

**Allegato alla domanda di partecipazione
Curriculum formativo, didattico, scientifico e professionale del candidato**

Dichiarazione sostitutiva di certificazioni

(Art. 46, D.P.R. 28 dicembre 2000 n. 445)

Dichiarazione sostitutiva dell'atto di notorietà

(da sottoscrivere davanti all'impiegato addetto o da presentare o spedire con la fotocopia di un documento di identità)

(Art. 47, D.P.R. 28 dicembre 2000 n. 445)

Estremi del bando di selezione	Assegno di ricerca "Algoritmi per l'accesso semplificato agli Open Data", R.S. Prof. Maurizio Atzori, D.R. n. 255/2020 del 28/02/2020
Informazioni aggiornate al	07/06/22
Nome e Cognome	Manuela Sanguinetti
Data di nascita	15/09/83

Esperienza professionale

Periodo	Ente	Principali attività e responsabilità
01/06/20 – In corso	Università degli Studi di Cagliari	Assegno di ricerca Titolo: <i>User-Friendly Access Algorithms to Open Data</i>
03/03/21 – 30/09/21	Università degli Studi di Cagliari	Tutor di laboratorio Insegnamento: <i>Programmazione II</i> Corso di Laurea Triennale in Informatica
21/02/20 – 31/05/20	Università degli Studi di Torino	Borsa di ricerca Titolo: <i>Natural Language Technologies for Intelligent Conversational Interfaces: Explanation and Emotions in Human-Chatbot Conversations</i>
01/08/18 - 01/01/20	Università degli Studi di Torino	Assegno di ricerca Titolo: <i>Data Annotation and Analysis for Developing Hate Speech Detection Systems and Methodologies</i>
01/07/17 - 01/07/18	Università degli Studi di Torino	Assegno di ricerca Titolo: <i>Developing Linguistic Resources for Sentiment Analysis Oriented to Hate Speech Detection</i>
01/08/16 - 31/03/17	Università di Pisa	Borsa di ricerca Titolo: <i>Development of a Parallel Multilingual Dependency Treebank in the Universal Dependency Format</i>

Istruzione, formazione (es. titoli di studio, certificazioni professionali/linguistiche/informatiche)

Data	Titolo / Principali tematiche	Ente
19/09/16	Dottorato di Ricerca in Informatica Titolo della tesi: <i>Combining Approaches for a Parallel Treebank Development and Alignment</i> ▪ 02/2014 - 05/2014: Periodo di ricerca presso l'Istituto di Linguistica Formale e Applicata (Charles	Università degli Studi di Torino

	University, Praga, Repubblica Ceca), finanziato dal programma Erasmus Placement <ul style="list-style-type: none"> ▪ 08/2013: European Summer School in Logic, Language and Information (ESLLI), Dusseldorf, Germany 03/2012: Bertinoro International Spring School (BISS), Bertinoro (FC), Italy 	
14/07/11	Laurea Specialistica in Traduzione	Università degli Studi di Torino
08/07/08	Laurea Triennale in Lingue e Culture Europee ed Extraeuropee	Università degli Studi di Cagliari

Publicazioni / Convegni

Publicazioni in Rivista:

- A. Mazzei, L. Anselma, M. Sanguinetti, A. Rapp, D. Mana, M. M. Hossain, V. Patti, R. Simeoni, L. Longo. Anticipating User Intentions in Customer Care Dialogue Systems. *IEEE Transactions on Human-Machine Systems* (in stampa), 2022
- E. Di Nuovo, M. Sanguinetti, A. Mazzei, C. Bosco, E. Corino. VALICO-UD: Treebanking an Italian Learner Corpus in Universal Dependencies. *IJCoL. Italian Journal of Computational Linguistics* (in stampa), 2022
- M. Sanguinetti, C. Bosco, L. Cassidy, Ö. Çetinoğlu, A. T. Cignarella, T. Lynn, I. Rehbein, J. Ruppenhofer, D. Seddah, A. Zeldes. Treebanking user-generated content: a UD based overview of guidelines, corpora and unified recommendations. *Language Resources and Evaluation*, 2022
- F. Poletto, V. Basile, M. Sanguinetti, C. Bosco, V. Patti. [Resources and benchmark corpora for hate speech detection: a systematic review](#). *Language Resources and Evaluation*, 2020
- A. T. E. Capozzi, M. Lai, V. Basile, F. Poletto, M. Sanguinetti, C. Bosco, V. Patti, G. Ruffo, C. Musto, M. Polignano, G. Semeraro, M. Stranisci. [“Contro L’Odio”: A Platform for Detecting, Monitoring and Visualizing Hate Speech against Immigrants in Italian Social Media](#). *IJCoL. Italian Journal of Computational Linguistics*, 2020

