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 – Third Year

Tutor: Prof. Antonio Picariello – co-Tutor: Prof. Antonio Rinaldi

UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II

1. Information

I graduated in Computer Science Engineering at "Politecnico di Bari" in 2006. I worked on my dissertation research activities in the Knowledge Factory of IBM Italy (Bari Office) and then I continued to work on the same subjects for further 6 months in a stage. Later on, I was hired in a private and public research consortium, namely Sintesi SCpA, where I have worked as computer programmer. Starting from March 2014, I am attending the XXIX cycle of Phd course in Intofmation Technology and Electrical Engineering, at the Department of Electrical Engineering and Information Technologies (DIETI), University of Naples "Federico II", collaborating with the group headed by Prof. Antonio Picariello, my tutor, and Prof. Antonio Rinaldi, my co-tutor. At the same time, since September 2013 I joined, as research fellow, the Institute of Industrial Technologies and Automation at the Italian National Council of Researches (specifically, the EVA group: Enterprise Engineering and Virtual Applications), in the Bari office. Here I conduct researches on semantic web related technologies or languages in order to make interoperable software and hardware devices. I am involved in several national and European research projects, such as:

-Design For All (http://www.d4all.eu/en/), which consists in designing and implementing a platform for software tools integration, hence the formalization of domestic environment available information, design and develop a software architecture able to both manage (as representation and exposure) the information and knowledge related to the domestic environment, and enabling the efficiency of information exchange among software tools.

-Pegaso: Personalised Guidance Services For Optimising Lifestyle In Teen-Agers (http://pegasof4f.eu/home), which, starting from the consideration that knowing how to stay healthy is not enough to motivate individuals to adopt healthy lifestyles, targets teen-agers exploiting technologies and approaches they are familiar with, like Gaming strategies etc., in order to leverage social networks and communities of interest, integrated in a participatory design methodology to make the difference.

-Apps4aME: Engineering Apps for advanced Manufacturing Engineering, which aims at the comprehensive consideration of ICT-based support of Manufacturing Engineering in all the factory domains toward an advanced Manufacturing Engineering (aME).

2. Study and Training activities

During the third year of my PHD, I have attended different training activities:

Information Retrieval Systems – held by Prof. A.M. Rinaldi, which is a Msc module from the degree course in Computer Science Engineering. The course content is: introduzione all'ir, modelli, valutazione, ruolo degli utenti, text retrieval e indicizzazione, architetture dei motori di ricerca, applicazioni emergenti.

To be attended “Le imprese e la ricerca” held by Dott. Frizzarin, at University of Naples, dept. Mechanical Engineering.

In addition to the MS module described above I’ve attended the following seminar (both internal or external):

Tutorials and keynotes at the Eighth International Conference on Information, Process, and Knowledge Management, eKnow 2016, Venice. In particular, I attended the following:

- Tutorial: Big data for Personalized and Persuasive Coaching via Self-monitoring Technology (2.5 h), held by Lisette van Gemert-Pijnen, Annemarie Braakman, Olga Kulyk, Liseth Siemons, Floor Sieverink from University of Twente-Enschede, The Netherlands;

- Tutorial: Big Data: Processing Anatomy (1.5 h), held by Petre Dini, Concordia University, Canada, IARIA, USA;

- Tutorial: Where Data Lives: Centricity with Complex Data and Advanced Computing (1.5 h), by Claus-Peter Rückemann, Leibniz Universität Hannover / Westfälische Wilhelms-Universität Münster / North-German Supercomputing Alliance (HLRN), Germany;

- Tutorial: Brain, Child, Self and Toy Robots: Enrobotment (1.5 h), held by Irini Giannopolu, from Pierre Marie Curie University in Paris, France.

- Keynote: Semantic Search (1 h), Prof. Dr. Lubomir Stanchev, California Polytechnic State University, USA;

- Keynote: Robust Evidence for the Impact of At-home Telemonitoring for Chronic Disease Management (1 h) held by Prof. Dr. Branko Celler, CSIRO Australian eHealth Research Centre, Australia;

-Seminar: Patient Role in Mobile Adaptable Healthcare: Awareness and Accessibility, Anne G. Ekeland, Norwegian Centre for Integrated Care and Telemedicine / University Hospital of North Norway | University of Tromsø, Norway Marianne Trondsen, Norwegian Centre for Integrated Care and Telemedicine / University Hospital of North Norway | University of Tromsø, Norway (1.5 h);

-Seminar: Learning- and Knowledge-based Adaptive Human-Machine Interactions, (2 h) held by Alf C. Zimmer, University of Regensburg / Engineering Psychology Unit, Germany.

