PhD in Information Technology and Electrical Engineering

Università degli Studi di Napoli Federico II

PhD Student: Enrico Caldarola

XXIX Cycle

Training and Research Activities Report – First Year

Tutor: Antonio Picariello – co-Tutor: Antonio Rinaldi

1. Information

I graduated in Computer Science Engineering at "Politecnico di Bari" in 2006. I have been working as computer programmer since 2006 and now I am attending the XXIX cycle at the Department of Electrical Engineering and Information Technologies (DIETI), University of Naples "Federico II", collaborating with the DIETI group headed by Prof. Antonio Picariello, my tutor, and Prof. Antonio Rinaldi, my co-tutor.

I am also research fellow at the Institute of Industrial Technologies and Automation, National Council of Researches in Bari in the EVA group (Enterprise Engineering and Virtual Applications).

2. Study and Training activities

During the first year of my PHD, I have attended the "*Site Reliability Engineering at Google"* seminar held by Ph.D. Marco Papa Manzillo, Senior SWE SRE, Google; the short course "*Three core issues for the Internet: things, security and economics*" held by Prof. Henning Schulzrinne. I have also attended the course: "Project Manager per la Ricerca" held by prof. Guido Capaldo. All seminars and courses have been organized by the department DIETI of University of Naples.

I am going to attend the following ad hoc modules in the next weeks: "*Semantic web reasoners: struttura, uso e ottimizzazioni"* held by professor Piero Bonatti and "*Laboratorio di software per l'ottimizzazione"* held by Antonio Sforza, Claudio Sterle at DIETI department, university of Naples.

Research activity

My research activities aim at using, combining and eventually enhancing ontology integration and semantic matching techniques also evaluating their scalability in terms of Big Data. The ontology integration is a broad field of research, which is concerned with determining and overcoming mismatchings between ontologies in order to allow the reuse of such ontologies. Ontology integration along with the paradigm shift involved by Big Data, as for theories and technologies in data management, offers a new perspective to deal with the huge pile of interconnected and (now even more) semantically related data over the Internet, fostering the concrete realization of the semantic web. In order to carry out the research topics above, I have spent much of my Phd first year doing a literature review of existing and well-known ontology matching techniques. I have collected and critically analysed several scientific papers describing linguistic or conceptual matching techniques, or more specifically ontology-oriented matching techniques, also detailing the similarity measures, which have been proposed throughout the past years from other researchers. I have also investigated existing frameworks dealing with the ontology integration problem from a methodological perspective. Parallel to researches on ontology matching techniques, I have carried out a literature review on Big Data, from the high-level characterization of the term "Big" throughout the Volume-Velocity-Variety model, to the existing technologies dealing with very large databases in terms of storage and computing. Furthermore, I have investigated the potential scenarios in which the Big Data paradigm shift matters, from data-intensive sciences to social networking, business intelligence and modern companies.

I have also conducted a study of the main international journals related with my research topics, annotating indexing and impact factors for each of them, in order to individuate proper journals for the forthcoming articles I plan to write.

Products

From the study of the art of Big Data, I have submitted two journal articles, which have been accepted and published, to Applied Computer Sciences (University of Lublin) and ACM Software Engineering Notes respectively. The first article[1] , after a comprehensive characterization of the Big Data definition in terms of Volume, Velocity and Variety of data, describe the most compelling scenarios of Big Data such as the grid computing for the LHC at CERN, the SLOAN Digital telescope image pipelining, Twitter, the Advanced manufacturing, etc. For each scenario, the paper highlights the technological challenges arising from the dimensions characterizing it. The second article[2] is a position paper pointing out the idea that modern companies cannot escape the paradigm shift coming from Big Data, but, on the contrary, must be able to get value from it, adopting new analytics strategies and tools, and exploiting any data coming from the company legacy databases or from the Web.

Starting from the literature review about the ontology matching, I have presented a paper[3] at "The Seventh International Conference on Information, Process, and Knowledge Management (eKNOW 2015)" in Lisbon. The study introduced in the paper address the efforts to examine the applicability of an approach to knowledge reuse based on the combination of existing techniques and methods proposed in the literature. It also proposes and uses a framework for knowledge reuse based on qualitative criteria and linguistic similarity measure exploiting some well-known linguistic resources like WordNet in the context of a real industrial scenario.

Conferences and Seminars

From the 22th to the 27th of February, 2015 I participated at "The Seventh International Conference on Information, Process, and Knowledge Management (eKNOW 2015). During the conference, I attended the following tutorials:

- *Patient Privacy and Security in eHEALTH* held by Dr. Güney Gürsel, Gülhane Military Medical Academy, Ankara, Turkey;
- *Cutting-Edge Technology for Patent Analysis* held by Prof. Dr.-Ing. Sigram Schindler, TELES Patent Rights International GmbH, Germany;

[1] Enrico Caldarola, Marco Sacco, Walter Terkaj, "*Big Data: The Current Wave Front of the Tsunami*, Applied Computer Science Volume 10, Number 4, 2014, Lublin University. [Online available] http://www.acs.pollub.pl/pdf/v10n4/1.pdf

[2] Enrico Caldarola, Antonio Picariello, Daniela Castelluccia, "*Modern Enterprises in the Bubble: Why Big Data Matters"*, ACM SIGSOFT Software Engineering Notes, Volume 40, Issue 1, January 2015, pp. 1-4, ACM New York, NY, USA, doi>10.1145/2693208.2693228

[3] Gianfranco Modoni, Enrico Caldarola, Walter Terkaj, Marco Sacco, *The ontology reuse in an industrial scenario: a case study,* In proceedings of eKNOW 2015: The Seventh International Conference on Information, Process, and Knowledge Management, pp. 66-71, Lisbon 2015. [online available] http://www.thinkmind.org/index.php?view=article&articleid=eknow_2015_3_30_60112.

I have also attended the following opening sessions and keynotes:

- *Someplace between Dreams and Potential: Reflections on the Current Status of Technological Innovation in Health Care* by Prof. Dr. Anthony Glascock, Drexel University, Philadelphia, USA;
- *What is Smart? Smart cities, Smart buildings, Smart people. Technology in a Human-centric Perspective* by Prof. Dr. Lasse Berntzen, HBV, Norway;
- *From Unmanned Vehicles to Cyber Security: More Than Twenty Years of CIRCA Research Towards Trusted Autonomy* by Dr. David Musliner, Smart Information Flow Technologies (SIFT), USA;

Finally the following panels:

- *Topic: Visions of the Future - Knowledge and Education 10 Years from Now*, moderated by
- Stephen White, University of Huddersfield, UK;
- *Topic: Digital Media Impact on Human-Computer Interaction*, moderated by Leslie Miller, Iowa State University, USA
- *Topic: Geo Measurements and Urban Challenges*, moderator Claus-Peter Rückemann, WWU Münster and Leibniz Universität Hannover and HLRN, Germany;
- *HealthCare Platforms: Lessons Learned and Future Challenges Moderator* Jean-Luc Strauss, Altran Research, France.

I have also presented my article with a PowerPoint presentation of 15-20 minutes in the scheduled session.

CS summary

Student: Enrico Caldarola	**Tutor: Antonio Picariello**	**Cycle XXIX**
enricogiacinto.caldarola@unina.it	antonio.picariello@unina.it	

	Credits year 1							
		1	2	3	4	5	6	
	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	**Summary**
Modules							3	**3**
Seminars		0	0	0	0	0,5	5	**5,5**
Research		6	6	8	10	7	7	**44**
	60	6	6	8	10	7,5	15	**53**