Publicazioni per Conferenze e Workshop:

- A. Cignarella, M. Lai, A. Marra, M. Sanguinetti. “La ministro è incinta”: A Twitter account of women’s job titles in Italian. In Proceedings of the Eighth Italian Conference on Computational Linguistics (CliC-It 2021), 2021.
- M. Sanguinetti, C. Bosco, L. Cassidy, Ö. Çetinoğlu, A. T. Cignarella, T. Lynn, I. Rehbein, J. Ruppenhofer, D. Seddah, A. Zeldes. *Treebanking User-Generated Content: A Proposal for a Unified Representation in Universal Dependencies*. Accepted paper at the 12th Conference on Language Resources and Evaluation (LREC 2020), 2020.
- A. T. Cignarella, M. Sanguinetti, C. Bosco, P. Rosso. *Marking Irony Activators in a Universal Dependencies Treebank: The Case of an Italian Twitter Corpus*. Accepted paper at the 12th Conference on Language Resources and Evaluation (LREC 2020), 2020.
- A. T. E. Capozzi, M. Lai, V. Basile, C. Musto, M. Polignano, F. Poletto, M. Sanguinetti, C. Bosco, V. Patti, G. Ruffo, G. Semeraro, and M. Stranisci. *Computational linguistics against hate: Hate speech detection and visualization on social media in the “Contro l’Odio” project*. In Proceedings of the Sixth Italian Conference on Computational Linguistics (CliC-it 2019), 2019

- A. T. Cignarella, M. Sanguinetti, C. Bosco, and P. Rosso. *Is this an effective way to annotate irony activators?* In Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019), 2019
- C. Francesconi, C. Bosco, F. Poletto, and M. Sanguinetti. *Error analysis in a hate speech detection task: The case of HaSpeeDe-TW at EVALITA 2018.* In Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019), 2019
- E. Di Nuovo, C. Bosco, A. Mazzei, and M. Sanguinetti. *Towards an Italian learner treebank in Universal Dependencies.* In Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019), 2019
- V. Basile, C. Bosco, E. Fersini, D. Nozza, V. Patti, F. M. Rangel Pardo, P. Rosso, and M. Sanguinetti. *Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in Twitter.* In Proceedings of the 13th International Workshop on Semantic Evaluation. Association for Computational Linguistics, 2019
- V. Basile, M. Lai, and M. Sanguinetti. *Long-term social media data collection at the University of Turin.* In Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), 2018
- C. Bosco, F. Dell'Orletta, F. Poletto, M. Sanguinetti, and M. Tesconi. *Overview of the EVALITA 2018 Hate Speech Detection task.* In Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018), 2018
- M. Sanguinetti, C. Bosco, A. Lavelli, A. Mazzei, Oronzo Antonelli, and F. Tamburini. *PoSTWITA-UD: an Italian Twitter treebank in Universal Dependencies.* In Proceedings of LREC 2018, Miyazaki, Japan. European Language Resources Association (ELRA), 2018.
- M. Sanguinetti, F. Poletto, C. Bosco, V. Patti, and M. Stranisci. *An Italian Twitter corpus of hate speech against immigrants.* In Proceedings of LREC 2018, Miyazaki, Japan. European Language Resources Association (ELRA), 2018
- F. Poletto, M. Stranisci, M. Sanguinetti, V. Patti, and C. Bosco. *Hate speech annotation: Analysis of an Italian Twitter corpus.* In Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-it 2017), Rome, Italy, 2017
- Daniel Zeman, et al. *Conll 2017 shared task: Multilingual parsing from raw text to Universal Dependencies.* In Proceedings of the CoNLL 2017 Shared Task, Vancouver, Canada. Association for Computational Linguistics, 2017
- M. Sanguinetti, C. Bosco, A. Mazzei, A. Lavelli, and F. Tamburini. *Annotating Italian social media texts in Universal Dependencies.* In Proceedings of the 4th Conference on Dependency Linguistics (DepLing2017), Pisa, Italy. Linkoping University Electronic Press, 2017
- E. Sulis, M. Lai, M. Vinai, and M. Sanguinetti. *Exploring sentiment in social media and official statistics: a general framework.* In Proceedings of the 2nd Workshop on Emotion and Sentiment in Social and Expressive Media (ESSEM 2015), Istanbul, Turkey, 2015
- M. Sanguinetti. *Experimenting the use of catenae in Phrase-Based SMT.* In Proceedings of the Second Italian Conference of Computational Linguistics (CLiC-it 2015), Trento, Italy, 2015
- C. Bosco, L. Allisio, V. Mussa, V. Patti, G. Ruffo, M. Sanguinetti, and E. Sulis. *Detecting happiness in Italian tweets: Towards an evaluation dataset for sentiment analysis in Felicitta.* In Proceedings of the International Workshop on Emotion, Social Signals, Sentiment and Linked Open Data (ES3LOD 2014), Reykjavik, Iceland. European Language Resources Association (ELRA), 2014
- C. Bosco and M. Sanguinetti. *Converting the parallel treebank ParTUT in Universal Stanford Dependencies.* In Proceedings of the First Italian Conference of Computational Linguistics (CLiC-

it), Pisa, Italy, 2014

- C. Bosco and M. Sanguinetti. *Towards a Universal Stanford Dependencies parallel treebank*. In Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13), Tübingen, Germany, 2014

