

Curriculum vitae

Name: Davide Buscaldi

E-mail: davide.buscaldi@lipn.univ-paris13.fr

Work Address: 99 av. Jean-Baptiste Clément, 93430 Villetaneuse, France

Phone: +33 6 84 57 49 65 (Mobile), +33 1 49 40 28 31 (Work)

Personal web page: <http://lipn.univ-paris13.fr/~buscaldi>

Current Positions

- Maître de Conférences (Associate Professor) at Université Paris 13, since 2012
- Part-time Assistant Professor (Chargé d'Enseignement) at Ecole Polytechnique, since Sept. 2020

Previous positions

- Post-doc researcher at the University of Orléans, laboratoire LIFO (Fundamental Informatics Lab, Orléans) January-December 2011. Automatic learning of medical and geographical ontologies using the pretopology formalism.
- Post-doc researcher at the IRIT Toulouse, December 2011-September 2012. Semantic Information Retrieval using ontologies and structured documents.

Education

- Ph.D. in Artificial Intelligence and Pattern Recognition, Ph.D. course with “quality mention” of the Spanish secretary of education, Universidad Politécnica de Valencia (Spain). Title of dissertation: “*Toponym Disambiguation in Information Retrieval*” Mark: « Cum Laude » (10/10). 2010. Panel composed by Mark Sanderson, Ross Purves, Paul Clough, Diana Santos and Emilio Sanchis. European mention. <http://dspace.upv.es/xmlui/handle/10251/8912>
- DEA degree (« Diploma de Estudios Avanzados », Degree of Advanced Studies) in Pattern Recognition and Artificial Intelligence, Universidad Politécnica de Valencia. 2007. Mark: “Sobresaliente” (9/10).
- Laurea degree (Roughly an M.Sc.) in Informatics, University of Genova (Italy), 2004. Graduation mark: 101/110

Research activities

Research topics:

Natural Language Processing, Machine Learning and Information Retrieval

Keywords:

- Information Extraction
- Knowledge Graphs
- Semantic Information Retrieval
- Geographic Information Retrieval (GIR)

- Toponym Disambiguation
- Word Sense Disambiguation
- Text Mining
- Question Answering
- Sentiment Analysis and Opinion Mining

Reviewing and PCs

- International Conferences and workshops: ESWC 2017-2022; SIGIR 2016, 2017 (short program); ADBIS 2013; IBERAMIA 2014; SOFSEM 16; ISWC 2016-2022; EKAW 2016; IJCAI-ECAI 2018, 2019, CoNLL 2019, 2020, AACL 2021, NLDB 2022
- Senior Area Chair for NAACL 2022 (Sentiment and Stylistic Analysis track)
- Area Chair for COLING 2022 (Sentiment and Opinion Mining track)
- National Conferences: CERI (2013-2015); CLiC-IT 2017, 2018; IC 2018, 2019,
- Workshops: RISE (2013-); SAVE-SD (2016-); KEGeoD @ SAC'2018; RECITAL 2017
- Journals: [The Computer Journal](#), [Data & Knowledge Engineering](#), [Computer Speech and Language](#), [International Journal of Geographical Information Science](#), [Knowledge and Information Systems](#), [Transactions on Information Systems](#), [Information Processing and Management](#), [Springer Nature Computer Science](#)

Scientific responsibilities

- Member of the Steering Committee for SemEval-2020
- Member of the Scientific Committee of the [Labex EFL](#)
- Member of the jury for the best PhD thesis prize (prix de la meilleure thèse) ATALA 2019
- Member of the Steering Committee for Human Language Technologies of the [AFIA \(French Association for Artificial Intelligence\)](#) (since 2020)
- External Reviewer at TU Dublin, Tallaght (Ireland)

Scientific Dissemination:

- Co-organizer of QAST 2009 for the Spanish language sub-task.
- Organization of the 1st International Workshop on the Extraction and Modelling of Knowledge from Scientific Texts (Emc-Sci), satellite workshop of IC 2017, Caen, France
- Organization of the Fourth International Workshop at ESWC on Sentic Computing, Sentiment Analysis, Opinion mining and Emotion Detection, satellite workshop at ESWC 2018, Crete, Greece
- Organization of task 7: Semantic Relation Extraction and Classification in Scientific Papers at SemEval 2018, co-located with NAACL-HLT 2018, New Orleans, USA
- Chair of the ESWC 2018 challenges track (together with D. Reforgiato)
- Organization of the 2nd Workshop VADOR (Valorisation des données de la recherche), CORIA-TALN-RJC 2018
- Organization of the series of Workshops on Deep Learning and Knowledge Graphs (DL4KG), at ESWC 2019-ESWC 2020 and on ISWC since 2021

PhD Students:

- Ehab Hassan (2013 - 2017) co-director with Aldo Gangemi - "Reconnaissance d'expériences vécues dans les avis d'utilisateurs : une méthode basée sur les événements."

