CURRICULUM DELL'ATTIVITÀ SCIENTIFICA E DIDATTICA

DR. STEFANO MARCHESIN

GIUGNO 2025

Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Padova
E-mail: stefano.marchesin@unipd.it, marches1@dei.unipd.it

Indice

1 Dati di sintesi **1**

2 Percorso accademico **2**

2.1 Ricercatore universitario 2

2.2 Assegni di ricerca 2

2.3 Borse di studio 2

2.4 Titoli di studio 2

2.5 Partecipazione alle attività di gruppi di ricerca nazionali e internazionali 3

2.6 Attività formative e di ricerca a livello internazionale 3

2.7 Abilitazioni 3

3 Attività scientifica **4**

3.1 Tematiche di ricerca 4

3.2 Partecipazione a progetti di ricerca a livello europeo 7

3.3 Presentazioni a congressi internazionali 7

3.4 Presentazioni a congressi nazionali 8

3.5 Presentazioni su invito 9

3.6 Attività di peer-reviewing 10

3.7 Organizzazione di eventi scientifici 13

3.8 Esperienze professionali 14

3.9 Librerie software 14

3.10 Premi 15

4 Attività didattica **15**

4.1 Insegnamenti 15

4.2 Didattica integrativa e servizio agli studenti 17

4.3 Lezioni su invito 17

4.4 Supervisione dottorato di ricerca 17

4.5 Supervisione tesi di laurea 18

5 Attività istituzionali, organizzative e di servizio **19**

6 Terza missione **19**

7 Pubblicazioni **20**

Riviste internazionali 20
Curatele 22
Atti di conferenze internazionali 22
Atti di workshop internazionali 25
Atti di conferenze nazionali 25
Tesi di dottorato 27

1 Dati di sintesi

Nome:	Stefano
Cognome:	Marchesin
Data di nascita:	29 ottobre 1992
Stato civile:	Coniugato
Cittadinanza:	Italiana
Codice fiscale:	MRCSFN92R29F241W
Cellulare:	347 649 77 42
E-mail:	`stefano.marchesin@unipd.it`, `marches1@dei.unipd.it`
Posizione:	Ricercatore universitario (RTDa)
Affiliazione:	Dipartimento di Ingegneria dell'Informazione Università degli Studi di Padova Via Gradenigo, 6/b 35131 Padova
Pagina Web:	`http://www.dei.unipd.it/~marches1/`

2 Percorso accademico

2.1 Ricercatore universitario

- *Dicembre 2023 - in corso*: **Ricercatore a tempo determinato di tipo A (RTDa)** presso il Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova. Finanziato dal Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova. RTDa a regime di tempo pieno.

2.2 Assegni di ricerca

- *Ottobre 2023 - Dicembre 2023*: **Assegnista di ricerca** di tipo Senior B presso il Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova.
  Assegno di ricerca dal titolo *"Efficient Methods to Estimate Knowledge Graph Accuracy"*. Finanziato dal Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova.

- *Novembre 2021 - Ottobre 2023*: **Assegnista di ricerca** di tipo Junior B presso il Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova.
  Assegno di ricerca dal titolo *"Unsupervised Neural Models for First-stage Medical Information Retrieval"*. Finanziato dal Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova. Referente scientifico: Prof. G. Silvello.

- *Febbraio 2021 - Novembre 2021*: **Assegnista di ricerca** di tipo A presso il Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova.
  Assegno di ricerca dal titolo *"Modelli di Ranking Neurale non Supervisionati per il Recupero di Informazioni Mediche"*. Finanziato dal progetto H2020 ExaMode. Referente scientifico: Prof. G. Silvello.

- *Aprile 2017 - Settembre 2017*: **Assegnista di ricerca** di tipo A presso il Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova.
  Assegno di ricerca dal titolo *"Progettazione e implementazione di un servizio di cross-site personalisation"*. Finanziato dal progetto DCH. Referente scientifico: Prof. M. Agosti.

2.3 Borse di studio

- **Borsa di dottorato** (XXXIII ciclo) finanziata dall'Università degli Studi di Padova dall'1 ottobre 2017 al 30 settembre 2020.

2.4 Titoli di studio

- *Ottobre 2017 - Marzo 2021*: **Dottorato di ricerca in Ingegneria dell'Informazione, Scuola di dottorato in Ingegneria dell'Informazione**, indirizzo in "Scienza e Tecnologia dell'Informazione", Ciclo: XXXIII, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova.
  Titolo della tesi: "Developing Unsupervised Knowledge-Enhanced Models to Reduce the Semantic Gap in Information Retrieval" (redatta in lingua inglese). Supervisore: Prof. M. Agosti.

- *Ottobre 2014 - Dicembre 2016*: **Laurea magistrale in Ingegneria Informatica, Università degli Studi di Padova**.
  Titolo della tesi: "An Adaptive Cross-Site User Modelling Platform for Information Exchange Techniques" (redatta in lingua inglese). Relatore: Prof. M. Agosti. Correlatori: Prof. V. Wade e Prof. S. Lawless.

- *Ottobre 2011 - Settembre 2014*: **Laurea triennale in Ingegneria dell'Informazione, Università degli Studi di Padova**.
  Titolo della tesi: "Ideal Secret Sharing Schemes Based on Integer Polymatroids" (redatta in lingua inglese). Relatore: Prof. N. Laurenti.

2.5 Partecipazione alle attività di gruppi di ricerca nazionali e internazionali

- *Marzo 2016 - oggi*: **Member** del gruppo Information Management Systems (IMS) coordinato dal prof. Nicola Ferro del Dipartimento di Ingegneria dell'Informazione dell'Università degli studi di Padova. Tematiche di ricerca: Information extraction, retrieval e quality.

- *Aprile 2017 - luglio 2017, marzo 2021 - agosto 2021*: **Visiting researcher** presso il MedGIFT group coordinato dal Prof. Henning Müller del Institute of Informatics della University of Applied Sciences and Arts of Western Switzerland, Sierre, Switzerland. Tematiche di ricerca: Information extraction e multimodal learning.

- *Agosto 2016 - novembre 2016*: **Internship**, tramite il progetto Erasmus + for traineeships, presso l'ADAPT Centre coordinato dal Prof. Vincent Wade del Department of Computer Science del Trinity College Dublin, Dublin, Ireland. Tematiche di ricerca: User modeling, adaptation e personalization.

2.6 Attività formative e di ricerca a livello internazionale

- *15 Luglio 2019 - 19 Luglio 2019*: **ESSIR 2019 - The 12th European Summer School in Information Retrieval**, Milano, Italia.

- *4 Settembre 2017 - 8 Settembre 2017*: **ESSIR 2017 - The 11th European Summer School in Information Retrieval**, Barcellona, Spagna.

- *Febbraio 2015 - Giugno 2015*: **Erasmus + for studies at Polytechnic University of Catalonia, Barcelona, Spain**. Corsi seguiti: Machine Learning; Open Data; Operating Systems; Concurrence, Parallelism and Distributed Systems; E-Business.

2.7 Abilitazioni

- *15 Luglio 2024 - 15 Luglio 2036*: **Abilitazione scientifica nazionale al ruolo di professore di II fascia per il settore concorsuale 01/B1 – Informatica**.

3 Attività scientifica

3.1 Tematiche di ricerca

Information Retrieval

Semantic gap. La ricerca mira a colmare il semantic gap, un problema critico nel reperimento delle informazioni (IR), che si verifica quando query e documenti utilizzano termini diversi per esprimere lo stesso concetto, impedendo ai modelli tradizionali basati su segnali lessicali di abbinarli. Per affrontare questo problema, sono emerse due linee principali di lavoro: 1) l'uso di risorse di conoscenza esterne per migliorare le rappresentazioni di query e documenti e 2) l'uso di modelli semantici per una corrispondenza semantica tra le rappresentazioni latenti di query e documenti.

