



UNIVERSITÀ DEGLI STUDI DI TORINO

DICHIARAZIONE SOSTITUTIVA DI CERTIFICAZIONE

(art 46 del D.P.R. n.445 del 28/12/2000)

DICHIARAZIONE SOSTITUTIVA DELL'ATTO DI NOTORIETÀ

(artt. 19 e 47 del D.P.R. n.445 del 28/12/2000)

IL SOTTOSCRITTO

DAVIDE COLLA, Codice fiscale CLLDVD92E04C665J, nato a Chivasso (TO) il 04/05/1992 e residente in Via Croce Pietra 5, Casalborgone (TO)

**consapevole delle sanzioni penali, nel caso di dichiarazioni non veritiere e falsità negli atti, richiamate dall'art. 76 del D.P.R. 445/2000 e dalle leggi speciali in materia**

**DICHIARA**

la veridicità del sotto riportato *curriculum vitae*.



## UNIVERSITÀ DEGLI STUDI DI TORINO

Curriculum vitae.....	3
Sintesi .....	3
Aree e interessi di ricerca.....	5
Formazione, titoli .....	5
Posizioni accademiche .....	5
ATTIVITÀ DI RICERCA E PUBBLICAZIONI SCIENTIFICHE .....	6
Pubblicazioni scientifiche .....	6
Riviste .....	6
Conference proceedings .....	7
Partecipazione a progetti di ricerca .....	8
Attività in conferenze, comitati di programma, e di revisione.....	10
Reviewer di riviste e conferenze internazionali.....	10
Comitati di programma in workshop e conferenze .....	11
Presentazioni a convegni.....	11
Premi, riconoscimenti e invited talks .....	12
Collaborazione con enti esterni .....	12
Altre attività: esperienze lavorative .....	12
ATTIVITÀ DIDATTICA.....	14
Attività di co-relazione tesi.....	14
Corsi.....	15
Didattica integrava.....	15
Seminari.....	15



UNIVERSITÀ DEGLI STUDI DI TORINO

# Davide Colla

## Curriculum vitae

### Sintesi

Davide Colla è titolare di un contratto da Ricercatore a Tempo Determinato di tipo A presso il Dipartimento di Studi Storici dell'Università degli Studi di Torino dal giugno 2023. Nel 2022 consegue il titolo di dottore di ricerca in Informatica, dopo avere ricevuto la laurea magistrale in informatica nel 2017 *summa cum laude*, e laurea in informatica nel 2014. Membro del gruppo di *Intelligenza Artificiale* del Dipartimento di Informatica, e in particolare del *Content-Centered Computing research group*.

L'**attività di ricerca** condotta da Davide Colla si colloca nell'ambito dell'elaborazione automatica del linguaggio naturale. Si è concentrata sullo sviluppo e l'applicazione di strumenti per l'analisi della semantica lessicale (sviluppo di embeddings, utilizzo di word/sense embeddings, acquisizione e utilizzo di language models e transformers), sulla rappresentazione della conoscenza. Ha contribuito e contribuisce attivamente allo sviluppo di diverse risorse per la semantica lessicale e di dataset per la sperimentazione su specifici task. Come testimoniato dalle pubblicazioni e dalla partecipazione a progetti, Davide Colla ha stabilito varie collaborazioni interdisciplinari: si occupa dell'estrazione e annotazione semantica di eventi estratti da documenti d'archivio storico (Proff. Annamaria Goy e Diego Magro, Dipartimento di Informatica, Università di Torino); dell'elaborazione del linguaggio di pazienti affetti da disturbi psichiatrici (Prof. Benedetto Vitiello, Dipartimento di Scienze della Sanità Pubblica e Pediatriche, Università di Torino); di analisi ed estrazione di informazioni da record di accesso al pronto soccorso (Dott. Marco Giustini, Istituto Superiore di Sanità, Roma); di estrazione e visualizzazione di relazioni di filiazione accademica di filosofi (Proff. Guido Bonino e Paolo Tripodi, Dipartimento di Filosofia, Università di Torino); e di elaborazione di documenti giuridici (Prof. Antonio Mastropaolo, Dipartimento di Scienze economiche e politiche, Università della Valle d'Aosta; Prof. Luisa Revelli, Dipartimento di Scienze umane e sociali, Università della Valle d'Aosta).