- C. Bosco, F. Dell'Orletta, S. Montemagni, M. Sanguinetti, and M. Simi. *The Evalita 2014 dependency parsing task*. In Proceedings of Evalita 2014. Evaluation of Natural Language and Speech Tools for Italian. Pisa University Press, 2014

- M. Sanguinetti, E. Sulis, V. Patti, G. Ruffo, L. Allisio, V. Mussa, and C. Bosco. *Developing corpora and tools for sentiment analysis: the experience of the University of Turin group*. In Proceedings of the First Italian Conference of Computational Linguistics (CLiC-it), Pisa, Italy, 2014.

- M. Sanguinetti, C. Bosco, and L. Cupi. *Exploiting catenae in a parallel treebank alignment*. In Proceedings of the 9th Conference on Language Resources and Evaluation (LREC2014), Reykjavik, Iceland. European Language Resources Association (ELRA), 2014

- M. Sanguinetti, C. Bosco, and L. Lesmo. *Dependency and constituency in translation shift analysis*. In Proceedings of the 2nd Conference on Dependency Linguistics (DepLing2013), Prague, Czech Republic. MATFYZPRESS/ACL, 2013

- M. Sanguinetti and C. Bosco. *Translational divergences and their alignment in a parallel multilingual treebank*. In Proceedings of the 11th Workshop on Treebanks and Linguistic Theories (TLT11), Lisbon, Portugal, 2012

- C. Bosco, M. Sanguinetti, and L. Lesmo. *The parallel-TUT: a multilingual and multiformat treebank*. In Proceedings of the 8th Conference on Language Resources and Evaluation (LREC2012), Istanbul, Turkey. European Language Resources Association (ELRA), 2012

- M. Sanguinetti, C. Bosco, *Building the Multilingual TUT Parallel Treebank*. In Proceedings of the 2nd Workshop on Annotation and Exploitation of Parallel Corpora (AEPC2), Hissar, Bulgaria, 2011

Capitoli:

- M. Sanguinetti and C. Bosco. *ParTUT: The Turin University Parallel Treebank*. In Italian Natural Language Processing within the PARLI Project, volume 589 of Studies in Computational Intelligence, pages 51–69. Springer, 2015

Curatela:

- A. Lieto, C. Battaglino, D. P. Radicioni, and M. Sanguinetti, editors. *Proceedings of the 3rd International Workshop on Artificial Intelligence and Cognition*, Turin, Italy, September 28-29, 2015, volume 1510 of CEUR Workshop Proceedings. CEUR-WS.org,2015

- L. Di Caro, C. Dodaro, A. Loreggia, R. Navigli, A. Perotti, and M. Sanguinetti, editors. *Proceedings of the Second Doctoral Workshop in Artificial Intelligence (DWAi 2014)*. An official workshop of the 13th Symposium of the Italian Association for Artificial Intelligence (AI*IA 2014), Pisa, Italy, December 11, 2014, volume 1334 of CEUR Workshop Proceedings. CEUR-WS.org, 2014

Altre attività scientifiche

Partecipazione a progetti:

- *High quality Open data Publishing and Enrichment (HOPE)*. Bando PRIN MIUR 2017

- *Tecnologie del Linguaggio Naturale per Interfacce Conversazionali Intelligenti: Spiegazione e*

Emozioni nelle Conversazioni Human-Chatbot. Convenzione di Ricerca con TIM S.p.A.
- *HatePrejudice. Immigrants, Hate and Prejudice in Social Media*. Finanziato dalla Fondazione Compagnia San Paolo, Torino
- *Hate Speech and Social Media. Analysis of Hate Speech Contents in Piedmont Region*. Finanziato dalla Fondazione CRT, Torino

Comitati:

Comitati Scientifici:

- Dal 2020: Scientific committee della Conference on Language Resources and Evaluation (LREC)
- Program committee del Workshop on Resources and Techniques for User and Author Profiling in Abusive Language (ResT UP, ResT-UP2)
- Program committee del Workshop on Perspectivist Approaches to Disagreement in NLP (NLPerspectives 2022)
- Track chair dell'area Machine Translation and Multilingualism all'8° Conferenza Italiana di Linguistica Computazionale (CliC-it 2021)
- Dal 2017: Program committee della Conferenza Italiana di Linguistica Computazionale (CliC-it)
- Program committee della IEEE DSAA Special Session on Sentiment, Emotion, and Credibility of Information in Social Data (SeCredISData) @DSAA 2018: International Conference on Data Science and Advanced Analytics