Attraverso una valutazione approfondita di modelli lessicali e semantici, abbiamo mostrato la necessità di combinare questi segnali nelle prime fasi della pipeline di IR al fine di affrontare efficacemente il semantic gap [12]. Sulla base di questa valutazione, abbiamo sviluppato modelli per il dominio medico – dove il semantic gap è prominente – mostrando che, quando rilevanti per la query, le relazioni semantiche sono più utili dei singoli concetti nel recupero della letteratura medica ed il Supporto Decisionale Clinico (CDS) [49, 67]. Tuttavia, è emerso che le espansioni di query possono avere effetti negativi se non pesate correttamente [41, 43]

Uno studio di validazione sui dati di TREC Precision Medicine (PM) 2019 ha confermato l'efficacia delle tecniche di riformulazione pesata delle query nel recupero di studi clinici, evidenziando però la necessità di tecniche più avanzate per combinare i segnali delle riformulazioni [40, 64]. Una visione completa e coerente degli studi sui modelli lessicali è stata presentata in un unico framework [11].

Per i modelli semantici, abbiamo proposto un framework che combina rappresentazioni implicite ed esplicite per ridurre il semantic gap [44]. In [13], il framework è stato perfezionato nel Semantic-Aware Neural Framework for IR (SAFIR), un framework neurale di self-supervised, multi-task learning ottimizzato su task di IR. SAFIR apprende congiuntamente rappresentazioni di parole, concetti e documenti. Tali rappresentazioni sono ottimizzate per task di IR e codificano sia polisemia che sinonimia al fine di affrontare il semantic gap tra query e documenti. Un'approfondita analisi sperimentale ha evidenziato che SAFIR recupera più documenti rilevanti rispetto ad approcci concorrenti, risultando inoltre adatto ad essere combinato con modelli lessicali tramite Pseudo Relevance Feedback (PRF). L'efficiacia nel combinare i segnali semantici di SAFIR con quelli lessicali per il semantic gap ha confermato quanto sostenuto in [12].

Query performance prediction. Data l'importanza del semantic gap in IR, abbiamo approfondito lo studio dei segnali lessicali e semantici anche in ambito predittivo. In particolare, dato che tali segnali sono complementari, risultano essere utili per diversi tipi di query in diversa misura. Pertanto, è cruciale identificare le query per cui il semantic gap è più rilevante e quindi dove i modelli semantici risultano più efficaci. A questo proposito, abbiamo condotto uno studio in [37] che ha quantificato l'impatto dell'uso di modelli semantici o lessicali su diverse query, proponendo una strategia per classificare le query in base alla loro difficoltà semantica.

Lo studio è proseguito poi focalizzandosi su approcci di Query Performance Prediction (QPP), pensati per stimare le prestazioni di un sistema di IR senza la necessità di avere giudizi di rilevanza manuali. Tradizionalmente, questi approcci si basano su segnali lessicali e sono stati applicati a modelli lessicali di IR con vari gradi di successo. Con l'avvento dei modelli neurali e dei Large Language Model (LLM), il paradigma di reperimento si è spostato verso segnali più semantici. Abbiamo quindi

analizzato in [34] l'efficacia degli approcci QPP nel prevedere le prestazioni dei modelli neurali di IR, mostrando che i QPP tradizionali sono meno efficaci con i modelli neurali, soprattutto in contesti dove i segnali semantici sono prominenti.

Motivati da ciò, in [31], abbiamo proposto un nuovo framework per sviluppare feature che possano essere usate dai QPP per i modelli neurali. Utilizzando il framework come una tavola periodica delle componenti QPP, i ricercatori possono sviluppare predittori più adatti a predire le prestazioni dei modelli neurali (i.e., semantici). Allo stesso tempo, il framework ci ha permesso di evidenziare le sfide ed opportunità che i QPP presentano per le diverse fasi della pipeline di IR neurale. Abbiamo mostrato il potenziale del framework attraverso lo sviluppo di due nuovi predittori che superano lo state dell'arte, dimonstrando la versatilità del framework nel descrivere diverse tipologie di QPP.

Information Extraction

La ricerca nell'ambito dell'Information Extraction (IE) si inserisce in un progetto a lungo termine volto a ridurre il semantic gap e migliorare l'efficacia dei metodi di support decisionale clinico, come descritto in [45]. L'obiettivo è sviluppare sistemi avanzati di CDS tramite metodi di Entity Linking (EL) e Relation Extraction (RE), mirati all'identificazione ed estrazione di concetti e relazioni da documenti testuali per costruire – o arricchire – risorse di conoscenza note come Knowledge Graph (KG).

Entity linking. Per quanto riguarda la ricerca su EL, abbiamo sviluppato il Semantic Knowledge Extractor Tool (SKET) [6, 8], un sistema ibrido e non supervisionato che si basa su regole e LLM per estrarre informazioni sul cancro da referti medici non strutturati. SKET è progettato per estrarre informazioni relative a cancro al colon, alla cervice, e ai polmoni. I concetti estratti possono essere usati per vari downstream task, come knowledge discovery, manual annotation support, e weakly supervised image classification. In particolare, in [9] SKET è stato usato per addestrare un modello neurale di image classification per il cancro al colon senza l'ausilio di annotazioni manuali – riducendo drasticamente i costi di sviluppo. Successivamente, abbiamo esteso SKET ad un nuovo caso d'uso: l'estrazione di informazioni rigaurdanti la Celiachia [56].

Relation extraction. Parallelamente alla ricerca su EL, lo studio sui metodi di RE si è concentrato sulle Gene Disease Association (GDA). In questo contesto, abbiamo identificato una mancanza di benchmark per sviluppare e valutare metodi di RE all'avanguardia per GDA. Così, in [10], abbiamo costruito il più grande dataset per lo sviluppo e la valutazione di metodi di estrazione di GDA, usato poi come benchmark per testare diversi modelli di RE stato dell'arte. Il benchmark ha evidenziato le limitazioni degli approcci correnti nell'estrazione delle GDA. Sia il dataset che i modelli sono stati resi disponibili per favorire lo sviluppo e la valutazione di nuovi modelli per l'estrazione di GDA.

Successivamente, abbiamo usato i modelli rilasciati con il benchmark come componente automatica di supporto per MetaTron [4], il primo strumento di annotazione web-based e open-source che consente di annotare relazioni semantiche tra dati biomedici in modo flessibile, interattivo, e collaborativo. MetaTron permette di annotare relazioni con il supporto di ontologie – funzionalità spesso trascurata dagli strumenti di annotazione disponibili in commercio. MetaTron è stato recentemente esteso in modo significativo e applicato all'annotazione di dati per task di IR, portando allo sviluppo di DocTron [20]. DocTron supporta l'evidenziazione di passaggi testuali, l'identificazione di relazioni, il tagging e linking di entità, la valutazione della rilevanza di un documento rispetto ad una query ed il riconoscimento di oggetti. Fornisce un ambiente collaborativo con gestione dei ruoli e include metriche di accordo tra annotatori per garantire coerenza e qualità dei dati.

Knowledge graph construction. Grazie alla ricerca su EL e RE, è stato possibile sviluppare il sistema Collaborative Oriented Relation Extraction (CORE) per la costruzione di KG con annotazioni manuali limitate [5]. CORE combina distant supervision e active learning, offrendo un'architettura modulare, trasparente e adatta all'elaborazione su larga scala. Eseguendo test iterativi sull'affidabilità dei dati estratti, CORE è in grado di offrire un servizio di versioning atto a migliorare continuamente – attraverso tecniche di active learning – i KG costruiti. Tramite CORE, abbiamo creato il più grande KG contenente associazioni "fine-grained" tra espressione genica e cancro.

Per facilitare l'accesso al KG generato da CORE, abbiamo sviluppato la piattaforma web COREKB [32], che permette la ricerca di associazioni gene-cancro tramite query di testo libero o strutturate. COREKB offre funzionalità come autocompletamento, schede riassuntive delle relazioni principali e panoramiche dettagliate sui contenuti – inclusa l'analisi dell'evidenza scientifica a supporto o contrasto del contenuto in esame.

Recentemente, si è progettato un metodo di multimodal learning in grado di minimizzare i costi di addestramento per la generazione di rappresentazioni efficaci per classificare e reperire immagini e documenti medici, e, per la prima volta, costruire automaticamente KG multimodali [1].