**Pubblicazioni scientifiche.** Davide Colla è autore di lavori apparsi in atti di conferenze, workshop e riviste internazionali, e collabora con gruppi di ricerca nazionali e internazionali (complessivamente 30 co-autori, in base a dati Scopus). Partecipa inoltre all'attività di revisione e vaglio scientifico per le



## UNIVERSITÀ DEGLI STUDI DI TORINO

comunità di AI e NLP; per cui ha servito come revisore varie conferenze e riviste, fra cui Natural Language Engineering e Journal of Experimental and Theoretical Artificial Intelligence, SIGNLL Conference on Computational Natural Language Learning, 29th International Conference on Computational Linguistics (ACL 2022) e The 61st Annual Meeting of the Association for Computational Linguistics (ALC2023).

Ha **collaborazioni** con ricercatori di università straniere, in particolare della University of Groningen (NL) su temi legati allo studio/estrazione di eventi da documenti testuali, e dei nessi di causazione fra eventi (semestre all'estero durante il dottorato, nel 2020). Partecipa inoltre attivamente alle collaborazioni interdisciplinari del gruppo di ricerca con ricercatori di università italiane (Torino: Dipartimenti di Informatica, Giurisprudenza, Filosofia, Psicologia, Scienze della Sanità Pubblica e Pediatriche; Valle d'Aosta; Piemonte Orientale) istituti di ricerca (Istituto Superiore di Sanità, ASL TO3: Servizio sovrazonale di Epidemiologia) e fondazioni (Fondazione Istituto piemontese Antonio Gramsci), coautori nelle pubblicazioni di DC.

Partecipa a **progetti di ricerca** nelle aree dell'elaborazione automatica del linguaggio naturale, dell'Information Extraction e del Text Mining (*Muoversi Meglio, MuMe*, 2018-2020), con applicazioni in ambiti diversi, come Cultural Heritage e Digital Humanities (*Turin Enhanced Philosophy Tree – TEPT*, 2022-2024; *Providing Rich Semantic Metadata for Historical Archives, PRiSMHA*, 2017-2019), legale-giuridico (*L'accessibilità dei testi normativi come dispositivo di inclusione: studio di caso e strumenti applicativi in Valle d'Aosta, Prima parte*, 2021-2023; seconda tranche di finanziamento: *L'accessibilità dei testi normativi come dispositivo di inclusione: studio di caso e strumenti applicativi in Valle d'Aosta, Seconda parte*, 2023-2025) e medico-clinico (progetti *Analisi automatica delle cartelle di Pronto Soccorso per la prevenzione del femminicidio, ViDeS*, 2019-2022; progetto di public engagement 'PAUSE, Prevention of Assault Under Scientific Evidence, interdisciplinari per contrastare la violenza interpersonale', 2021-2022).

Davide Colla contribuisce alla **didattica integrativa** (bandi ex artt. 76 e 11) predisposta dai Corsi di Studio in Informatica, sia su tematiche di informatica di base (Linguaggi di Programmazione), sia su tematiche legate all'attività di ricerca (NLP). È stato co-relatore di 10 tesi di laurea (di cui 7 magistrali, 3 triennali e 1 tesi di master).



## UNIVERSITÀ DEGLI STUDI DI TORINO

### Aree e interessi di ricerca

*Natural Language Processing and Computational Linguistics* (word embeddings, sense embeddings, transformer networks, word sense disambiguation, word sense induction, supersense tagging, entity linking, semantic similarity), *Knowledge Representation and Reasoning* (formal ontology, conceptual categorization), *Information Extraction and Information Retrieval* (text mining, text categorization).