- Simon David Hernandez Pérez (2016 - 2019) co-director with Thierry Charnois - "Topic detection and similarity in scientific literature"
- Vijini Liyanage (2021 -) co-director with Adeline Nazarenko "Knowledge-guided Detection of Automatically Generated Texts"
- Nabil Moncef Boukhatem (2020 -) co-director with Leo Liberti (Ecole Polytechnique), CIFRE thesis
- Hadi Abdine (2020 -) co-director with Michalis Vazirgiannis (Ecole Polytechnique)
- Fernando Lovera (2019 -) co-director with Thierry Charnois and Yudith Cardinale (Venezuela USB)

Research Projects

Current active projects:

- LabEx EFL (Empirical Foundation of Linguistics), axis 5: text semantics
<http://www.labex-efl.org/?q=en>
- Projet d'établissement (BQR) "Plongements lexicaux pour l'acquisition et la structuration de connaissances" (Porteur)

Previous projects:

- PhilHumans EU Project H2020-MSCA-ITN-2018: local coordinator for the LIPN
- CampusFrance "Polonium" project on sentiment analysis and irony detection (international project with Pedagogical University of Kraków, Poland)
- "Cartographie de la diaspora des savoirs par fouille sémantique des résultats recherche sur le web" ([projet ECOS-NORD](#))
- Golfred:G Machine reading for narrative generation of spatial experiences in service robots
- MOANO: Modèles et Outils pour Applications NOmades de découvertes de territoire.
- TEXT-MESS 2.0 (subproject UPV: TEXT-ENTERPRISE 2.0: Text comprehension techniques applied to the needs of the Enterprise 2.0)
- Red Temática TIMM: Tratamiento de Información Multilingüe y Multimodal (Automatic Processing of Multilingual and Multimodal Information)
- Extensión y Mejora de Recursos Léxicos en Árabe (Extension and Improvement of Lexical Resources in the Arabic Language)
- Answer Extraction for Definition Questions in Arabic
- TEXT-MESS: Minería de Textos Inteligente, Interactiva y Multilingüe basada en Tecnología del Lenguaje Humano (subproject UPV: MiDEs) (Intelligent, Interactive and Multilingual Text Mining, based on Human Language Technologies)
- Sistema de Búsqueda de Respuestas Inteligente basado en Agentes (AraEsp) (Agent-based Intelligent Question Answering System)
- ICT for EU-India Cross-Cultural Dissemination
- R2D2: Recuperación de Respuestas en Documentos Digitalizados (Answer Extraction in Digital Texts)
- CIAO SENSO: Combining Corpus-Based and Knowledge-Based Methods for Word Sense Disambiguation

Invited Talks:

- *NLP and Geographical Information Systems*, 3th Workshop on Semantic Web and New Technologies, Benemérita Universidad Autónoma de Puebla, Mexico, 12 July 2010 (Keynote Speaker)

- *Análisis Semántico*, LKE2015 Symposium, BUAP, Mexico, 27 October 2015 (Keynote Speaker)
- *Computational Semantics for Next-generation Natural Language Processing applications*, 2a Escuela de otoño de tecnologías del lenguaje, Cholula, Mexico, 27-28 September 2016
- *Information Géographique, textes et média sociaux*, Atelier extraction de connaissances en données spatialisés (EXCES), SAGEO 2017, Rouen, 6 November 2017 ([slides](#))
- [Improving access to scientific literature: a semantic IR perspective](#), Conferencia Española de Recuperación de la Información (CERI 2018), Zaragoza, 27 June 2018
- [Towards the Automatic Generation of States of the art: current trends and challenges in the analysis of scientific texts](#), 7th LKE Symposium, Dublin 28-31 October 2019

Selected Publications

a.) Journals:

1. Dessì D., Osborne F., Reforgiato D., Buscaldi D., Motta E. **Generating knowledge graphs by employing Natural Language Processing and Machine Learning techniques within the scholarly domain**. *Future Gener. Comput. Syst.* 116: 253-264 (2021)
2. Di Rocco L., Dassereto F., Bertolotto M., Buscaldi D., Catania B., Guerrini G. **Sherloc: a knowledge-driven algorithm for geolocating microblog messages at sub-city level**. In: *International Journal of Geographical Information Science*, pp.84-115, issue 35 (2021)
3. Reforgiato D., Alam M., Buscaldi D., Grezka A., Tavazoe F. **Frame-Based Detection of Figurative Language in Tweets**. In: *IEEE Computational Intelligence Magazine*, 14(4) pp.77-88. [Digital edition](#)
4. Reyes A., Rosso P., Buscaldi D. (2012) **From Humor Recognition to Irony Detection: The Figurative Language of Social Media**. In: *Data & Knowledge Engineering*, 74:1-12 <http://dx.doi.org/10.1016/j.datak.2012.02.005>
5. Buscaldi D., Rosso P., Gómez J.M., Sanchis E. (2010) **Answering Questions with an n-gram based Passage Retrieval Engine**. In: *Journal of Intelligent Information Systems*, 34(2) pp. 113--134
6. Buscaldi, D., Rosso, P. (2008) **A conceptual density-based approach for the disambiguation of toponyms**. *International Journal of Geographical Information Systems*, 22(3), pp.301-313, Taylor and Francis
7. Buscaldi D. (2011) **Approaches to Disambiguating Toponyms**. In: *SIGSPATIAL Special*. ACM, New York. Accepted, to be published.

b.) Peer-reviewed Intl. Conferences and workshops:

1. Gonzalez J.-A., Buscaldi D., Sanchis E., Hurtado L. **Transformer-based models for the Automatic Indexing of Scientific Documents in French**. *NLDB 2022*, Valencia
2. Liyanage V., Buscaldi D., Nazarenko A. **A Benchmark Corpus for the Detection of Automatically Generated Text in Academic Publications**. *LREC 2022*, Marseille
3. Lovera F., Cardinale Y., Buscaldi D., Charnois T., Homsí M.N. **Deep Learning Enhanced with Graph Knowledge for Sentiment Analysis**. *Joint proceedings of DeepOntoNLP and X-SENTIMENT workshops @ESWC 2021*, pp. 74-86, June 6th - 7th, 2021, Hersonissos, Grece

4. Dessì D., Osborne F., Reforgiato Recupero D., Buscaldi D., Motta E., Sack H. **AI-KG: An Automatically Generated Knowledge Graph of Artificial Intelligence**. ISWC (2) 2020: 127-143, International Semantic Web Conference, Athens 2020.
5. Ghanem B., Buscaldi D. and Rosso P. **TexTrolls: Identifying Trolls on Twitter with Textual and Affective Features**. Workshop on Online Misinformation- and Harm-Aware Recommender Systems, RecSys2020, pp.4-22, 25 September 2020, Rio de Janeiro, Brazil
6. Gabor K., Zargayouna H., Tellier I., Buscaldi D., Charnois T. **Exploring Vector Spaces for Semantic Relations**. Proceedings of EMNLP 2017, pp.1814-1823, Copenhagen, Denmark ([CORE A](#))
7. Buscaldi D., Hernandez, S., Charnois T. **Classification of keyphrases from scientific publications using WordNet and word embeddings**. Proceedings of VADOR@Inforsid 2017, pp.50-57, Toulouse, France
8. Gábor K., Zargayouna H., Tellier I., Buscaldi D., Charnois T. (2016) **Unsupervised Relation Extraction in Specialized Corpora Using Sequence Mining**. IDA 2016: 237-248 ([CORE A](#))
9. Hassan E., Buscaldi D., Gangemi A. (2015) **Correlating Open Rating Systems and Event Extraction from Text**. In: ICONIP 2015. Istanbul, Turkey. ([CORE A](#))
10. Buscaldi D., Bessagnet M., Royer A., Sallaberry C. (2013) **Using the semantics of texts for information retrieval: a concept- and domain relation-based approach**. In: New Trends in Databases and Information Systems Advances in Intelligent Systems and Computing, vol. 241, p.257-266. Genoa, Italy. ([CORE B](#))
11. Buscaldi D., Dessì D., Motta E., Osborne F., Reforgiato Recupero D. **Mining Scholarly Data for Fine-Grained Knowledge Graph Construction**. Deep Learning for Knowledge Graph workshop, DL4KG@ESWC 2019, pp.21-30, Portorož, Slovenia
12. Di Rocco L., Buscaldi D., Bertolotto M., Catania B., Guerrini G. **The role of geographic knowledge in sub-city level geolocation**. Proceedings of the ACM SAC 2019 conference, Pages 687–689, Limassol, Cyprus 2-7 April 2019
<https://doi.org/10.1145/3297280.3297557>
13. Gabor K., Buscaldi D., Schumann A.-K., QasemiZadeh B., Zargayouna H., Charnois T. **SemEval-2018 Task 7: Semantic Relation Extraction and Classification in Scientific Papers**. Proceedings of SemEval2018, co-located with NAACL-HLT 2018, New Orleans, USA, 4th June 2018
14. Gabor K., Zargayouna H., Tellier I., Buscaldi D., Charnois T. (2016) **A Typology of Semantic Relations Dedicated to Scientific Literature Analysis**. In: SAVE-SD Workshop. Montreal, Canada.
15. Gabor K., Zargayouna H., Buscaldi D., Tellier I., Charnois T. (2016) **Semantic Annotation of the ACL Anthology Corpus for the Automatic Analysis of Scientific Literature**. In: LREC 2016. Portoroze, Slovenia. ([CORE C](#))
16. Buscaldi D., Hernández-Farías I. (2015) **Sentiment Analysis on Microblogs for Natural Disasters Management: a Study on the 2014 Genoa Floodings**. In: Proceedings of the 3rd International Workshop on Social Web for Disaster Management (SWDM'15), co-located with the WWW'15 conference. pp.1185-1188. Firenze, Italia. [link](#)
17. Cleuziou G., Buscaldi D., Dias G., Levorato V. **A Pretopological Framework for Automatic Construction of Semantic-Lexical Structures from Texts**. In: Proc. of the 20th ACM Conference on Information and Knowledge Management (CIKM 2011), pp. 2453-2456, Glasgow (UK), October 24 - 28, 2011 ([CORE A](#))
18. Buscaldi, D., Rosso, P. (2008) **Geo-WordNet: Automatic Georeferencing of WordNet**. In: Proc. 5th Int. Conf. on Language Resources and Evaluation, LREC-2008, Marrakech, Morocco, May 2008 ([CORE C](#))