Information Quality

Le informazioni contenute nei KG vengono quotidianamente utilizzate in una moltitudine di servizi e applicazioni popolari. Tuttavia, i metodi odierni di costruzione dei KG non sono perfetti, generando così risorse con un alto grado di sparsità dei dati, numerose inaccurattezze, e fatti errati. Pertanto, verificare la qualità dei KG è essenziale per prendere decisioni informate nei servizi o nelle applicazioni che ne fanno uso. Tuttavia, valutare manualmente la qualità dei KG è estremamente costoso date le loro dimensioni. Infatti, i KG possono arrivare a contenere miliardi di fatti differenti.

Per risolvere questo problema, la ricerca si concentra sullo sviluppo di tecniche di campionamento efficienti per stimare la qualità dei KG. In questo contesto, abbiamo affrontato in [3] le limitazioni degli approcci correnti, che si basano sul metodo di Wald per costruire intervalli di confidenza e che presentano problemi di affidabilità come intervalli ad ampiezza nulla e sovrastime. La nostra soluzione, basata sul metodo di Wilson e adattata ad approcci di campionamento complesso, supera queste limitazioni e garantisce un'applicabilità estesa a vari scenari di valutazione. Successivamente, abbiamo proposto in [22] l'utilizzo di intervalli credibili, basati sul framework Bayesiano, per ovviare ai problemi interpretativi degli intervalli di confidenza, basati sul framework frequentista. Gli intervalli credibili garantiscono maggiore affidabilità e solidità in scenari *one-shot*, come la valutazione della qualità dei KG, dove il processo di annotazione viene tipicamente condotto una sola volta a causa dei costi elevati. In questo contesto, abbiamo introdotto un approccio adattivo, chiamato aHPD, che non richiede di scegliere una specifica *prior* a priori – una delle più note resistenze all'utilizzo di metodi Bayesiani in statistica – e che risulta sia efficiente che robusto.

Lo studio di tecniche efficienti per stimare la qualità dei KG ha portato anche allo sviluppo di applicazioni pratiche, come la valutazione su larga scala della qualità di *DBpedia* [26], e l'implementazione di un approccio stratificato per valutare la qualità, progettato per ottimizzare l'utilità dei KG in applicazioni di entity-oriented search [25].

Recentemente, si sono studiate metodologie automatiche basate su LLM per valutare la qualità dei KG, con l'obiettivo di ridurre ulteriormente i costi di verifica manuale. Questo ha portato allo sviluppo di FactCheck [21], un sistema di verifica che combina LLM e IR, integrato in una piattaforma web interattiva. La piattaforma consente una verifica trasparente dei fatti nei KG, confrontando le interpretazioni di diversi modelli, ed è pensata sia per utenti tecnici che non esperti.

3.2 Partecipazione a progetti di ricerca a livello europeo

HEREDITARY (Contratto n. 101137074, durata del progetto: 2024 - 2027) HEREDITARY (HetERogeneous sEmantic Data integratIon for the guT-bRain interplaY), è un progetto finanziato dalla Commissione Europea nel contesto del programma Horizon Europe. Il progetto si inserisce nella linea di ricerca "Tools and technologies for a healthy society" della Commissione Europea.

Tra gli obiettivi il progetto mira a trasformare il modo in cui si affronta l'individuazione delle malattie, si prepara la risposta terapeutica e si esplorano le conoscenze mediche costruendo un framework solido, interoperabile, affidabile e sicuro che integri dati sanitari multimodali garantendo al contempo il rispetto della privacy.

Ruolo. Responsabile del comitato "Communication and Dissemination" per l'Università degli Studi di Padova; referente per il task 4.6 "Evidence-based knowledge graph creation and exploration"; membro dell'unità di ricerca dell'Università degli Studi di Padova.

BRAINTEASER (Contratto n. 101017598, durata del progetto: 2021 - 2024) BRAINTEASER (BRinging Artificial INTelligencE home for a better cAre of amyotrophic lateral sclerosis and multiple SclERosis), è un progetto finanziato dalla Commissione Europea nel contesto del programma H2020. Il progetto si inserisce nella linea di ricerca "Societal Challenges - Health, demographic change and well-being" della Commissione Europea.

Tra gli obiettivi il progetto mira a sviluppare un sistema di sensori indossabili per consentire previsioni nonché per far progredire processi decisionali clinici e prevenzione. In questo modo, si potrà assistere i medici nel suggerire interventi che possano ritardare la progressione della malattia.

Ruolo. Membro dell'unità di ricerca dell'Università degli Studi di Padova.

EXAMODE (Contratto n. 825292, durata del progetto: 2019 - 2023) EXAMODE (EXtreme-scale Analytics via Multimodal Ontology Discovery & Enhancement), è un progetto finanziato dalla Commissione Europea nel contesto del programma H2020. Il progetto si inserisce nella linea di ricerca "Big Data technologies and extreme-scale analytics" della Commissione Europea.

Tra gli obiettivi il progetto mira a progettare e sviluppare metodi e strumenti di analisi su larga scala per dati medici e di sfruttare metodi "weakly supervised" di deep learning per l'estrazione di informazione da basi di conoscenza mediche.

Ruolo. Task leader per il task 2.1 "Semantic knowledge extractor prototype"; task leader per il task 2.3 "Automatic knowledge discovery system prototype and user study outcome"; Membro dell'unità di ricerca dell'Università degli Studi di Padova.

3.3 Presentazioni a congressi internazionali

1. *Credible Intervals for Knowledge Graph Accuracy Estimation*, 2025 ACM SIGMOD/PODS International Conference on Management of Data (SIGMOD/PODS 2025), Berlino, Germania, 22-27 giugno 2025.

2. *Veracity Estimation for Entity-Oriented Search with Knowledge Graphs*, 33rd ACM International Conference on Information and Knowledge Management (CIKM 2024), Boise, Idaho, USA, 21-25 ottobre 2024.

3. *Utility-Oriented Knowledge Graph Accuracy Estimation with Limited Annotations: A Case Study on DBpedia*, 12th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2024), Pittsburgh, Pennsylvania, USA, 16-19 ottobre 2024.

4. *Efficient and Reliable Estimation of Knowledge Graph Accuracy*, 50th International Conference on Very Large Databases (VLDB 2024), Guangzhou, Cina, 26-30 agosto 2024.

5. *Terminology Extraction in Electronic Health Records. The ExaMode Project*, 1st International Conference on Multilingual Digital Terminology Today (MDTT 2022), Padova, Italia, 16-17 giugno 2022.

6. *Filter, Transform, Expand, and Fuse. The IMS Unipd at TREC 2021 Clinical Trials*, 30th Text REtrieval Conference (TREC 2021), Gaithersburg, Maryland, USA, 15-19 novembre 2021.

7. *Exploring how to Combine Query Reformulations for Precision Medicine*, 28th Text REtrieval Conference (TREC 2019), Gaithersburg, Maryland, USA, 13-15 novembre 2019.

8. *An Analysis of Query Reformulation Techniques for Precision Medicine*, 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019), Parigi, Francia, 21-25 luglio 2019.

9. *Knowledge Enhanced Representations to Reduce the Semantic Gap in Clinical Decision Support*, 9th PhD Symposium on Future Directions in Information Access (FDIA 2019), Milano, Italia, 17 luglio 2019.

10. *A Relation Extraction Approach for Clinical Decision Support*, 12th International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO 2018) co-located with the 27th ACM International Conference on Information and Knowledge Management (CIKM 2018), Torino, Italia, 22 ottobre 2018.

11. *Implicit-Explicit Representations for Case-Based Retrieval*, 1st International Conference on Design of Experimental Search & Information REtrieval Systems (DESIRES 2018), Bertinoro, Italia, 28-31 agosto 2018.

12. *Case-Based Retrieval Using Document-Level Semantic Networks*, 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018), Ann Arbor, Michigan, USA, 8-12 luglio 2018.

13. *Towards open-source shared implementations of keyword-based access systems to relational data*, 1st International Workshop on Keyword-based Access and Ranking at Scale (KARS 2017) co-located with the EDBT/ICDT 2017 Joint Conference (EDBT/ICDT 2017), Venezia, Italia, 21 marzo 2017.