### Formazione, titoli

- Luglio 2022: **Dottorato di ricerca** in Informatica (XXXIV ciclo, Università degli Studi di Torino);
- 2017: **Laurea Magistrale** in Informatica, 110/110 *cum laude* (Università degli Studi di Torino);
- 2014: **Laurea** in Informatica: 110/110 (Università degli Studi di Torino);
- 2011: **Diploma** di perito informatico, 100/100 (Istituto Tecnico Piero Calamandrei, Santhià (TO)).

### Posizioni accademiche

- Giugno 2023: Ricercatore a tempo determinato di tipo A
- Febbraio 2023 - Maggio 2023: **Assegno di ricerca** sul progetto “Sviluppo di tecnologie per l’analisi semantica del linguaggio naturale, e per l’annotazione, l’estrazione, la rappresentazione e la visualizzazione di informazioni da documenti testuali”.
- Febbraio 2022 - Gennaio 2023: **Assegno di ricerca** sul progetto “Azioni per la prevenzione della violenza interpersonale: sviluppo di sistema per l’estrazione, analisi e integrazione di informazioni provenienti da campi narrativi di referti di Pronto Soccorso”.
- Novembre 2017 - Settembre 2018: **Borsa di studio** sul progetto PRiSMHA (Providing Rich Semantic Metadata for Historical Archives).



UNIVERSITÀ DEGLI STUDI DI TORINO

## ATTIVITÀ DI RICERCA E PUBBLICAZIONI SCIENTIFICHE

### Publicazioni scientifiche

Lista completa su IRIS APERTO: [https://iris.unito.it/simple-search?query=&location=&filter field=author&filter type=authority&filter value=rp68093&filter value display=COLLA%2C+DAVIDE&rpp=100&sort by=bi sort 2 sort &order=desc](https://iris.unito.it/simple-search?query=&location=&filter%20field=author&filter%20type=authority&filter%20value=rp68093&filter%20value%20display=COLLA%2C+DAVIDE&rpp=100&sort%20by=bi%20sort%202%20sort%20&order=desc)

### Riviste

1. Colla, D., Delsanto, M., & Radicioni, D. P. (2022). Semantic Coherence Dataset: Speech Transcripts. *Data in Brief*, vol.46, 108799. <https://doi.org/10.1016/j.dib.2022.108799>
2. Colla, D., Delsanto, M., Agosto, M., Vitiello, B., & Radicioni, D. P. (2022). Semantic coherence markers: The contribution of perplexity metrics. *Artificial Intelligence in Medicine*, 134, 102393. <https://doi.org/10.1016/j.artmed.2022.102393>; <https://www.sciencedirect.com/science/article/pii/S0933365722001440>
3. Colla, D., Goy, A., Leontino, M., Magro, D., & Picardi, C. (2021). Bringing Semantics into Historical Archives with Computer-aided Rich Metadata Generation. *ACM Journal on Computing and Cultural Heritage*.
4. Colla, D., Goy, A., Leontino, M., & Magro, D. (2021). Wikidata Support in the Creation of Rich Semantic Metadata for Historical Archives. *Applied Sciences*, 11(10), 4378.
5. Mensa, E., Colla, D., Dalmaso, M., Giustini, M., Mamo, C., Pitidis, A., & Radicioni, D. P. (2020). Violence detection explanation via semantic roles embeddings. *BMC medical informatics and decision making*, 20(1), 1-13.
6. Colla, D., Mensa, E., & Radicioni, D. P. (2020). Sense identification data: A dataset for lexical semantics. *Data in Brief*, 32, 106267.
7. Colla, D., Mensa, E., & Radicioni, D. P. (2020). Novel metrics for computing semantic similarity with sense embeddings. *Knowledge-Based Systems*, 206, 106346.
8. Goy, A., Colla, D., Magro, D., Accornero, C., Loreto, F., & Radicioni, D. P. (2020). Building semantic metadata for historical archives through an ontology-driven user interface. *Journal on Computing and Cultural Heritage (JOCCH)*, 13(3), 1-36.