Comitati organizzatori:

- 2nd Hate Speech Detection (HaSpeeDe2) task a EVALITA 2020
- Multilingual detection of hate speech against immigrants and women in Twitter (HatEval) al International Workshop on Semantic Evaluation (SemEval) 2019
- Hate Speech Detection (HaSpeeDe) task at EVALITA 2018
- Third International Workshop on Artificial Intelligence and Cognition (AIC 2015), Torino
- Doctoral Consortium on Artificial Intelligence, co-located with the XIII International Symposium on Advances in Artificial Intelligence (AI*IA 2014), Pisa
- Shared task on Dependency Parsing a EVALITA 2014

Relazioni su invito:

"Introducing the ParTUT parallel treebank": relazione su invito al workshop ACor4French at TALN 2017, Orleans, Francia

Ulteriori informazioni pertinenti

Lingue:

Italiano: Madrelingua

Inglese: C1

Francese: B2/C1

Russo: A2/B1

Supervisione di tesi:

- Tesi Triennale (da discutere). Titolo: *"Set Expansion and Labeling Using Masked Language Models"*. Corso di Laurea Triennale in Informatica, Università degli Studi di Cagliari (Co-relatrice). Studente: Michela Lecca. A.A. 2021/2022
- Tesi Triennale. Titolo: *"Studio e implementazione di un sistema di videosorveglianza in Python"*

con individuazione e riconoscimento di volti e feature facciali". Corso di Laurea Triennale in Informatica, Università degli Studi di Cagliari (Co-relatrice). Studente: Leonardo Sanna. A.A. 2020/2021

- Tesi Triennale. Titolo: *"A Data-to-Text Conversational Agent for Dieting Support"*. Corso di Laurea Triennale in Informatica, Università degli Studi di Cagliari (Co-relatrice). Studenti: Matteo Gioele Collu e Federico Sanna. A.A. 2020/2021

- Tesi Triennale. Titolo: *"Strumenti per l'analisi e il miglioramento di conversazioni utente-chatbot"*. Corso di Laurea Triennale in Informatica, Università degli Studi di Torino (Co-relatrice). Studente: Marco Scalerandi. A.A. 2019/2020

Seminari:

- 19-23 Maggio 2022: *"Datasets for NLP"- "Question Answering over Knowledge Graphs"*. Insegnamento: Artificial Intelligence and Natural Language Processing, Dipartimento di Matematica e Informatica, Università degli Studi di Cagliari (Titolare del corso: Prof. Maurizio Atzori)

- Maggio 2021 - *"Datasets for NLP"*. Insegnamento: Artificial Intelligence and Natural Language Processing, Dipartimento di Matematica e Informatica, Università degli Studi di Cagliari (Titolare del corso: Prof. Maurizio Atzori)

- 29 Ottobre 2021 - *"NLP & Corpora"*. Insegnamento: Linguistica Applicata, Dipartimento di Lettere, Lingue e Beni Culturali, Università degli Studi di Cagliari (Titolare del corso: Prof. Nicoletta Puddu)

- 4 Maggio 2021 - *"Elaborazione Automatica del Linguaggio Naturale e Applicazioni nei Social Media"*: presentazione online in occasione dell'OpenDay UniCa

- 1 Giugno 2020: *"Hate Speech Detection and Corpora: An Overview of the Italian Scenario"*. Insegnamento: Artificial Intelligence and Natural Language Processing, Dipartimento di Matematica e Informatica, Università degli Studi di Cagliari (Titolare del corso: Prof. Maurizio Atzori)

- 17-18 Marzo 2015: *"Inside MT systems. Alignment issues"*. Insegnamento: Informatica Applicata alla Comunicazione Multimediale. Dipartimento di Studi Umanistici, Università degli Studi di Torino (Titolare del corso: Prof. Cristina Bosco)

- 18 Marzo 2013: *"Building a parallel multilingual treebank: a case study"*. Insegnamento: Informatica Applicata alla Comunicazione Multimediale. Dipartimento di Studi Umanistici, Università degli Studi di Torino (Titolare del corso: Prof. Cristina Bosco)

- Maggio 2012: *"Annotazione sintattica di corpora multilingue paralleli"*. Insegnamento: Informatica Applicata alla Comunicazione Multimediale. Dipartimento di Studi Umanistici, Università degli Studi di Torino (Titolare del corso: Prof. Cristina Bosco)

Cagliari, 07/06/2022

Manuela Sanguinetti
(sottoscritto con firma digitale)