3.4 Presentazioni a congressi nazionali

1. *SKET: an Unsupervised Knowledge Extraction Tool to Empower Digital Pathology Applications*, 19th Conference on Information and Research science Connecting to Digital and Library science (IRCDL 2023), Bari, Italia, 23-24 febbraio 2023.

2. *Assessing the Semantic Difficulty of Queries*, 12th Italian Information Retrieval Workshop (IIR 2022), Milano, Italia, 29-30 giugno 2022.

3. *Exploiting Curated Databases to Train Relation Extraction Models for Gene-Disease Associations*, 30th Symposium on Advanced Database Systems (SEBD 2022), Tirrenia, Italia, 19-22 giugno 2022.

4. *SAFIR: a Semantic-Aware Neural Framework for IR*, 11th Italian Information Retrieval Workshop (IIR 2021), Bari, Italia, 13-15 settembre 2021.

5. *A Post-Analysis of Query Reformulation Methods for Clinical Trials Retrieval*, 28th Italian Symposium on Advanced Database Systems (SEBD 2020), Villasimius, Italia, 21-24 giugno 2020.

6. *Reproducibility of the Neural Vector Space Model via Docker*, 16th Italian Research Conference on Digital Libraries (IRCDL 2020), Bari, Italia, 30-31 gennaio 2020.

7. *Knowledge Enhanced Representations for Clinical Decision Support*, 10th Italian Information Retrieval Workshop (IIR 2019), Padova, Italia, 16-18 settembre 2019.

8. *Medical Retrieval using Structured Information Extracted from Knowledge Bases*, 27th Italian Symposium on Advanced Database Systems (SEBD 2019), Castiglione della Pescaia, Italia, 16-19 giugno 2019.

9. *Thirty Years of Digital Libraries Research at the University of Padua: The User Side*, 14th Italian Research Conference on Digital Libraries (IRCDL 2018), Udine, Italia, 25-26 gennaio 2018.

10. *An Adaptive Cross-Site User Modelling Platform for Cultural Heritage Websites*, 13th Italian Research Conference on Digital Libraries (IRCDL 2017), Modena, Italia, 26-27 gennaio 2017.

3.5 Presentazioni su invito

1. *Efficient Evaluation of Knowledge Graph Quality: Challenges and Opportunities*, nell'ambito della serie di seminari "Tales on Data Science and Big Data" organizzata dal CINI Lab on Data Science, Online, 9 maggio 2025. Disponibile al seguente link. (Invited talk)

2. *From Building Knowledge Bases with Nanopubs to Evaluating them: The Role of Data Quality*, Knowledge Pixels, Online, 25 giugno 2024. (Invited talk)

3. *Preserving Clinicians' Time: Empowering Healthcare AI with Minimal Expert Annotations*, Artificial Intelligence for Healthcare (AI4H), Padova, 21-22 marzo 2024. (Invited poster)

4. *Annotated Data Doesn't Grow on Trees, but Nanopublications Do: Exploiting Nanopublications to Train Relation Extraction Models for Gene-Disease Associations*, Knowledge Pixels, Online, 16 maggio 2023. (Invited talk)

5. *Learning Unsupervised Knowledge-Enhanced Representations to Reduce the Semantic Gap in Information Retrieval*, 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2021), Online, 11-15 luglio 2021. (Invited presentation for ACM TOIS paper [13])

3.6 Attività di peer-reviewing

Comitati editoriali

- Membro dell'Editorial Review Board della rivista ACM Journal of Data and Information Quality dal 14 marzo 2025. (SCImago 2024: Q2 Information Systems)
- Guest Editor della rivista Information Processing & Management per la special issue "Large Language Models and Data Quality for Knowledge Graphs". Special issue disponibile all'indirizzo: `https://www.sciencedirect.com/special-issue/10C05PM18JB`. (SCImago 2024: Q1 Information Systems)
- Membro del comitato editoriale e Review Editor per la sezione Machine Learning and Artificial Intelligence delle riviste Frontiers in Big Data e Frontiers in Artificial Intelligence dal 24 luglio 2023. (SCImago 2024: Q2 Information Systems; Q2 Artificial Intelligence)
- Membro del comitato editoriale e Review Editor per la sezione Emerging Technologies and Transformative Paradigms in Research della rivista Frontiers in Research Metrics and Analytics dal 14 aprile 2023. (SCImago 2024: Q1 Library and Information Sciences)

Riviste

- ACM Computing Surveys (CSUR)
- ACM Transactions on Information Systems (TOIS)
- Applied Intelligence
- Artificial Intelligence in Medicine
- BMC Medical Informatics and Decision Making
- Data Science and Engineering
- Expert Systems With Applications (ESWA)
- Frontiers in Artificial Intelligence
- IEEE Access
- Intelligent Systems With Applications (ISWA)
- International Journal of Digital Libraries (IJDL)
- International Journal of Information Technology and Decision Making (IJITDM)
- Journal of Medical Internet Research (JMIR)
- Journal of Supercomputing (J SUPERCOMPUT)
- Nature Scientific Data
- Nature Scientific Reports
- PeerJ Computer Science

Membro di comitati di programma di congressi internazionali

- **The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2025):**
  - Full papers track;
  - Short papers track;
  - Doctoral Consortium track. (Revisore e Mentore)
- **The 47th European Conference on Information Retrieval (ECIR 2025):**
  - Full papers track;
  - Short papers track;
  - Demo papers track.
- **The 31st International Conference on Computational Linguistics (COLING 2025):** Regular papers track.
- **The 18th ACM International Conference on Web Search and Data Mining (WSDM 2025):** Regular papers track.
- **The 18th ACM Conference on Recommender Systems (RecSys 2024):**
  - Late-breaking results papers track;
  - Reproducibility papers track.
- **The 33rd ACM International Conference on Information and Knowledge Management (CIKM 2024):**
  - Full papers track;
  - Short papers track.
- **The 28th International Conference on Theory and Practice of Digital Libraries (TPDL 2024):**
  - Full papers track;
  - Short papers track.
- **The 15th Conference and Labs of the Evaluation Forum (CLEF 2024):** Intelligent Disease Progression Prediction at CLEF (iDPP@CLEF) papers track.
- **The 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2024):**
  - Full papers track;
  - Resource and reproducibility papers track.
- **The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024):** Regular papers track.
- **The 46th European Conference on Information Retrieval (ECIR 2024):**
  - Full papers track;

  - Short papers track;
  - Demonstration papers track.

- **The 17th ACM International Conference on Web Search and Data Mining (WSDM 2024):** Regular papers track.
- **The 32nd ACM International Conference on Information and Knowledge Management (CIKM 2023):**
  - Full papers track;
  - Short papers track.
- **The 27th International Conference on Theory and Practice of Digital Libraries (TPDL 2023):** Full papers track.
- **The 14th biennial conference on Recent Advances in Natural Language Processing (RANLP 2023):** Regular papers track.
- **The 14th Conference and Labs of the Evaluation Forum (CLEF 2023):** Intelligent Disease Progression Prediction at CLEF (iDPP@CLEF) papers track.
- **The 18th Conference on Computational Intelligence Methods for Bioinformatics & Biostatistics (CIBB 2023):** Regular papers track.
- **Query Performance Prediction and Its Evaluation in New Tasks (QPP++ 2023):** Regular papers track.
- **The 45th European Conference on Information Retrieval (ECIR 2023):**
  - Full papers track;
  - Short papers track;
  - Reproducibility papers track;
  - Workshop proposals track.
- **The 16th ACM International Conference on Web Search and Data Mining (WSDM 2023):** Regular papers track.
- **The 13th Conference and Labs of the Evaluation Forum (CLEF 2022):** Lab overviews.
- **The 31st ACM International Conference on Information and Knowledge Management (CIKM 2022):** Full papers track.
- **The 26th International Conference on Theory and Practice of Digital Libraries (TPDL 2022):**
  - Research papers track;
  - Practitioner full papers track.
- **The 13th Language Resources and Evaluation Conference (LREC 2022):**
  - Evaluation and validation methodologies track;
  - Information extraction and information retrieval track.
- **The 44th European Conference on Information Retrieval (ECIR 2022):**

  - Full papers track;
  - Short papers track;
  - Reproducibility papers track.