Other internal seminars are:

MINIX3: A Reliable and Secure Operating System – held by Prof. Tannenbaum at University of Naples;

“Neuro-Robotics: Modelling Sensor-Motor Control Functionalities with Spiking Neural Networks — The Renaissance of Bio-Cybernetics?”, (1.5 h), held by prof. Rüdiger Dillmann, DIETI dept., University of Naples.Seminar - “Neuro-Robotics: Modelling Sensor-Motor Control Functionalities with Spiking Neural Networks — The Renaissance of Bio-Cybernetics?”, (1.5 h), held by prof. Rüdiger Dillmann, DIETI dept., University of Naples.

Exploiting machine learning techniques in software development processes, held by Ing. Domenico Amalfitano, at University of Naples, dept. DIETI.

I've also attended the 3rd International Winter School on Big Data, at University of Bari, Italy, from 13th to the 17th of February, 2017. Here some of the courses attended at the school:

Using R for Mixed-effects (Multilevel) Models, held by Prof. Paul Bliese, Decision Trees for Big Data Analytics, held by Hendrik Blockeel, Big Data Approaches in Astronomy, held by Tamás Budavári, Data-aware Processes: Modeling and Verification, held by Diego Calvanese, Using High Performance Computing for Big Data Analytics, held by Geoffrey C. Fox, Streaming Big Data Analytics, held by Minos Garofalakis; Data Visualization with R held by Georgios B. Giannakis, Signal Processing Tools for Big Network Data Analytics, held by Prof. Sander Klous; We Are Big Data; Analysis of Large Social Networks, , held by Prof Laks V.S. Lakshmanan, Ontology-based Data Management, held by Prof. Maurizio Lenzerini, Swarm Intelligence Methods and Optimization Problems in Big Data Analytics, held by Prof. Soumya D. Mohanty, Introduction to Data Stream Mining for Big Data, held by Prof. Bernhard Pfahringer, Harnessing Big Data for Building Smart Things, held by Prof. Krithi Ramamritham, Coupled Oscillators Approach in Time Series Analysis, held by Prof, Michael Rosenblum, Data Security and Privacy in the Cloud, held by Prof. Pierangela Samarati, Big Data and Cybersecurity, held by Prof. V.S. Subrahmanian, Recommender Systems and Big Data, held by Alexander S. Tuzhili , Big Data Algorithms that Aren't Machine Learning, held by Prof. Jeffrey Ullman, Sentiment Mining from User Generated Content, held by Prof. Lyle Ungar; Knowledge Discovery from Relational and Multimedia Data, held by Prof. Zhongfei Zhang

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 mismatching between ontologies (or in general knowledge bases) in order to allow the reuse of such ontologies. Knowledge 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 first years 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.

With the second year of my phd, I conducted further developments and researches in my doctoral thesis about the integration of very large knowledge bases specifically focused on: collecting the state-of-the-art of matching techniques: string matching, linguistic/semantic matching and structural matching; exploiting the above techniques and metrics in order to find a general measure of similarity to relate entities coming from different knowledge bases each other. Moreover, I did experiments on matching two well-known large, generalistic knowledge bases, namely: DBpedia and Wordnet, as a benchmark or test suite to evaluate performances (in terms of precision and recall) of my matching algorithms and techniques. I learnt and used Neo4J as a new GraphDB to import, query and visualize very large databases. In particular, some experiments have been conducted in visualizing the WordNet lexical DB within Neo4J and Cytoscape. In addition, Neo4J has been used for a rapid prototyping of a graph structure to handle the similarity measure between entities coming from WordNet and DBpedia tested matching algorithms. In this regard, at the end of the second year I found a new way to visualize WordNet synsets in Neo4J and Cytoscape using a tag clouds-based approach.

