

Fabio Massimo Zanzotto - CV

EDUCATION

- 2002 PhD in Computer Science and Automation, Faculty of Engineering, Univ. of Rome “Tor Vergata”, Italy PhD Supervisor: *Prof. Maria Teresa Pazienza*
- 2002 Master in Business Administration (MBA), Profingest Business School, Bologna
- 1997 MS in Computer Science Engineering, Univ. of Rome “Tor Vergata”, Italy (*cum Laude*)

CURRENT POSITIONS

- 2013- Coordinator of the Master Program: *Indicizzazione di documenti cartacei, multimediali ed elettronici in ambiente digitale*, Dept. of Humanities, Univ. of Rome “Tor Vergata”, Italy
- 2006- Associate Professor, Dept. of Enterprise Engineering, Univ. of Rome “Tor Vergata”, Italy

Research Statement and Research Interests

Fabio Massimo Zanzotto is an Associate Professor at Department of Enterprise Engineering of the University of Rome “Tor Vergata”. Since 1998, he has interests in the research endeavor of Artificial Intelligence. He is active in the area of Natural Language Processing and Machine Learning, mainly working in four topics: distributed/distributional models for NLP, AI applied to precision medicine, recognizing textual entailment and syntactic parsing for Italian. Additionally, he has recently got interested in the field of Ethics&AI.

Area of distributed/distributional models for NLP Representing syntax and meaning in vectors and tensors is becoming an very active research area in these years due to the recent renaissance of deep neural networks. Zanzotto is then interested in understanding the implication of distributed representations and compositional distributional semantics models (CDSMs) in natural language processing [18, 11, 15, 21]. He has started by proposing one of the first model to estimate full additive CDMSs [18]. Then, he has investigated the relation between syntax and semantics in CDSMs. As a result, he has hypothesized that CDSMs have a clear separation between syntax and semantics [6, 17]. Hence, he has started to investigate how to encode [21] and decode [?] syntactic structures in small vectors called *distributed trees*. Then, he proposed a way to clearly merge syntactic and semantic information in the same tensor [7]. Finally, he has started to investigate how these vectors can be directly produced from sentences by trying to replace syntactic parsers with distributed syntactic parsers [16, ?].

For this research area, he has been involved in the program committee of GEMS: GEometrical Models of Natural Language Semantics in (2009,2010, 2011) and *Sem (2012).

Area of AI applied to precision medicine In recent years, the approach to medicine has substantially changed: global approaches have been pressured by a growing availability of electronic health records (EHR) and by the consequent demand to provide precision medicine. Hence, Zanzotto and colleagues proposed a different approach from that generally used in the development of risk assessment models based on the arbitrary assignment of a score according to association analyses. To this purpose, Zanzotto used kernel learning machines and random

optimization (RO) to produce VTE risk predictors in a population of consecutive ambulatory cancer patients. The risk predictors exploit significant patterns in and can be used in the development of clinical decision support systems [9, 8, 10].

Area of Recognizing Textual Entailment The term recognizing textual entailment (RTE) has been introduced in NLP to systematically foster studies in defining computational models that replicate the human ability to determine whether or not texts imply sentences. A simple RTE example is determining whether or not “Acme bought BigT” entails “Acme owns BigT”. Zanzotto has been interested in this scientific endeavor since its beginning in 2005 and published a book on this topic [5] co-authored with Ido Dagan Mark Sommons and Dan Roth. He is mainly interested in studying the application of machine learning models to the RTE. After 2005, a first exploratory year [13], during the investigation of the application of ML models to RTE task, Zanzotto produced a major innovation for the specific RTE field [20, 19, 14] as well as for the related fields of natural language processing, machine learning [12], and graph analysis applied to text [22]. It is still under analysis if the innovation can generate an interest in the research area of graph theory as it involves graph isomorphism on a particular class of graphs. The innovation is the following. In the context of machine learning models based on kernel functions, he proposed a novel class of feature spaces encoding first-order rewrite rules. With this new class of feature spaces and the related kernel functions, Zanzotto and his colleagues produced a system which scored at the 3rd place in the 2006 worldwide RTE Challenge (scored 1st among the academic systems) and in the 5th place in 2007 RTE Challenge. The exploration of full potential of this idea is still an on-going research as the class of first-order rewrite rule feature spaces can be applied in many areas of natural language processing (machine translation, document summarization, stylistic control systems, question-answering, and dialogue models) as well as in many other research areas.

For these interests, Zanzotto has been invited to the program committees of all the textual entailment recognition workshops and challenges after the first. In 2009, Zanzotto co-chaired the ACL workshop of Applied Textual Inference that has a program committee with researchers in the area of textual entailment recognition and natural language processing. In 2009, he co-organized the Italian chapter of the RTE challenge in Evalita 2009. Zanzotto, together with Ido Dagan and Dan Roth, gave a tutorial titled “Textual Entailment Recognition” in the 45th Association of Computational Linguistics (ACL) Annual Conference. He co-organized two editions of the TextGraphs Workshop series (2010,2011). He steadily participated to the program committees of the TextGraphs (2007, 2008, 2012, 2013) and he participated to the program committee of the workshop MLG-2010: Mining and Learning from Graphs. He has been area chair for Textual Entailment in *Sem (2013).