- **The 13th biennial conference on Recent Advances in Natural Language Processing (RANLP 2021):** Regular papers track.
- **The 30th ACM International Conference on Information and Knowledge Management (CIKM 2021):**
  - Full papers track;
  - Applied papers track.
- **The 43rd European Conference on Information Retrieval (ECIR 2021):** Reproducibility papers track.

3.7 Organizzazione di eventi scientifici

Attività scientifiche

- **Program co-chair**. The 15th Italian Information Retrieval Workshop (IIR 2025), Cagliari, Italia, 3-5 settembre 2025.
- **Doctoral Consortium co-chair**. The 31st Symposium on Advanced Database System (SEBD 2023), Galzignano Terme (Padova), Italia, 2-5 luglio 2023.
- **Program co-chair**. The 19th Conference on Information and Research science Connecting to Digital and Library science (IRCDL 2019), Bari, Italia, 23-24 febbraio 2023.

Attività organizzative

- **Publicity co-chair**. The 21st Conference on Information and Research science Connecting to Digital and Library science (IRCDL 2025), Udine, Italia, 20-21 febbraio 2025.
- **Proceedings chair**. The 27th International Conference on Theory and Practice of Digital Libraries (TPDL 2023), Zadar, Croazia, 26-29 settembre 2023.
- **Proceedings chair**. The 2nd Design of Experimental Search & Information REtrieval Systems Conference (DESIRES 2021), Padova, Italia, 15-18 settembre 2021.
- **Publicity chair**. The 11th Italian Information Retrieval Workshop (IIR 2021), Bari, Italia, 13-15 settembre 2021.
- **Organization chair**. The 17th Italian Research Conference on Digital Libraries (IRCDL 2021), Padova, Italia, 18-19 febbraio 2021.

Campagne di valutazione

Le campagne di valutazione sono iniziative per valutare e confrontare in modo sistematico le prestazioni di diversi sistemi in ambiti come l'information retrieval, il natural language processing e la computer vision. Ogni campagna di valutazione richiede la definizione di task e metriche rigorose, nonché la creazione e condivisione di dataset al fine di garantire equità e riproducibilità nella misurazione dei risultati. I partecipanti sviluppano soluzioni che vengono testate in condizioni standardizzate, consentendo un confronto strutturato tra approcci diversi ed offrendo così un benchmark di riferimento per l'avanzamento scientifico. Le campagne di valutazione durano un anno e culminano in un workshop dedicato, dove i risultati vengono analizzati e discussi.

- **Co-organizer** di the 13th BioASQ Challenge on Large-Scale Biomedical Semantic Indexing and Question-Answering (BioASQ 2025), organizzata nel contesto di the 16th Conference and Labs of the Evaluation Forum (CLEF 2025). La campagna di valutazione si concluderà con un workshop internazionale dedicato, co-locato con CLEF 2025, 9-12 settembre 2025, Madrid, Spagna.
- **Co-organizer** di the 3rd Intelligent Disease Progression Prediction Challenge (iDPP 2024), organizzata nel contesto di the 15th Conference and Labs of the Evaluation Forum (CLEF 2024). La campagna di valutazione si è conclusa l'11-12 settembre 2024 con un workshop internazionale dedicato, co-locato con CLEF 2024, 9-12 settembre 2024, Grenoble, Francia.
- **Co-organizer** di the 2nd Intelligent Disease Progression Prediction Challenge (iDPP 2023), organizzata nel contesto di the 14th Conference and Labs of the Evaluation Forum (CLEF 2023). La campagna di valutazione si è conclusa il 21 settembre 2023 con un workshop internazionale dedicato, co-locato con CLEF 2023, 18-21 settembre 2023, Salonicco, Grecia.

3.8 Esperienze professionali

- *19 Maggio 2021 - 18 Giugno 2021*: **Prestazione di lavoro autonomo** per l'attività dal titolo "Unsupervised entity linking for cervical cancer reports". Finanziata dal progetto DCH. Referente scientifico: Prof. G. Silvello.
- *21 Aprile 2021 - 20 Maggio 2021*: **Prestazione di lavoro autonomo** per l'attività dal titolo "Scrittura di testi scientifici di carattere divulgativo su tematiche legate al dominio medico-informatico oggetto del progetto Examode". Finanziata dal progetto ExaMode. Referente scientifico: Prof. G. Silvello.
- *19 Ottobre 2020 - 18 Novembre 2020*: **Prestazione di lavoro autonomo** per l'attività dal titolo "Semantic Knowledge extraction and linking from medical reports". Finanziata dal progetto DCH. Referente scientifico: Prof. G. Silvello.

3.9 Librerie software

Si sono sviluppate librerie software open source per molti dei lavori pubblicati; si riportano di seguito le librerie per cui si è svolto il ruolo di collaboratore principale. Elenco in ordine di pubblicazione decrescente.

- https://github.com/KGAccuracyEval/credible-intervals4kg-estimation/ – software per implementare metodi di stima della qualità dei KG basati su intervalli credibili [22].
- https://github.com/KGAccuracyEval/kg-accuracy4entity-search/ – software per implementare metodi di stima della qualità dei KG usati nella ricerca orientata alle entità [25].
- https://github.com/KGAccuracyEval/dbpedia-accuracy-estimation/ – software per implementare metodi collaborativi di annotazione e stima della qualità di DBpedia [26].
- https://github.com/KGAccuracyEval/reliable-kg-estimation/ – software per implementare metodi di stima efficiente ed affidabile della qualità dei KG [3].
- https://github.com/GDAMining/core/ – software per implementare CORE [5], il sistema di creazione automatica di KG.
- https://github.com/ExaNLP/sket/ – software per implmentare SKET [6, 8], il sistema di estrazione automatica delle informazioni da referti medici non strutturati.
- https://github.com/GDAMining/gda-extraction/ – software per implementare i metodi di RE per GDA [10].
- https://github.com/stefano-marchesin/SAFIR/ – software per implementare SAFIR [13], il framework neurale di self-supervised, multi-task learning ottimizzato su task di IR e IE.

3.10 Premi

- **Best paper award** alla 21st Conference on Information and Research Science Connecting to Digital and Library Science (IRCDL 2025) per il paper "Extending Nanopublications with Knowledge Provenance for Multi-Source Scientific Assertions".
- **Best paper honorable mention award** alla 12th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2024) per il paper "Utility-Oriented Knowledge Graph Accuracy Estimation with Limited Annotations: A Case Study on DBpedia".
- **SIGIR student travel grant** finanziato da ACM SIGIR per partecipare alla 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018).
- **ESSIR 2017 scholarship** finanziata da ACM SIGIR per partecipare alla 11th European Summer School in Information Retrieval (ESSIR 2017).
- **Premio di studio 2016 per laureati magistrali in Ingegneria Informatica** assegnato dalla Fondazione Luciano Iglesias.

4 Attività didattica

4.1 Insegnamenti

Anno accademico 2024/2025

- "Methods for Scalable Graph Analytics", **2 CFU** per il corso di Dottorato in Ingegneria dell'Informazione dell'Università degli Studi di Padova. Co-titolarità del corso.

- “Foundations of Databases”, **6 CFU** per i corsi di Laurea Magistrale in ICT for Internet and Multimedia e Cybersecurity dell’Università degli Studi di Padova. Titolarità del corso.
  - Indicatore soddisfazione (media): 8.83; Indicatore azione didattica (media): 8.91; Indicatore organizzazione (media): 9.08.
- “Basi di Dati 1”, **3 CFU** per il corso di Laurea Triennale in Ingegneria Informatica dell’Università degli Studi di Padova. Titolarità del corso.
  - Indicatore soddisfazione (media): 7.73; Indicatore azione didattica (media): 7.75; Indicatore organizzazione (media): 8.07.

Anno accademico 2023/2024

- “Foundations of Databases”, **6 CFU** per i corsi di Laurea Magistrale in ICT for Internet and Multimedia, Computer Engineering e Cybersecurity dell’Università degli Studi di Padova. Titolarità del corso.
  - Indicatore soddisfazione (media): 8.82; Indicatore azione didattica (media): 8.83; Indicatore organizzazione (media): 8.97.