UNIVERSITÀ DEGLI STUDI DI TORINO

9. Colla, D., Mensa, E., & Radicioni, D. P. (2020). LessLex: Linking multilingual embeddings to SenSe representations of LEXical items. *Computational Linguistics*, 46(2), 289-333.

**Conference proceedings**

1. Colla, D., Delsanto, M., & Di Nuovo, E. (2023). ELICODE at MultiGED2023: fine-tuning XLM-RoBERTa for multilingual grammatical error detection. In *LINKÖPING ELECTRONIC CONFERENCE PROCEEDINGS* (Vol. 197, pp. 24-34). Linköping University Electronic Press.
2. Mensa E., Liberatore D., Colla D., Delsanto M., Giustini M. & Radicioni D. P. (2022). Road Accidents: Information Extraction from Clinical Reports. *HC@AlxIA 2022: 1st AlxIA Workshop on Artificial Intelligence For Healthcare, November 28 – December 02, 2022, Udine*.
3. Demaria R., Delsanto M., Colla D., Mensa E., Pasini E., & Radicioni D. P. (2022). Shuffling-Based Data Augmentation for Argument Mining. *AI3@AlxIA 2022: 6th Workshop on Advances in Argumentation in Artificial Intelligence, November 28 – December 02, 2022, Udine*.
4. A. Goy, C. Re, D. Colla, M. Leontino. Turning 1968 memories into usable texts, *Proceedings of the First International Conference on Recent Advances in Digital Humanities (RADH 2021)*, University of Bucharest, Romania, 2021
5. Mensa, E., Marino, G. M., Colla, D., Delsanto, M., & Radicioni, D. P. (2021). A resource for detecting misspellings and denoising medical text data. In *CLiC-it 2020 Italian Conference on Computational Linguistics 2020* (Vol. 2769, pp. 1-7). CEUR
6. D. Colla, T. Caselli, V. Basile, J. Mitrovic, and M. Granitzer (2020, December). GruPaTo at SemEval-2020 Task 12: Retraining mBERT on social media and fine-tuned offensive language models . In *Proceedings of the Fourteenth Workshop on Semantic Evaluation* (pp. 1546-1554).
7. D. Colla, M. Leontino, E. Mensa, D. P. Radicioni (2019). From Sartre to Frege in three steps: A\* search for enriching semantic text similarity measures. In *Proceedings of the Sixth Italian Conference on Computational Linguistics* (pp. 131-138). CEUR.
8. A. Goy, C. Accornero, D. Astrologo, D. Colla, M. D'Ambrosio, R. Damiano, M. Leontino, A. Lieto, F. Loreto, D. Magro, E. Mensa, A. Montanaro, V. Mosca, S. Musso, D. P. Radicioni and C. Re (2019). Fruitful Synergies between Computer Science, Historical Studies and Archives: The Experience in the PRiSMHA Project. In *Proceedings of the 11th*



## UNIVERSITÀ DEGLI STUDI DI TORINO

*International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - Volume 3: KMIS*, Vienna. September 2019 (pp. 225-230).

