the computation of overriding”, P. A. Bonatti, I. M. Petrova, L. Sauro, CoRR abs/1507.04630 (2015)

D. Papers in preparation

“Improving Module Extraction for Nonmonotonic and Classical DL”, P. A. Bonatti, I. M. Petrova, L. Sauro, to be submitted to Journal of Automated Reasoning.

5. Conferences and Seminars

A. Conferences

I participated in the 28th International Workshop on Description Logics (DL 2015), held in Athens, Greece from June 7th to June 10st, 2015. In that place I have made a poster presentation of the paper “Optimized Construction of Secure Knowledge-Base Views”, joint work with Piero Andrea Bonatti and Luigi Sauro.

B. Workshop

“1st Workshop of the Project Cluster on Data Protection, Security and Privacy in the Cloud”, 23 February 2016, Naples, Italy

6. Activity abroad

7. Tutorship

- I am teaching assistant for the courses of Web Technologies and Programming Languages I, courses of the Bachelor's and the Master's degree in Computer science, Department of Electrical Engineering and Information Technologies, Università degli Studi di Napoli Federico II.

- On 26 October 2015 I have been awarded with "Assegno di attività di tutorato" within the "Scuola Politecnica e delle Scienze di Base" of University of Naples Federico II. In this occasion I have been involved also in the course Programmazione II: Laboratorio, course of the Bachelor's degree in Computer science, Department of Electrical Engineering and Information Technologies, Università degli Studi di Napoli Federico II.