Anno accademico 2022/2023

- “Foundations of Databases”, **6 CFU** per i corsi di Laurea Magistrale in ICT for Internet and Multimedia, Computer Engineering e Cybersecurity dell’Università degli Studi di Padova. Titolarità del corso.
  - Indicatore soddisfazione (media): 8.65; Indicatore azione didattica (media): 8.95; Indicatore organizzazione (media): 8.94.

Anno accademico 2021/2022

- “Foundations of Databases”, **3 CFU** per i corsi di Laurea Magistrale in ICT for Internet and Multimedia, Computer Engineering e Cybersecurity dell’Università degli Studi di Padova.
  - Indicatore soddisfazione (media): 8.08; Indicatore azione didattica (media): 8.22; Indicatore organizzazione (media): 8.77; Indicatore organizzazione online (media): 8.76.
- “Basi di Dati 1”, **3 CFU** per il corso di Laurea Triennale in Ingegneria Informatica dell’Università degli Studi di Padova.
  - Indicatore soddisfazione (media): 6.90; Indicatore azione didattica (media): 6.63; Indicatore organizzazione (media): 7.25; Indicatore organizzazione online (media): 7.20.

Anno accademico 2020/2021

- “Foundations of Databases”, **3 CFU** per i corsi di Laurea Magistrale in ICT for Internet and Multimedia e Computer Engineering dell’Università degli Studi di Padova.
  - Indicatore soddisfazione (media): 8.32; Indicatore azione didattica (media): 8.02; Indicatore organizzazione online (media): 9.00.

4.2 Didattica integrativa e servizio agli studenti

Anno accademico 2019/2020

- "Database Management Systems", 22 ore di **didattica integrativa** per i corsi di Laurea Magistrale in Computer Engineering e ICT for Internet and Multimedia dell'Università degli Studi di Padova.

Anno accademico 2018/2019

- "Web Applications", 20 ore di **attività di tutoring** per i corsi di Laurea Magistrale in Computer Engineering e ICT for Internet and Multimedia dell'Università degli Studi di Padova.
- "Database Management Systems", 20 ore di **didattica integrativa** per i corsi di Laurea Magistrale in Computer Engineering e ICT for Internet and Multimedia dell'Università degli Studi di Padova.

4.3 Lezioni su invito

- *Efficient Knowledge Graph Quality Evaluation – Applications for Entity-Oriented Search*, lezione per il corso di Search Engines della Magistrale di Computer Engineering dell'Università degli Studi di Padova, tenuta in presenza, 6 marzo 2025.
- *Efficient Methods to Estimate Knowledge Graph Accuracy*, serie di tre lezioni per il corso di Database 2 della Magistrale di Computer Engineering dell'Università degli Studi di Padova, tenute in presenza, 13-15 dicembre 2023.
- *Introduction to Neural Networks*, lezione per il corso di Introduction to Machine Learning della Triennale in Ingegneria dell'Informazione dell'Università degli Studi di Padova, tenuta in presenza, 11 aprile 2022, 4 aprile 2023, e 27 maggio 2024.
- *Word Embeddings: Application Experiments*, lezione per il corso di Deep Learning Applied to Neuroscience and Rehabilitation della Magistrale di Bioingegneria dell'Università degli Studi di Padova, tenuta in presenza, 19 novembre 2021 e 2 dicembre 2022.
- *Query Reformulation Techniques for Precision Medicine*, lezione per il corso di Information Retrieval del Master in Computer Science dell'Università di Copenhagen, tenuta online, 13 maggio 2020.

4.4 Supervisione dottorato di ricerca

- **Marco Martinelli**, 40°ciclo del dottorato in Ingegneria dell'Informazione del Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova. Tematica: "Cross-Document Relation Extraction". Data di inizio: novembre 2024. Ruolo: co-supervisore (supervisore: Prof. G. Silvello).
- **Laura Menotti**, 39°ciclo del dottorato in Ingegneria dell'Informazione del Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova. Tematica: "Long-Tail Relation Extraction". Data di inizio: ottobre 2023. Ruolo: co-supervisore (supervisore: Prof. G. Silvello).

4.5 Supervisione tesi di laurea

Laurea Magistrale

- **Laura Settimo** (mat. 2121705), ICT for Internet and Multimedia, Università degli Studi di Padova. Tirocinio presso InfoCamere S.C.p.A. Tematica: "Analysis and Experimentation of Change Data Capture Technologies". Prevista per marzo 2026. Ruolo: relatore.
- **Irfan Ullah Khan** (mat. 2054601), Computer Engineering, Università degli Studi di Padova. Tematica: "Knowledge Graph Embeddings for Link Prediction". Prevista per dicembre 2025. Ruolo: relatore.
- **Saverio Fincato** (mat. 2087926), Computer Engineering, Università degli Studi di Padova. Tematica: "Fast and Accurate Query Processing on Dirty Data". Prevista per settembre 2025. Ruolo: relatore.
- **Matteo Baldoni** (mat. 2125716), Computer Engineering, Università degli Studi di Padova. Tirocinio presso EssilorLuxottica S.A. Tematica: "Integrating Text-to-SQL Techniques into Chatbots for Enhanced Question Answering". Prevista per settembre 2025. Ruolo: relatore.
- **Roberto Vicentini** (mat. 2055532), Data Science, Università degli Studi di Padova. Tirocinio presso Spindox S.p.A. Tesi: "Integrating Knowledge Graphs into RAG-Based LLMs to Improve Fact-Checking". Aprile 2025. Ruolo: relatore.
- **Farzad Shami** (mat. 2090160), Computer Engineering, Università degli Studi di Padova. Tesi: "FactCheck: Knowledge Graph Fact Verification Through Retrieval-Augmented Generation Using a Multi-Model Ensemble Approach". Marzo 2025. Ruolo: relatore (correlatore: Prof. G. Silvello).
- **Andrea Segala** (mat. 2082154), Computer Engineering, Università degli Studi di Padova. Tesi: "Large Language Models as Knowledge Graph Accuracy Estimators: An Exploratory Analysis". Dicembre 2024. Ruolo: relatore (correlatore: Prof. G. Silvello).
- **Asma Bakhtiariazad** (mat. 2050788), ICT for Internet and Multimedia, Università degli Studi di Padova. Tirocinio presso InfoCamere S.C.p.A. Tesi: "PostgreSQL in Kubernetes environment, implementation and testing of HA solutions in clouds". Ottobre 2024. Ruolo: relatore.
- **Borwoei Huang** (mat. 2044019), Computer Engineering, Università degli Studi di Padova. Tesi: "A Survey of Biomedical Relation Extraction Systems: Techniques, Domains, and Benchmarks". Luglio 2024. Ruolo: relatore.
- **Laura Menotti** (mat. 2019172), Computer Engineering, Università degli Studi di Padova. Tesi: "Reproducibility and Generalization of a Relation Extraction System for Gene-Disease Associations". Ottobre 2022. Ruolo: correlatore (relatore: Prof. G. Silvello).

Laurea Triennale

- **Pierlorenzo Ambroso** (mat. 2066689), Ingegneria Informatica, Università degli Studi di Padova. Tirocinio: Samo S.p.A. Tematica: "Design and Implementation of a Frontier Database for SAP-WMS Integration". Prevista per marzo 2026. Ruolo: relatore (correlatore: Dr. O. Irrera).
- **Samuele Volpato** (mat. 2074966), Ingegneria Informatica, Università degli Studi di Padova. Tematica: "Entity Resolution for Big Data". Prevista per settembre 2025. Ruolo: relatore.