9. D. Colla, E. Mensa, D. P. Radicioni and A. Lieto (2018). Conceptual abstractness: from nouns to verbs. In *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLIC-IT 2018)*, Torino. December 2018 (pp. 70-75). CEUR Workshop Proceedings.
10. Colla, D., Mensa, E., Radicioni, D. P., & Lieto, A. (2018). Tell me why: Computational explanation of conceptual similarity judgments. In *Proceedings of the 17th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, Special Session on Advances on Explainable Artificial Intelligence, Cadiz, Spain, June, 2018 (pp. 74-85). Springer, Cham.
11. A. Goy, R. Damiano, F. Loreto, D. Magro, S. Musso, D. P. Radicioni, C. Accornero, D. Colla, A. Lieto, E. Mensa, M. Rovera, D. Astrologo, B. Boniolo and M. D'Ambrosio (2017). PRISMHA (Providing Rich Semantic Metadata for Historical Archives): background, motivations, and challenges. In *Proceedings of the Joint Ontology Workshops 2017 Episode 3: The Tyrolean Autumn of Ontology*, Bolzano, Italy, September, 2017.
12. D. Colla, E. Mensa and D. P. Radicioni (2017, November). Semantic measures for keywords extraction. In *Proceedings of AI\*IA 2017 Advances in Artificial Intelligence: XVIth International Conference of the Italian Association for Artificial Intelligence*. Bari, Italy, November, 2017 (pp. 128-140). Springer, Cham.

### Partecipazione a progetti di ricerca

- Partecipazione a progetto 'L'accessibilità dei testi normativi come dispositivo di inclusione: studio di caso e strumenti applicativi in Valle d'Aosta. Seconda parte', progetto finanziato da Fondazione CRT in collaborazione con l'Università della Valle d'Aosta. Responsabile scientifico: Prof. Antonio Mastropaolo. Erogazione RF= 2022.1796. Durata: 03.2023-06.2025.
- Progetto "Turin Enhanced Philosophy Tree" (TEPT), progetto finanziato da Compagnia CRT in collaborazione con l'Università di Torino. Responsabile scientifico: Prof. Guido Bonino. **Ruolo del candidato: sviluppo di algoritmi per il matching dei filosofi presenti nell'albero**, sulla base di informazioni contenute in triple store enciclopedici (e.g., WikiData) e dell'analisi di eventuali frammenti di opere ivi reperibili. Inoltre sviluppo del database per la



## UNIVERSITÀ DEGLI STUDI DI TORINO

memorizzazione dei dati forniti dai partner di progetto e dell'interfaccia grafica per la visualizzazione dell'albero dei filosofi, oltre che di tutte le funzionalità di servizio per la gestione del database da interfaccia grafica. Durata: 01.10.2021–08.01.2024

- Progetto di Public Engagement “PAUSE (Prevention of Assault Under Scientific Evidence) interdisciplinari per contrastare la violenza interpersonale”. Progetto presentato all'interno del Bando Public Engagement 2021 di UNITO. Responsabile scientifico: Prof. Daniele Radicioni. **Ruolo del candidato: addestramento e utilizzo di modelli probabilistici del linguaggio per l'information extraction.** Durata: 01.10.2021–31.07.2023
- Progetto “L'accessibilità dei testi normativi come dispositivo di inclusione: studio di caso e strumenti applicativi in Valle d'Aosta”, progetto finanziato da Compagnia CRT in collaborazione con l'Università della Valle d'Aosta. Responsabile scientifico: Prof. Antonio Mastropaolo. **Ruolo del candidato: elaborazione di strumenti per l'estrazione di informazioni da documenti giuridici, ed elaborazione di modello di descrizione della complessità sintattica e semantica di diversi tipi di testo giuridico e normativo.** Durata: ottobre 2021-08.01.2024
- Progetto 'Analisi automatica delle cartelle di Pronto Soccorso per la prevenzione del femminicidio', finanziato da Fondazione CRT e Università degli Studi di Torino. Responsabile scientifico Prof. Daniele Radicioni. **Ruolo del candidato: contributo allo sviluppo di modello semantico e implementazione di sistema per l'estrazione di informazioni da campi narrativi dei referti cliniche.** Durata: febbraio 2020 – giugno 2022
- Progetto “RESEARCH STUDY IN THE DEVELOPMENT OF CHATBOTS FOR THE DOMAIN OF CAR SHARING”, all'interno del progetto 'Muoversi Meglio' (MuMe), finanziato nel quadro del POR FESR 2014/2020, PI Dott. Ing. Luigi Magnanini. **Ruolo del candidato: supporto nella progettazione dell'architettura di un Assistente Conversazionale con funzioni di interfaccia con gli utenti del servizio di car sharing.** Durata: da novembre 2018 a ottobre 2020
- Progetto di Ateneo e Compagnia di San Paolo 'PRiSMHA: Providing Rich Semantic Metadata for Historical Archives', responsabile scientifica la Prof. Annamaria Goy. **Ruolo del candidato: elaborazione di un modello per l'estrazione di eventi da documenti d'archivio, a partire da un modello ibrido basato su ontologie e spazi concettuali. Sviluppo di un sistema di**