In this third year of the phd course, I conducted research and development about the multi-strategy methodology for ontology integration and reuse that represents the subject of my doctoral thesis. In detail, my activities were: designing and development of software modules to adapt some reference model ontologies (retrieved from the Web) in the proposed framework for ontology integration; Designing and implementing the Matching and Alignment modules. Designing and development of the Integrator module for the proposed software framework for ontology integration and reuse. Designing and development of an Evaluation module for the matching strategy adopted in the proposed approach. Evaluation of the proposed methodology by studying and applying some methodologies for evaluating multi-class classifier. In particular, a binarization technique has been adopted for the binarization of the multi-class classifier in order to establishing, through a thresholding strategy, the optimum values for the thresholds used within the classifier function. Also, have studied and used some curves and diagrams in order to shows and evaluate the experimental results of the classifier in terms of precision and recall. In particular, P/R curves and ROC (Receiver Operator characteristic curves) have been investigated. Furthermore, semantically-grounded evaluation of the alignment has been investigated in order to improve the precision of the Aligner.

Products

During the third year, I submitted and published four conference papers coming from the research activities above described:

[conference paper, best paper award] Caldarola, E., Picariello, A. and Rinaldi, A., WordNet exploration and visualization in Neo4J.A Tag Cloud-based approach, In proceeding of Eighth International Conference on Information, Process, and Knowledge Management, eKNOW 2016, Venice, Italy;

[conference paper] Enrico G. Caldarola, Antonio M. Rinaldi, Improving the Visualization of WordNet Large Lexical Database through Semantic Tag Clouds, IEEE BigData Congress 2016, San Francisco, June 2016, DOI: 10.1109/BigDataCongress.2016.14;

[conference paper] Enrico G. Caldarola, Antonio M. Rinaldi, An Approach to Ontology Integration for Ontology Reuse, 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI), Pittsburgh, Pennsylvania, USA, 2016. DOI: 10.1109/IRI.2016.58.

[conference paper] Enrico G. Caldarola, Antonio Picariello, Antonio Rinaldi, Marco Sacco, Exploration and Visualization of Big Graphs - The DBpedia Case Study, Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Porto - Portugal, Volume: Vol 1. DOI: 10.5220/0006046802570264;

[journal article] Enrico G. Caldarola, Antonio Picariello, A. M. Rinaldi, M. Sacco, Enhancing theWordNet Exploration and Visualization in Neo4J with a Tag Cloud Based Approach, Journal on Advances in Intelligent Systems, vol 9 no 3 & 4, year 2016, http://www.iariajournals.org/intelligent_systems/

[journal article] Daniela Castelluccia, Enrico G. Caldarola, Nicola Boffoli (additional authors: Antonio M. Rinaldi, Marco Sacco and Nicola Ungaro), Environmental Big Data: a systematic mapping study, ACM SIGSOFT Software Engineering Notes archive, Volume 41 Issue 6, November 2016 , Pages 1-4 , doi>10.1145/3011286.3011307;

[journal article] Enrico G. Caldarola, Antonio Picariello, A. M. Rinaldi, Experiences in WordNet Visualization with Labeled Graph Databases, Communications in Computer and Information Science 631, pp. 80–99, 2016., DOI: 10.1007/978-3-319-52758-1 6;

[journal article] Enrico G. Caldarola, A.M. Rinaldi, A Multi-strategy Approach for Ontology Reuse through Matching and Integration Techniques, Advances in Intelligent Systems and Computing, Springer 2017;

Conferences

Participation as presenter at the Eighth International Conference on Information, Process, and Knowledge Management, eKnow 2016 held in Venice from 24th to 28th of April, 2016.

Preparing a video presentation for the 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI), Pittsburg, Pennsylvania, USA.

Tutorship

Seminar: Introduction to Apache Lucene™, held during the M.Sc. course in Information Retrieval Systems at University of Naples, Prof. A.M. Rinaldi.

Several tutorship activities carried out for the MSc degree candidates which have attended the MSc course in Information Retrieval Systems.

CS summary

The following table reports my credits for all the three years of my phd activities.

	Credits year 1								Credits year 2								Credits year 3									
		1	2	3	4	5	6			1	2	3	4	5	6			1	2	3	4	5	6			
	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Estimated	bimonth	bimonth	bimonth	bimonth	bimonth	bimonth	Summary	Total	Check
Modules							5	**5**		4			6		9	**19**					6		4	**10**	**34**	**30-70**
Seminars						0.5	6	**6.5**		1.2	6.2	0.4		0.5	5	**13**		2.5				0.4	5.2	**8.1**	**28**	**10-30**
Research		8	7	8	10	8	7	**48**		5	3	7	5	6	2	**28**		8	8	9	7	6	4	**42**	**118**	**80-140**
								60								**60**								**60**	**180**	**180**