- **Enoc Ke** (mat. 2066605), Ingegneria Informatica, Università degli Studi di Padova. Tematica: "An Analysis of Link Prediction Techniques". Prevista per settembre 2025. Ruolo: relatore (correlatore: Dr. O. Irrera).
- **Enrico De Carli** (mat. 2032595), Ingegneria Informatica, Università degli Studi di Padova. Tematica: "Using Knowledge Graphs to Enhance Fact Checking". Prevista per settembre 2025. Ruolo: relatore (correlatore: Dr. O. Irrera).
- **Alessandro Salvagno** (mat. 1195657), Ingegneria Informatica, Università degli Studi di Padova. Tesi: "Design and Development of a Web Application for the Management of Student Self-Assessments". Settembre 2022. Ruolo: correlatore (relatore: Prof. N. Ferro)
- **Francesca Cipolla** (mat. 594031), Ingegneria Informatica, Università degli Studi di Padova. Tesi: "Design and implementation of a database for text simplification: The SimpleText project". Settembre 2022. Ruolo: correlatore (relatore: Prof. G.M. Di Nunzio)

5 Attività istituzionali, organizzative e di servizio

- **Membro della commissione di verifica candidature international students** per il corso di Laurea Triennale in Information Engineering dell'Università degli Studi di Padova. Anno accademico 2025/2026.
- **Responsabile flusso Erasmus per la Faculty of Information Technology del Politecnico di Praga** presso il Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova. Inizio mobilità (previsto): anno accademico 2025/2026.
- **Membro della commissione VQR 2020-2024** del Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova. Ruolo: selezione prodotti per il gruppo informatici (INFO-01/A e IINF-05/A).
- **Membro del collegio dei ricercatori** per il Centro Studi in Terminologia Computazionale (CENTRICO) del Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova.
- **Coordinatore tutor junior** per il Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova. Anno accademico 2018/2019.
- **Rappresentante dei dottorandi** nel Consiglio di Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova. Novembre 2017 - Novembre 2019.

6 Terza missione

- **Seminario nell'ambito del progetto "SGUARDI: Nuove tecnocrazie?"** ospitato dal Dipartimento di Studi Linguistici e Letterari dell'Università degli Studi di Padova, tenuto l'8 aprile 2025. Titolo: "Intelligenza Artificiale: Tra Generazione e Creazione". Durata: 1h45m.
- **Workshop nell'ambito del Global Citizenship Programme 2024/2025** dell'Università degli Studi di Padova, tenuto il 7 dicembre 2024. Titolo: "Generative AI: Navigating the Future of Content Creation in Society". Durata: 2h30m.

7 Pubblicazioni

Sommario delle pubblicazioni:

- Riviste internazionali peer-reviewed: 13
- Conferenze internazionali peer-reviewed: 26
- Workshop internazionali peer-reviewed: 5
- Riviste internazionali: 2
- Conferenze nazionali peer-reviewed: 20
- Curatele: 4

Riviste internazionali (peer-reviewed)

[1] N. Marini, S. Marchesin, M. Wodzinski, A. Caputo, D. Podareanu, B. Cardenas Guevara, S. Boytcheva, S. Vatrano, F. Fraggetta, F. Ciompi, G. Silvello, H. Müller, and M. Atzori. Multimodal Representations of Biomedical Knowledge from Limited Training Whole Slide Images and Reports Using Deep Learning. *Medical Image Analysis*, Volume 97, Issue 1, October 2024, 103303, 2024. DOI: 10.1016/j.media.2024.103303 (SCImago 2024: Q1 Computer Vision and Pattern Recognition; Health Informatics)

[2] G. Faggioli, L. Menotti, S. Marchesin, A. Chiò, A. Dagliati, M. de Carvalho, M. Gromicho, U. Manera, E. Tavazzi, G. M. Di Nunzio, G. Silvello, and N. Ferro. An Extensible and Unifying Approach to Retrospective Clinical Data Modeling: the BrainTeaser Ontology. *Journal of Biomedical Semantics*, Volume 15, Issue 1, August 2024, 16, 2024. DOI: 10.1186/s13326-024-00317-y (SCImago 2024: Q2 Information Systems)

[3] S. Marchesin and G. Silvello. Efficient and Reliable Estimation of Knowledge Graph Accuracy. *Proceedings of the VLDB Endowment*, Volume 17, Issue 9, June 2024, pages 2392-2404, 2024. DOI: 10.14778/3665844.3665865 (SCImago 2024: Q1 Computer Science)

[4] O. Irrera, S. Marchesin, and G. Silvello. MetaTron: Advancing Biomedical Annotation Empowering Relation Annotation and Collaboration. *BMC Bioinformatics*, Volume 25, Issue 112, March 2024. DOI: 10.1186/s12859-024-05730-9 (SCImago 2024: Q1 Computer Science Applications)

[5] S. Marchesin, L. Menotti, F. Giachelle, G. Silvello, and O. Alonso. Building a Large Gene Expression-Cancer Knowledge Base with Limited Human Annotations. *Database: The Journal of Biological Databases and Curation*, Volume 2023, Issue 2023, September 2023, baad061, 2023. DOI: 10.1093/database/baad061 (SCImago 2024: Q1 Information Systems)

[6] L. Menotti, G. Silvello, M. Atzori, S. Boytcheva, F. Ciompi, G.M. Di Nunzio, F. Fraggetta, F. Giachelle, O. Irrera, S. Marchesin, N. Marini, H. Müller, and T. Primov. Modelling digital health data: The ExaMode ontology for computational pathology. *Journal of Pathology Informatics*, Volume 14, Issue 1, August 2023, 100332, 2023. DOI: 10.1016/j.jpi.2023.100332 (SCImago 2024: Q2 Computer Science Applications; Q1 Pathology and Forensic Medicine)

[7] G.M. Di Nunzio, S. Marchesin, and G. Silvello. A systematic review of Automatic Term Extraction: What happened in 2022? *Digital Scholarship in the Humanities*, Volume 38, Issue Supplement_1, June 2023, pages i41-i47, 2023. DOI: 10.1093/llc/fqad030 (SCImago 2024: Q3 Information Systems; Q1 Linguistics and Language)

[8] S. Marchesin, F. Giachelle, N. Marini, M. Atzori, S. Boytcheva, G. Buttafuoco, F. Ciompi, G.M. Di Nunzio, F. Fraggetta, O. Irrera, H. Müller, T. Primov, S. Vatrano, and G. Silvello. Empowering Digital Pathology Applications through Explainable Knowledge Extraction Tools. *Journal of Pathology Informatics*, Volume 13, Issue 1, September 2022, 100139, 2022. DOI: 10.1016/j.jpi.2022.100139 (SCImago 2024: Q2 Computer Science Applications; Q1 Pathology and Forensic Medicine)

[9] N. Marini, S. Marchesin, S. Otálora, M. Wodzinski, A. Caputo, M. van Rijthoven, W. Aswolinskiy, J. M. Bokhorst, D. Podareanu, E. Petters, S. Boytcheva, G. Buttafuoco, S. Vatrano, F. Fraggetta, J. der Laak, M. Agosti, F. Ciompi, G. Silvello, H. Müller, and M. Atzori. Unleashing the potential of digital pathology data by training computer-aided diagnosis models without human annotations. *npj Digital Medicine*, Volume 5, Issue 1, July 2022, 102, 2022. DOI: 10.1038/s41746-022-00635-4 (SCImago 2024: Q1 Computer Science Applications; Health Informatics)

[10] S. Marchesin and G. Silvello. TBGA: A Large-Scale Gene-Disease Association Dataset for Biomedical Relation Extraction. *BMC Bioinformatics*, Volume 23, Issue 1, March 2022, 111, 2022. DOI: 10.1186/s12859-022-04646-6 (SCImago 2024: Q1 Computer Science Applications)

[11] S. Marchesin, G.M. Di Nunzio, and M. Agosti. Simple but Effective Knowledge-Based Query Reformulations for Precision Medicine Retrieval. *MDPI Information*, Volume 12, Issue 10, Article 402, September 2021, 28 pages, 2021. DOI: 10.3390/info12100402. (SCImago 2024: Q2 Information Systems)

[12] S. Marchesin, A. Purpura and G. Silvello. Focal Elements of Neural Information Retrieval Models. An Outlook through a Reproducibility Study. *Information Processing & Management (IP&M)*, Volume 57, Issue 6, November 2020, 102109, 2020. ISSN: 2330-1643. DOI: 10.1016/j.ipm.2019.102109 (SCImago 2024: Q1 Information Systems)