## UNIVERSITÀ DEGLI STUDI DI TORINO

annotazione semantica per eventi estratti da documenti d'archivio.  
Elaborazione di un modello di allineamento fra ontologia sviluppata nel  
contesto del progetto e WikiData. Durata: 2017 – 2019.

### Attività in conferenze, comitati di programma, e di revisione

#### Reviewer di riviste e conferenze internazionali

- Reviewer for The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023) – 2023
- Reviewer for The Tenth Annual Conference on Advances in Cognitive Systems (ACS 2022)– 2022
- Reviewer for The 29th International Conference on Computational Linguistics (ACL 2022)– 2022
- Reviewer for 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2023) - 2022
- Reviewer for The The SIGNLL Conference on Computational Natural Language Learning (CoNLL 2022) - 2022
- Reviewer for The 37th ACM/SIGAPP Symposium On Applied Computing - 2022
- Reviewer for Natural Language Engineering –2021
- Reviewer for the Eighth Annual Conference on Advances in Cognitive Systems – 2020
- Reviewer for the Semantic Spaces at the Intersection of NLP, Physics, and Cognitive Science - 2020
- Reviewer for the 25th International Conference on Database Systems for Advanced Applications - 2020
- Reviewer for the 31st International Conference on Tools with Artificial Intelligence - 2019
- Reviewer for the 28th ACM International Conference on Information and Knowledge Management - 2019
- Reviewer for 24th International Conference on Database Systems for Advanced Applications - 2018
- Reviewer for the Knowledge and Language Processing Track @The 34th ACM Symposium on Applied Computing - ACM SAC 2019 - 2018



## UNIVERSITÀ DEGLI STUDI DI TORINO

### Comitati di programma in workshop e conferenze

- **Co-Organizer** of the Shared Task 'Assessing DIScourse COherence in Italian TEXTs (**DisCoTeX**)' at EVALITA 2023, the 8th Evaluation Campaign of NLP and Speech Tools for Italian
- **PC member** in Semantics, Knowledge Representation at Eighth Italian Conference on Computational Linguistics (CLIC-it) – 2021
- **PC member**, CONcreteness in conTEXT (CONcreTEXT) at Evaluation of NLP and Speech Tools for Italian (EVALITA) – 2020
- **PC member** in Semantics, Knowledge Representation at Seventh Italian Conference on Computational Linguistics (CLIC-it) – 2020

### Presentazioni a convegni

- Presentazione ELICODE at MultiGED2023: fine-tuning XLM-RoBERTa for multilingual grammatical error detection. NLP4CALL 2023: 12th Workshop on Natural Language Processing for Computer Assisted Language Learning. May 22, 2023, Tórshavn, Faroe Islands (<https://spraakbanken.gu.se/en/research/themes/icall/nlp4call-workshop-series/nlp4call2023#program>).
- Presentazione Third **DR2** Conference in Rome – 2023 (<https://www.di.unito.it/~colla/proofs/dr2.pdf>).
- Presentazione Road Accidents: Information Extraction from Clinical Reports. HC@AixIA 2022: 1st AixIA Workshop on Artificial Intelligence For Healthcare, November 28 – December 02, 2022, Udine (<https://sites.google.com/unical.it/hcaixia2022/program-and-accepted-papers>).
- Presentazione LessLex: Linking Multilingual Embeddings to SenSe Representations of Lexical Items (published in the Computational Linguistics Journal). 58th annual meeting of the Association for Computational Linguistics (ACL 2022), **ACL** 2020 in Seattle, WA, USA ([https://virtual.acl2020.org/paper\\_cl.1543.html](https://virtual.acl2020.org/paper_cl.1543.html)).
- Presentazione Grupato at semeval-2020 task 12: Retraining mbert on social media and fine-tuned offensive language models. Proceedings of the Fourteenth Workshop on Semantic Evaluation, 12 December 2020 (<https://semeval.github.io/semeval2020-sched.html>).
- Presentazione From Sartre to Frege in three steps: A\* search for enriching semantic text similarity measures. In *Proceedings of the Sixth Italian Conference on Computational Linguistics* (<http://clit2019.di.uniba.it/programma.html>).