Area of Syntactic Parsing In the early stages of his career, Zanzotto developed a robust natural language syntactic parser [1, 2, 3, 4] based on two principles: lexicalization and modularization. He proposed a novel model for representing syntactic information for modular parsing system that combined the two alternative theories: the dependency and the constituency based model. The resulting syntactic parser is still one of the only available syntactic parsing system for Italian.

For these interests, Zanzotto has been invited as program committee member in the two workshops on ROBust Methods for Analyzing Natural language Data (ROMAND) (2004,2006). Due to his expertise in Italian parsing technologies, he participated as co-organizer in the “parsing track” of the Evalita 2007 Challenge, he wrote the parsing technology chapter of an Italian Artificial Intelligence Book, and he gave a tutorial in the Italian Association for Artificial Intelligence in 2001.

General information Zanzotto is author of more than 100 publications. He is a reviewer for the major conferences in the area of NLP and AI (ACL, NAACL, EAACL, EmNLP, CoLing, LREC, IJCAI, ECAI, CLEF) and for journals (Computational Linguistics, Journal of Natural Language Engineering, Cognitive Computation Journal, ACM Transactions on Intelligent Systems, IEEE Data and Knowledge Engineering). He is member of ACL and of the Italian Association on Artificial Intelligence (AIIA).

PREVIOUS POSITIONS

- 2004-2006 Full-Time Researcher, Univ. of Milano-Bicocca, Italy
- 2002-2004 Post-Doc (Assegnista di Ricerca), Univ. of Rome, “Tor Vergata”, Italy
- 1998-1999 Visiting Researcher, Xerox Research Centre Europe, Grenoble, France

FELLOWSHIPS AND AWARDS

- 2010 Yahoo! Faculty Research and Engagement Award
- 1998-2001 PhD Scholarship, Faculty of Engineering, Univ. of Rome “Tor Vergata”, Italy (Univ. RTV)
- 1999 Scholarship European Program “Leonardo da Vinci” given by Univ. RTV

CITATIONS IN THE CULTURAL PAGES OF NATIONAL PRESS

- Sole 24 Ore Sanità, Innovazione 2.0 per il Diabete, 13/12/2016
- Repubblica: Tradurre Pinocchio in Emoticon, 17/4/2016
- Corriere del Trentino, 20/5/2014, p. 13, Scrivere: Abbreviazioni da Tirone ai Tweet
- Corriere della Sera: inserto domenicale La Lettura 13 Ottobre 2013, L’IMU è una frase d’amore
- Espresso On Line, 22/6/2011, Sms, il ’tvb’ entra nell’Oxford di Maria Simonetti
- Corriere della Sera, 11/5/2011, Perché studiare lo slang giovanile

ORGANISATION OF SCIENTIFIC MEETINGS

- 2018 Senior Area Chair of the Association for Computational Linguistic Conference (ACL)
- 2015 Co-Chair, The first Italian Computational Linguistics Conference (CLIC)
- 2014 Area Co-Chair, The first Italian Computational Linguistics Conference (CLIC)
- 2013 Area Chair, Second Joint Conference on Lexical and Computational Semantics (* Sem)
- 2013 Special Session Organizer, International Conf. on Pattern Recognition in Bioinformatics, France
- 2012 Program Co-Chair, Brain Informatics, Macau
- 2011 Scritture Brevi (Short Writing): 3 workshops, Univ. RTV
- 2011 Co-Organizer of TextInfer Workshop co-located at EMNLP
- 2011 Co-Organizer of TextGraphs-6 co-located at ACL
- 2010 Co-Organizer of TextGraphs-5 co-located at ACL
- 2009 Co-Organizer of TextInfer co-located at ACL 2009, Singapore

INSTITUTIONAL RESPONSIBILITIES

- 2012- Faculty Member, PhD Program Science of Information, Univ. RTV
- 2006-2012 Faculty Member, PhD Program Computer Science and Automation, Univ. RTV
- 2013- Elected Member of the Independent Quality Commission, Dept. of Enterprise Eng., Univ. RTV
- 2011- Member of the Dept. of Enterprise Engineering, Univ. RTV
- 2006-2012 Member of the Faculty of Humanities, Univ. RTV
- 2006-2011 Member of the Dept. of Comp.Sci., Systems, and Production, Univ. RTV
- 2004-2006 Member of the Dept. of Comp.Sci., Systems, and Information, Univ. of Milano-Bicocca, Italy

TEACHING ACTIVITIES

- 2013- *Programming Languages and Methodologies* (6CFU), Computer Science Degree, Univ. RTV
- 2012- *Natural Language Processing* (6CFU), Computer Science (MS), Univ. RTV
- 2006- *Computer Science and Knowledge Representation* (12CFU), *Basic Computer Science for Humanities* (6CFU), Faculty of Humanities, Univ. RTV
- 2005-2006 *Artificial Intelligence* (6CFU) and *Data Bases* (6CFU), Computer Science, Univ. of Milano-Bicocca, Milan, Italy

Selected Publications

- [1] Roberto Basili, Maria Teresa Pazienza, and Fabio Massimo Zanzotto. Efficient parsing for information extraction. In *Proc. of the ECAI98*. Brighton, UK, 1998.

