

Pierpaolo Basile

Department of Computer Science, University of Bari Aldo Moro
Via E. Orabona, 4, 70125 Bari – ITALY

www.di.uniba.it/~swap/index.php?n=Membri.Basile • [in pierpaolobasile](#)
[pippokill](#)

Current Position

Assistant Professor (type B researcher) at the Department of Computer Science, University of Bari Aldo Moro. ORCID: 0000-0002-0545-1105.

Curriculum Studiorum

Education.....

August 2017 National Academic Qualification as Associate Professor (01/B1 - Computer Science)

April 2017 National Academic Qualification as Associate Professor (09/H1 - Information Processing Systems)

May 2009 Ph.D. in Computer Science at Department of Computer Science, University of Bari Aldo Moro, Italy. Title :“Word Sense Disambiguation and Intelligent Information Access”. *Advisor: Professor Giovanni Semeraro.*

July 2005 “Laurea in Informatica Vecchio Ordinamento” (equivalent to a Master degree in Computer Science) at Department of Computer Science, University of Bari Aldo Moro, Italy. Title: “JIGSAW: un algoritmo di disambiguazione diversificato per ogni categoria grammaticale” (JIGSAW: a word sense disambiguation algorithm with different approaches for each part-of-speech). *Supervisor: Professor Giovanni Semeraro.*

Visiting, Internship and Training courses.....

March 2018-April 2018 visiting researcher at the Alan Turing Institute (London, UK) for the project: “Diachronic Analysis of Language”.

2011 InnovAction Lab Puglia - Regional final 2011. I participated in the regional final of InnovAction Lab Puglia with “Shearch”. Our project aimed at developing a full-content search functionality for cloud-based storage services.

May 2008-July 2008 internship at the IXA Research Group, Department of Computer Science, University of Basque Country, under the supervision of the Professor Eneko Agirre. The aim of this collaboration was the combination of several word sense disambiguation approaches by exploiting meta learning strategies.

2007 Participation in the 3rd BiCi International PhD School on Advanced Retrieval and Web Mining, Varenna (CO) - Italy.

2006 Participation in the 18th European Summer School In Logic Language and Information

(ESSLLI 2006), University of Malaga - Spain.

Professional Experience

January 2016 - December 2018 Assistant Professor (type A researcher) at the Department of Computer Science, University of Bari Aldo Moro, project title: "Multilingual Entity Linking".

July 2013-January 2016 Post-Doctoral Researcher (Assegno di Ricerca - INF/01) at the Department of Computer Science, University of Bari Aldo Moro, project title: "Compositional operators in Distributional Semantic Models".

June 2009-June 2013 Post-Doctoral Researcher (Assegno di Ricerca - INF/01) at the Department of Computer Science, University of Bari Aldo Moro, project title: "Methods and techniques for the semantic indexing of textual documents".

December 2008-May 2009 Employment Contract with Department of Computer Science, University of Bari Aldo Moro within the regional project DIPIS for the evaluation of a recommender system.

October 2005 Employment Contract with Department of Computer Science, University of Bari Aldo Moro within the European project VIKEF for the development of a content-based recommender system.

Scientific Activity

- Natural Language Processing
 - Word Sense Disambiguation and Entity Linking
 - Distributional Semantic Models and Compositional Semantics
 - Diachronic Analysis
 - Statistical methods for Natural Language Processing
 - Sentiment Analysis
- Intelligent Information Access
 - Natural Language Processing for Information Retrieval
 - Information Filtering
 - Recommender Systems
 - Conversational Recommender Systems
 - Machine Learning Techniques for Recommender Systems

Research Projects

- European Project
 - VIKEF** Virtual Integrated Knowledge Environment Framework. I participated in the project with the role of developing a Conference Participant Advisor, a service aimed at providing personalized information in the Scientific Congress Scenario.