c.) As Editor:

- Mehwish Alam, Davide Buscaldi, Michael Cochez, Francesco Osborne, Diego Reforgiato Recupero, Harald Sack: **Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG2019)** Co-located with the 16th Extended Semantic Web Conference 2019 (ESWC 2019), Portoroz, Slovenia, June 2, 2019. CEUR Workshop Proceedings 2377, CEUR-WS.org 2019
- Davide Buscaldi, Aldo Gangemi, Diego Reforgiato Recupero: **Semantic Web Challenges - 5th SemWebEval Challenge at ESWC 2018, Heraklion, Greece, June 3-7, 2018, Revised Selected Papers.** Communications in Computer and Information Science 927, Springer 2018, ISBN 978-3-030-00071-4
- Diego Reforgiato Recupero, Mauro Dragoni, Davide Buscaldi, Mehwish Alam, Erik Cambria: **Proceedings of 4th Workshop on Sentic Computing, Sentiment Analysis, Opinion Mining, and Emotion Detection (EMSASW 2018)** Co-located with the 15th Extended Semantic Web Conference 2018 (ESWC 2018), Heraklion, Greece, June 4, 2018. CEUR Workshop Proceedings 2111, CEUR-WS.org 2018

Teaching

- Ecole Polytechnique, course INF-582: Natural Language Processing, Master M1 CS Engineering
- Ecole Polytechnique, course INF-554: Introduction to Machine Learning, Master M1 CS Engineering
- Ecole Polytechnique, course CSE-204 (Bachelor program): Machine Learning
- Université Sorbonne Paris Nord, “Statistical Programming 1 and 2” Data Science department, IUT de Villetaneuse (since 2019)
- Université Sorbonne Paris Nord, “Introduction to Information Systems”, “Principles of Operating Systems”, “Systems and Computer Networks” at the IUT of Villetaneuse, Computer Science Department, Université Paris 13 (since 2012)

Grants and Awards

- Best poster at the 5th Franco-Polish forum of Innovation (2019)
- [Visiting professor](#) grant from the University of Cagliari, May 2017
- 2009: Grant (“beca EFPI”) by the local administration of the Comunidad Valenciana (Spain) for a 3 months internship at the HLT Group at Fondazione Bruno Kessler (FBK), Trento, Italy (European doctorate program)
- 2006: 4 years grant (“beca FPI”) by the local administration of the Valencia regional government for the completion of the Ph.D. Thesis
- 2003: Grant for the “Training of researcher in highly qualified foreign research institutions” (Addestramento di Ricercatori in centri stranieri ad alta qualificazione, regional decree n.78) by the local administration of the Liguria region (Italy)