- [2] Roberto Basili, Maria Teresa Pazienza, and Fabio Massimo Zanzotto. Lexicalizing a shallow parser. In *Proc. of the Traitement Automatique de la Langue Naturelle, TALN99*. Cargese, FR, 1999.
- [3] Roberto Basili, Maria Teresa Pazienza, and Fabio Massimo Zanzotto. Customizable modular lexicalized parsing. In *Proc. of the 6th International Workshop on Parsing Technology, IWPT2000*. Trento, Italy, 2000.
- [4] Roberto Basili and Fabio Massimo Zanzotto. Parsing engineering and empirical robustness. *Natural Language Engineering*, (8/2-3):1245–1262, 2002.
- [5] Ido Dagan, Dan Roth, Mark Sammons, and Fabio Massimo Zanzotto. *Recognizing Textual Entailment: Models and Applications*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2013.
- [6] Lorenzo Ferrone and Fabio Massimo Zanzotto. Linear compositional distributional semantics and structural kernels. In *Proceedings of the Joint Symposium of Semantic Processing (JSSP)*, 2013.
- [7] Lorenzo Ferrone and Fabio Massimo Zanzotto. Towards Syntax-Aware Compositional Distributional Semantic Models. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 721–730, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics.
- [8] P. Ferroni, F.M. Zanzotto, N. Scarpato, S. Riondino, F. Guadagni, and M. Roselli. Validation of a machine learning approach for venous thromboembolism risk prediction in oncology. *Disease Markers*, 2017.
- [9] P. Ferroni, F.M. Zanzotto, N. Scarpato, S. Riondino, U. Nanni, M. Roselli, and F. Guadagni. Risk Assessment for Venous Thromboembolism in Chemotherapy-Treated Ambulatory Cancer Patients. *Medical Decision Making*, 37(2), 2016.
- [10] F. Guadagni, F.M. Zanzotto, N. Scarpato, A. Rullo, S. Riondino, P. Ferroni, and M. Roselli. *RISK: A random optimization interactive system based on kernel learning for predicting breast cancer disease progression*, volume 10208 LNCS. 2017.
- [11] Yashar Mehdad, Alessandro Moschitti, and Fabio Massimo Zanzotto. Syntactic/semantic structures for textual entailment recognition. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10*, pages 1020–1028, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
- [12] Alessandro Moschitti and Fabio Massimo Zanzotto. Fast and effective kernels for relational learning from texts. In *Proceedings of the International Conference of Machine Learning (ICML)*. Corvallis, Oregon, 2007.
- [13] Maria Teresa Pazienza, Marco Pennacchiotti, and Fabio Zanzotto. Textual entailment as syntactic graph distance: a rule based and a svm based approach. In *Proceedings of the 1st Pascal Challenge Workshop*, Southampton, UK, 2005.
- [14] Fabio Massimo Zanzotto and Lorenzo Dell’Arciprete. Efficient kernels for sentence pair classification. In *Conference on Empirical Methods on Natural Language Processing*, pages 91–100, 6-7 August 2009.
- [15] Fabio Massimo Zanzotto and Lorenzo Dell’Arciprete. Distributed structures and distributional meaning. In *Proceedings of the Workshop on Distributional Semantics and Compositionality*, pages 10–15, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.
- [16] Fabio Massimo Zanzotto and Lorenzo Dell’Arciprete. Transducing sentences to syntactic feature vectors: an alternative way to “parse”? In *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, pages 40–49, 8 August 2013.
- [17] Fabio Massimo Zanzotto, Lorenzo Ferrone, and Marco Baroni. When the whole is not greater than the combination of its parts: A decompositional look at compositional distributional semantics. *Comput. Linguist.*, 2015.
- [18] Fabio Massimo Zanzotto, Ioannis Korkontzelos, Francesca Fallucchi, and Suresh Manandhar. Estimating linear models for compositional distributional semantics. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, August, 2010.
- [19] Fabio Massimo Zanzotto and Alessandro Moschitti. Automatic learning of textual entailments with cross-pair similarities. In *Proceedings of the 21st Coling and 44th ACL*, pages 401–408. Sydney, Australia, July 2006.
- [20] Fabio Massimo Zanzotto, Marco Pennacchiotti, and Alessandro Moschitti. A machine learning approach to textual entailment recognition. *NATURAL LANGUAGE ENGINEERING*, 15-04:551–582, 2009.

- [21] F.M. Zanzotto and L. Dell'Arciprete. Distributed tree kernels. In *Proceedings of International Conference on Machine Learning*, pages 193–200, 2012.
- [22] FM Zanzotto, L Dell'Arciprete, and A Moschitti. Efficient graph kernels for textual entailment recognition. *Fundamenta Informaticae*, 107(2-3):199 – 222, 2011.