DELOS European Network of Excellence on Digital Libraries. I participated in the project with the role of developing personalized services for Digital Libraries.

o National Project

MAIVISTO Massive Adaptive Internet Video Streaming using the cloud. The aim of the project is the development of an adaptive streaming framework for delivering video contents using a cloud platform. I participated in the project with the role of designing and developing a semantic recommender system for video contents.

o Regional Project

HEALTHNET - Apulian ICT Living Lab The aim of the project is the development of an integrated platform able to coordinate and monitor both the access and the delivery of health and social services. I participated in the project with the role of developing a semantic content-based recommender system able to suggest similar patients, doctors, clinics or therapies adopted by other patients.

Ask2me Question answering systems for supporting the customer care. The aim of the project is the development of a question answering system in the context of a customer care platform. I participated in the project with the role of developing the natural language processing pipeline for both Italian and English.

AskPuglia The aim of the project is the development of a framework for the evaluation of natural language processing technologies into regional web sites. I participated in the project with the role of designing the system architecture and developing a crawler able to extract information from regional web sites.

Natural Browsing The aim of the project is the development of a framework able to automatically build a knowledge base from unstructured data. I participated in the project with the role of designing the knowledge extraction process from unstructured data.

DIPIS Distributed Production as Innovative System. The aim of the project is to provide a framework for the development of distributed systems. I participated in the project with the role of designing intelligent methods for user personalization.

DDTA Apulian Textile-Clothing Digital District. DDTA has the goal of digitalizing the regional districts of textiles and clothing production. I participated in the project with the role of designing a filtering scenario to support the textiles and clothing production.

JUMP Performance Support System for dynamic organizations. I participated in the project with the role of designing the component able to extract concepts from the user query.

@SSO_NET Virtual Community of Small and Medium Enterprises and Apulian Artisans. I participated in the project with the role of developing a semantic recommender system for delivering personalized textual documents.

Participation in Evaluation Campaigns.....

MultiLing-2017 Single Document Summarization, *1st position in the final rank*

Kaggle The Allen AI Science Challenge - Team: A pure logical approach (5th position in the public leaderboard).

#Microposts2015 Making Sense of Microposts - Named Entity Recognition and Linking (NEEL) Challenge

SemEval-2015 Sentiment Analysis in Twitter

- SemEval-2015** Multilingual All-Words Sense Disambiguation and Entity Linking
- EVALITA 2014** SENTiment POLarity Classification (SENTIPOLC), *1st position in the final rank*
- SemEval-2014** Cross-Level Semantic Similarity
- ESWC 2014** Linked Open Data-Enabled Recommender System Challenge: *1st position in the final rank for the Top-N recommendation from binary user feedback task*
- *SEM 2013** shared task on Semantic Textual Similarity
- SemEval-2012** Semantic Textual Similarity
- EVALITA 2011** Evaluation of NLP and Speech Tools for Italian 2011 - Super Sense Tagging, *1st position in the final rank*
- SemEval-2 Task 2** Cross-Lingual Lexical Substitution, Uppsala (Sweden), July 15-16, 2009, *1st position in the final rank*
- EVALITA 2009** Evaluation of NLP and Speech Tools for Italian 2009 - Lexical Substitution Task, *1st position in the final rank*
- CLEF-2009** Ad-hoc Robust WSD Task, *1st position in the final rank of one of the four organized subtasks*
- CLEF-2008** Ad-hoc Robust WSD Task
- EVALITA 2007** Evaluation of NLP and Speech Tools for Italian 2007 - WSD All-Word Task, *1st position in the final rank*
- SemEval-1 Task 1** Evaluating Word Sense Disambiguation on Cross-Language Information Retrieval

Academic Service.....