## UNIVERSITÀ DEGLI STUDI DI TORINO

- Presentazione Evaluating the MuMe dialogue system with the IDIAL protocol. In *Proceedings of the Sixth Italian Conference on Computational Linguistics* (<http://clic2019.di.uniba.it/programma.html>).
- Presentazione Conceptual abstractness: from nouns to verbs. In *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLIC-IT 2018)*, Torino. December 2018 (<https://clic2018.di.unito.it/it/programma-dettaglio/>).

### Premi, riconoscimenti e invited talks

- **Invited talk** at Third **DR2 Conference** in Rome – 2023 (<https://www.di.unito.it/~colla/proofs/dr2.pdf>)
- **Invited talk** at 58th annual meeting of the Association for Computational Linguistics (**ACL 2022**), ACL 2020 in Seattle, WA, USA to present ‘LessLex: Linking Multilingual Embeddings to SenSe Representations of LEXical Items’ published in the Computational Linguistics Journal, ([https://virtual.acl2020.org/paper\\_cl.1543.html](https://virtual.acl2020.org/paper_cl.1543.html)) – 2020

### Collaborazione con enti esterni

- DC ha partecipato alle attività del gruppo nella **convenzione di ricerca** fra Dipartimento di Informatica dell’Università degli Studi di Torino e SSB Progetti Srl per la ‘Progettazione e implementazione di prototipo di agente conversazionale per prenotazione di autoveicoli elettrici’. Marzo 2018 – marzo 2020
- DC ha partecipato alle attività del gruppo nella convenzione di ricerca fra il Dipartimento di Informatica dell’Università di Torino e il Servizio sovrazonale di Epidemiologia dell’ASL TO3, finalizzata all’applicazione di tecniche di estrazione delle informazioni e di elaborazione automatica del linguaggio naturale per l’analisi delle cartelle cliniche e di documenti medici/clinici. Dicembre 2018 – dicembre 2021

### Altre attività: esperienze lavorative

- Website manager e sviluppatore software. Collaborazione con l’Archivio di Diritto e Storia Costituzionali, diretto dai Proff. Mario Dogliani e Mario Losano, per lo sviluppo di componenti software per l’elaborazione e marcatura semantica di documenti giuridici all’interno del Portale ‘Archivio di



## UNIVERSITÀ DEGLI STUDI DI TORINO

Diritto e Storia Costituzionali', <http://dircost.di.unito.it/index.shtml> (Dicembre 2016 — in corso).

- Web designer e Website manager. Collaborazione con la rivista Nuvole per la rielaborazione e la gestione del sito [www.nuvole.it](http://www.nuvole.it), rivista digitale. Il lavoro prevede, oltre alla normale gestione del sito, la creazione di strumenti che rendano gli articoli facilmente pubblicabili anche da più utenti contemporaneamente (Dicembre 2014 — in corso).



UNIVERSITÀ DEGLI STUDI DI TORINO

## ATTIVITÀ DIDATTICA

### Attività di co-relazione tesi

Davide Colla è stato co-relatore di varie tesi con stage interno su temi connessi alla propria attività di ricerca, fra cui 3 tesi di tirocinio triennale in Informatica e 7 tesi di laurea magistrale in informatica e 1 tesi di master.