Chair

- 1st International Workshop on REbooting the COnVERSational Recommender Systems (RECOVER 2018), co-located at ACM RecSys 2018
- 2nd Workshop on Natural Language for Artificial Intelligence (NL4AI 2018), co-located at AI*IA 2018 17th International Conference of the Italian Association for Artificial Intelligence
- 1st Workshop on Natural Language for Artificial Intelligence (NL4AI 2017), co-located at AI*IA 2017 16th International Conference of the Italian Association for Artificial Intelligence
- 1st Workshop on Temporal Dynamics in Digital Libraries (TDDL 2017), co-located at TPD 2017
- EVALITA 2016, the 5th evaluation campaign of NLP and Speech Tools for Italian

Co-editor

- Proceedings of the 1st Workshop on Natural Language for Artificial Intelligence (NL4AI 2017), co-located with the 16th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2017), Bari, Italy, November 16-17, 2017. Edited by: Pierpaolo Basile, Danilo Croce, Marco Guerini, CEUR-WS Vol-1983.
- Proceedings of Third Italian Conference on Computational Linguistics & Evaluation Campaign of Natural Language Processing and Speech Tools for Italian Final Workshop (CLiC-it &

EVALITA 2016), Napoli, Italy, December 5-7, 2016. Edited by: Pierpaolo Basile, Anna Corazza, Franco Cutugno, Simonetta Montemagni, Malvina Nissim, Viviana Patti, Giovanni Semeraro, Rachele Sprugnoli, CEUR-WS Vol-1749.

PC Member ISWC 2018 Posters and Demos, EMNLP 2018, IJCAI 2018, ACL 2018, NLDB2018, Emotions@AISB 2018, ICSC 2018, ISWC 2017 Posters and Demos, LREC2018, EMNLP 2017, ICSC 2017, NLDB 2017, ESWC 2017Posters&Demos, WWW2016 Poster Track, NLDB 2016, LD4IE2016, ISWC 2016 Posters and Demos, ICSC 2016, ESWC 2016, DART 2016, CLIC-it 2016, CIKM 2016, Microposts2016, DART 2015, EC-Web 2015, NLDB 2015, ESWC 2015, Microposts2015, ICSC 2015, CLIC-it 2014, DART 2014, LD4IE 2014, *SEM 2014, ESWC 2014, Microposts2014, LD4IE2013, ISWC 2014 Posters and Demos, ICSC 2014, IIR 2014, ESWC 2013, EC-Web 2013, Microposts2013, IIR 2013, DART2012, DiDaS 2012, EC-Web 2012, Microposts2012, EC-Web 2011, DART 2011, EC-Web 2010, EC-Web 2009, HIS 2010

Organizer

- o Tutorial on Linked Open Data for Semantics-aware Recommender Systems at ESWC 2017
- o NEEL-IT: 1st Named Entity rEcognition and Linking in Italian Tweets task at EVALITA 2016
- o Laboratorio dal Basso - Semantics: theory, technologies and applications http://www.questioncube.com/ldb_semantics/
- o Intelligent Information Access - IIA 2008. Cagliari, Italy, December 9-11, 2008

Local Organizer

- o IIR2012: 3rd Italian Information Retrieval workshop, Bari, Italy, January 26-27, 2012
- o SWAP 2007: 4th Italian Workshop on Semantic Web. Bari, Italy, December 18-20, 2007

Journal Review International Journal on Semantic Web and Information Systems, ACM Transactions on Information Systems, Knowledge and Information Systems Journal, Journal of Intelligent Information Systems, Information Sciences Journal

Additional Review ECIR 2018, ECIR 2016, RecSys (2012, 2011, 2010, 2009), ECIR (2012, 2011, 2010, 2009, 2008), LREC (2012, 2010), AI*IA 2013, AI*IA 2011, K-CAP 2011, ,EC-WEB 2009, ISDA 2009, K-CAP 2009, PAKDD 2009, SWAP 2007, WSC 2008, KES 2007, ECML/PKDD 2007, SIGIR 2007, SEMEVAL 2007 (1st Workshop on SEMantic EVALuation), DAWAK 2007 (10th International Conference on Data Warehousing and Knowledge Discovery), HIS 2007, ADMA 2006 , ECML/PKDD 2006, ISMIS 2006

Seminars

- 04/04/2017** "Introduction to Natural Language Processing and Diachronic Analysis", PhD School in Arts and Literature, Italianist and comparative cultures, University of Bari Aldo Moro.
- 05/02/2015** "Linked Open Data-enabled Strategies for Top-N Recommendations", PhD School in Information Technology and Electrical Engineering, University of Naples Federico II.
- 23/07/2012** "Accesso alle informazioni con Apache Lucene" (Information Access with Apache Lucene), F5 Hit Refresh Open Innovation Camp on Natural Language Processing e Media Screening, San Vito dei Normanni (BR).
- 24/05/2010-25/05/2010** "Sostituzione lessicale in corpora multilingua" (Multilingual Lexical Substitution), University of Bari Aldo Moro - Master in Technical-Scientific Translation.

18/12/2008 “Open Source for Intelligent Information Access”, at “Conferenza Open Source” (Open Source Conference), Politecnico di Bari.

Skills

Main Computing Skills

Programming languages JAVA, C, Arduino, Python, PASCAL, PROLOG, BASIC, 6502 Assembly, x86 Assembly.

Operating systems Linux, Windows.

General skills word2vec, SemanticVectors, Liblinear, SVMlib, SVDLIB, WEKA, TensorFlow, scikit-kg, XML, RDF, JSP, LaTeX, SQL, SPARQL.

IR & NLP tools Apache Lucene, Apache Tika, WordNet, BabelNet, MaltParser, StanfordCoreNLP, Apache OpenNLP, MyMediaLite.

GitHub <https://github.com/pippokill>

Software

SentIT: a system for the detection of subjectivity and polarity in tweets. The system was developed for the Italian language and reports the best performance in the EVALITA 2014 evaluation campaign [?].

TRI: a system for the analysis of meaning variation over time. The system implements the algorithm described in [?]. TRI is freely available on **github**: <https://github.com/pippokill/tri>.

TEE: a system for the retrieval of temporal events in a large collection of documents. The system implements the algorithm described in [?]. TEE is freely available on **github**: <https://github.com/pippokill/TAIR>.

Lesk-DSM-WSD: a system for Word Sense Disambiguation that exploits an enhanced version of the Lesk algorithm based on DSMs. The algorithm is described in [?]. Lesk-DSM-WSD is freely available on **github**: <https://github.com/pippokill/lesk-wsd-dsm>.

JIGSAW: a knowledge-based Word Sense Disambiguation developed during the Ph.D. The algorithm is described: [?, ?, ?, ?]. JIGSAW is freely available on **github**: <https://github.com/pippokill/JIGSAW>.

Foreign Languages

Italian mother tongue.

English good (spoken/written).

Teaching and Supervision

Teaching

2019-2020/2020-2021 Professor for the course “Metodi Avanzati di Programmazione” (Advanced Programming Methods) at the Department of Computer Science, University of Bari Aldo Moro.

2018-2019 Professor for the course “Algoritmi e Strutture Dati” (Algorithms and Data Structures) at the Department of Computer Science, University of Bari Aldo Moro.

2017-2018/2016-2017 Professor for the course “Metodi per il Ritrovamento dell’Informazione” (Information Retrieval) at the Department of Computer Science, University of Bari Aldo Moro.

2012-2013 Adjoin professor for the course “Gestione della Conoscenza d’Impresa” (Knowledge Management) at the Department of Computer Science, University of Bari Aldo Moro.

2011-2012 Adjoin professor for the course “Accesso Intelligente all’Informazione - B” (Intelligent Information Access - B) at the Department of Computer Science, University of Bari Aldo Moro.

2009-2010 Adjoin professor for the course “Gestione della Conoscenza d’Impresa” (Knowledge Management) at the Department of Computer Science (Brindisi), University of Bari Aldo Moro.

Teaching support and tutoring activities.....

2015-2016 Tutoring activities within the course “Linguaggi di Programmazione” (Programming Languages) at the Department of Computer Science, University of Bari Aldo Moro.

2009-2015 Teaching support and tutoring activities within the course “Accesso Intelligente all’Informazione” (Intelligent Information Access) at the Department of Computer Science, University of Bari Aldo Moro.

2006-2007/2014-2015 Teaching support and tutoring activities within the course “Linguaggi di Programmazione” (Programming Languages) at the Department of Computer Science, University of Bari Aldo Moro.