Segue elenco delle tesi seguite con tirocinio interno, su temi connessi all'attività di ricerca.

- Tesi 'Studio della complessità testuale a partire dalla misura della perplexity dei language models', Laurea **Magistrale** in Informatica;
- Tesi 'Studio della complessità testuale a partire dall'analisi dei movimenti oculari', Laurea triennale in Informatica;
- Tesi 'Strumenti per la ricomposizione dei sensi in sense inventory frammentate', Laurea triennale in Informatica;
- Tesi 'Verso un approccio semantico profondo: utilizzo combinato di embedding contestuali e informazione sintattica nel text entailment', Laurea **Magistrale** in Informatica;
- Tesi 'Approcci dell'Elaborazione del Linguaggio Naturale per la Predizione della Psicosi', Laurea **Magistrale** in Informatica;
- tesi 'Neural Supersense Tagging', Laurea **Magistrale** in Informatica;
- tesi 'Esplorazione di Benchmark per l'NLP con Integrazione di Transformers e Applicazioni Alternative per il Textual Entailment', Laurea **Magistrale** in Informatica;
- tesi 'Esplorazione di Benchmark per l'NLP con Integrazione di Transformers e Applicazioni Alternative per il Textual Entailment', Laurea **Magistrale** in Informatica;
- tesi 'Data Science per la Filantropia Agenda 2030: studio pilota per un classificatore dei progetti presentati alla Fondazione CRC', Master di I livello in Analisi dei Dati per la Business Intelligence e Data Science;
- tesi 'Denoising Emergency Room Reports: detection and correction of abbreviations, acronyms and misspellings.', Laurea **Magistrale** in Informatica;
- tesi 'Segmentazione testuale con tecniche di analisi semantica', Laurea Triennale in Informatica.



## UNIVERSITÀ DEGLI STUDI DI TORINO

### Corsi

- **Titolare** del corso “Approfondimenti di Tecnologie Web”, MICA\_SAA0073, laurea in Management dell’informazione e della comunicazione aziendale dell’Università degli Studi di Torino – Dipartimento di Management – 2022 (<https://www.di.unito.it/~colla/proofs/mica.pdf>). 40 ore di didattica frontale. — 2022

### Didattica integrava

- didattica integrativa (bandi ex artt. 76 e 11) predisposta dai Corsi di Studio in Informatica, sia su tematiche di informatica di base (Linguaggi di Programmazione), sia su tematiche legate all'attività di ricerca (NLP). Anni accademici 2018-2019, 2020-2021, 2021-2022, 2022-2023. 160 ore (4 anni x 40 ore).

### Seminari

- **Invited** seminar on ‘LessLex and Sense Identification’ within the course “Techniques for Natural Language Understanding” (Master in AI, Turin) — 2022.
- **Invited** seminar on ‘LessLex and Sense Identification’ within the course “Techniques for Natural Language Understanding” (Master in AI, Turin) — 2021.
- **Invited** Seminar all’interno del Master Universitario M11033 in Analisi Dati per la Business Intelligence e Data Science, su introduzione del vector space model — 2019.
- **Invited** seminar on ‘Sperimentazioni sul calcolo della distanza semantica fra concetti’ within the course “Techniques for Natural Language Understanding” (Master in AI, Turin) — 2018.
- **Invited** Seminar all’interno del Master in Data Science (MADAS) presso il Collegio Carlo Alberto (Università degli Studi di Torino) sul tema ‘Semantic Resources for Text Documents Processing’ — 2018.



UNIVERSITÀ DEGLI STUDI DI TORINO

**Informativa ai sensi dell'art.13 del D.Lgs 196/2003:**

i dati sopra riportati sono prescritti dalle disposizioni vigenti ai fini del procedimento per il quale sono richiesti e verranno utilizzati esclusivamente per tale scopo.

Torino, 01/06/2023

FIRMA