2007-2008/2009-2010/2011-2012 Teaching support and tutoring activities within the course “Gestione della Conoscenza di Impresa” (Knowledge Management) at the Department of Computer Science, University of Bari Aldo Moro.

Theses supervision.....

2006-Present I have been a supervision team member for 40 theses in Computer Science (Master/Bachelor) about Information Retrieval, Natural Language Processing and Recommender System.

Awards

Alan Turing Institute - Visiting Researcher Application My application about the “Diachronic Analysis of Language” has been accepted as visiting research project at the Alan Turing Institute.

Future in Research Author of the project “Multilingual Entity Linking” (cod. IAVIN27) funded by the Apulia Region (Future In Research program) for the funding of 170 fixed-term research positions in the Apulia universities.

CLiC-it 2014 Best Young Paper Award “Analysing Word Mining over Time by Exploiting Temporal Random Indexing” assigned by the First Italian Conference on Computational Linguistics (CLiC-it 2014).

Linked Open Data-Enabled Recommender System Challenge - ESWC 2015 first system in the Top-N recommendation from binary user feedback task.

Distinguished Ph.D. theses 2009-2010 assigned by the Special Interest Group on Natural Language Processing of the Italian Association for Artificial Intelligence.

IJCAI2009 Grant awarded for attending the *IJCAI 2009 - 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA*

BiCi 2007 Grant awarded for attending the *3rd BiCi International Ph.D. School on Advanced Retrieval and Web Mining. Varenna (CO), August 26-31, 2007*

ECIR2007 Grant awarded for attending the *European Conference on Information Retrieval 2007 - ECIR 2007*

January 2006-December 2008 Fellowship for the Ph.D. School of Computer Science of the University of Bari.

Technology Transfer

Research contract.....

September 2017-September 2018 I am responsible of the research contract between the Digital Transformation Team (<https://teamdigitale.governo.it/en/>) and the Department of Compute Science (University of Bari Aldo Moro). The contract has the aim of developing a question answering system for Public Administrations.

Participation in Innovative Startups.....

2013 - March 2017 Co-founder of Zenfeed (<http://www.zenfeed.me/>). Zenfeed is a smart news reader able to sort news according to user interests. I was responsible of the research unit. Zenfeed was financed by the Principi Attivi 2012 program of the Apulia Region. Zenfeed won the business competition "Valore Assoluto" organized by the Camera di Commercio of Bari. Zenfeed received an early stage funding by United Ventures and LVenture Group. Zenfeed has been participated in the 2014 Startup School organized by the Mind the Bridge Foundation at San Francisco.

2011 Co-founder of QuesitonCube (<http://www.questioncube.com/>). QuestionCube develops solutions for natural language processing, semantic search and question answering. I am the CTO and responsible for the research unit about natural language processing. QuestionCube won the Lamarck prize for young innovative startups, and was one of the winners of the Apps4Italy competition organized by the Forum PA.

Awards and Funding for technology transfer activities.....

- Seed funding from United Ventures and LVenture Group for the Zenfeed start-up
- Mind The Bridge Start-up School award
- Acceleration Program at LUISS ENLABS for the Zenfeed start-up
- Lamarck Award for young entrepreneurs (SMAU 2013) with the QuestionCube start-up
- Valore Assoluto 2012 Award (business plan competition)
- Principi Attivi 2012 Award (business plan competition)

Associations

- AILC, Italian Association for Computational Linguistics
- AI*IA, Italian Association for the Artificial Intelligence
- SIGLEX, Special Interest Group on the Lexicon of the Association for Computational Linguistics
- Alumni Mathematica (<http://alumnimathematica.org/>). It is a nonprofit organization for the mathematical sciences dissemination and the independent scientific research. It organizes activities for promoting mathematical sciences, innovative teaching in schools and scientific research in collaboration with industrial partners. I have organized several courses about Arduino and Robotics for higher schools.

Publications

- Google Scholar: <https://scholar.google.it/citations?user=dNznumkAAAAJ&hl=it>
- DBLP: <https://dblp.org/pid/98/5082.html>
- Scopus: <https://www.scopus.com/authid/detail.uri?authorId=23392182500>