[13] M. Agosti, S. Marchesin, and G. Silvello. Learning Unsupervised Knowledge-Enhanced Representations to Reduce the Semantic Gap in Information Retrieval. *ACM Transactions on Information Systems (TOIS)*, Volume 38, Issue 4, Article 38, October 2020, 48 pages, 2020. DOI: 10.1145/3417996 (SCImago 2024: Q1 Information Systems)

Riviste internazionali

[14] O. Alonso, S. Marchesin, M. Najork, and G. Silvello. Report on the 2nd International Conference on Design of Experimental Search & Information REtrieval Systems (DESIRES 2021). *SIGIR Forum*, Volume 55, Issue 2, Article 14, December 2021, 13 pages, 2021. DOI: 10.1145/3527546.3527563

[15] S. Marchesin. Developing Unsupervised Knowledge-Enhanced Models to Reduce the Semantic Gap in Information Retrieval. *SIGIR Forum*, Volume 55, Issue 1, Article 18, June 2021, 2 pages, 2021. DOI: 10.1145/3476415.3476433

Curatele

[16] O. Alonso, H. Cousijn, G. Silvello, M. Marrero, C. Teixeira Lopes, and S. Marchesin (2023). TPDL 2023: Linking Theory and Practice of Digital Libraries. Proceedings of the 27th International Conference on Theory and Practice of Digital Libraries. Lecture Notes in Computer Science 14241, Springer. Zadar, Croatia, September 26-29, 2023.

[17] D. Calvanese, C. Diamantini, N. Ferro, S. Marchesin, G. Silvello, and L. Tanca (2023). SEBD 2023: 31st Italian Symposium on Advanced Database Systems. Proceedings of the 31st Italian Symposium on Advanced Database Systems. CEUR Workshop Proceedings 3478. Galzignano Terme (Padova), Italy, July 2-5, 2023.

[18] A. Bardi, A. Falcon, S. Ferilli, S. Marchesin, and D. Redavid (2023). IRCDL 2023: Information and Research science Connecting to Digital and Library science. Proceedings of the Conference on Information and Research science Connecting to Digital and Library science. CEUR Workshop Proceedings 3365. Bari, Italy, February 23-24, 2023.

[19] O. Alonso, S. Marchesin, M. Najork, and G. Silvello (2021). DESIRES 2021: Design of Experimental Search & Information Retrieval Systems. Proceedings of the Second International Conference on Design of Experimental Search & Information REtrieval Systems. CEUR Workshop Proceedings 2950. Padova, Italy, September 15-18, 2021.

Atti di conferenze internazionali (peer-reviewed)

[20] O. Irrera, S. Marchesin, F. Shami, and G. Silvello. Doctron: A web-based collaborative annotation tool for Ground Truth creation in IR. To appear in *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2025)*, Padua, Italy, July 13-17, 2025, pp. 10. (CORE 2023: A*)

[21] F. Shami, S. Marchesin, and G. Silvello. Fact Verification in Knowledge Graphs Using LLMs. To appear in *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2025)*, Padua, Italy, July 13-17, 2025, pp. 5. (CORE 2023: A*)

[22] S. Marchesin and G. Silvello. Credible Intervals for Knowledge Graph Accuracy Estimation. *Proceedings of the ACM on Management of Data*, Volume 3, Issue 3 (SIGMOD), Article 142 (June 2025), Berlin, Germany, June 22-27, 2025, pp. 26. (CORE 2023: A*)

[23] A. Nentidis, G. Katsimpras, A. Krithara, M. Krallinger, M. Rodriguez Ortega, N. Loukachevitch, A. Sakhovskiy, E. Tutubalina, G. Tsoumakas, G. Giannakoulas, A. Bekiaridou, A. Samaras, G. M. Di Nunzio, N. Ferro, S. Marchesin, L. Menotti, G. Silvello, and G. Paliouras. BioASQ at CLEF2025: The thirteenth edition of the large-scale biomedical semantic indexing and question answering challenge. *Proceedings of the 47th European Conference on Information Retrieval (ECIR 2025)*, Lucca, Italy, April 6-10, 2025, pp. 407-415. (CORE 2023: A)

[24] O. Irrera, S. Marchesin, and G. Silvello. MetaTron: Streamlining Collaborative Annotation for Biomedical Documents. *Proceedings of the 16th International SWAT4HCLS Conference (SWAT4HCLS 2025)*, Barcelona, Spain, February 24-27, 2025, pp. 2.

[25] S. Marchesin, G. Silvello, and O. Alonso. Veracity Estimation for Entity-Oriented Search with Knowledge Graphs. *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM 2024)*, Boise, Idaho, USA, October 21-25, 2024, pp. 1649-1659. (CORE 2023: A)

[26] S. Marchesin, G. Silvello, and O. Alonso. Utility-Oriented Knowledge Graph Accuracy Estimation with Limited Annotations: A Case Study on DBpedia. *Proceedings of the 12th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2024)*, Pittsburgh, Pennsylvania, USA, October 16-19, 2024, pp. 105-114. **Best paper honorable mention.** (CORE 2023: B)

[27] G. Birolo, P. Bosoni, G. Faggioli, H. Aidos, R. Bergamaschi, P. Cavalla, A. Chiò, A. Dagliati, M. de Carvalho, G. M. Di Nunzio, P. Fariselli, J. M. García Dominguez, M. Gromicho, A. Guazzo, E. Longato, S. Madeira, U. Manera, S. Marchesin, L. Menotti, G. Silvello, E. Tavazzi, E. Tavazzi, I. Trescato, M. Vettoretti, B. Di Camillo and N. Ferro. Overview of iDPP@CLEF 2024: The Intelligent Disease Progression Prediction Challenge. *25th Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024)*, Grenoble, France, September 9-12, 2024, pp. 1312-1331.

[28] G. Birolo, P. Bosoni, G. Faggioli, H. Aidos, R. Bergamaschi, P. Cavalla, A. Chiò, A. Dagliati, M. de Carvalho, G. M. Di Nunzio, P. Fariselli, J. M. García Dominguez, M. Gromicho, A. Guazzo, E. Longato, S. Madeira, U. Manera, S. Marchesin, L. Menotti, G. Silvello, E. Tavazzi, E. Tavazzi, I. Trescato, M. Vettoretti, B. Di Camillo and N. Ferro. Intelligent Disease Progression Prediction: Overview of iDPP@CLEF 2024. *Proceedings of the 15th International Conference of the Cross-Language Evaluation Forum for European Languages (CLEF 2024)*, Grenoble, France, September 9-12, 2024, pp. 118-139.

[29] G. Faggioli, A. Guazzo, S. Marchesin, L. Menotti, I. Trescato, H. Aidos, R. Bergamaschi, G. Birolo, P. Cavalla, A. Chiò, A. Dagliati, M. de Carvalho, G. M. Di Nunzio, P. Fariselli, J. M. García Dominguez, M. Gromicho, E. Longato, S. Madeira, U. Manera, G. Silvello, E. Tavazzi, E. Tavazzi, M. Vettoretti, B. Di Camillo and N. Ferro. Overview of iDPP@CLEF 2023: The Intelligent Disease Progression Prediction Challenge. *24th Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023)*, Thessaloniki, Greece, September 18-21, 2023, pp. 1123-1164.

[30] G. Faggioli, A. Guazzo, S. Marchesin, L. Menotti, I. Trescato, H. Aidos, R. Bergamaschi, G. Birolo, P. Cavalla, A. Chiò, A. Dagliati, M. de Carvalho, G. M. Di Nunzio, P. Fariselli, J. M. García Dominguez, M. Gromicho, E. Longato, S. Madeira, U. Manera, G. Silvello, E. Tavazzi, E. Tavazzi, M. Vettoretti, B. Di Camillo and N. Ferro. Intelligent Disease Progression Prediction: Overview of iDPP@CLEF 2023. *Proceedings of the 14th International Conference of the Cross-Language Evaluation Forum for European Languages (CLEF 2023)*, Thessaloniki, Greece, September 18-21, 2023, pp. 343-369.

[31] G. Faggioli, T. Formal, S. Lupart, S. Marchesin, S. Clinchant, N. Ferro, and B. Piwowarski. Towards Query Performance Prediction for Neural Information Retrieval: Challenges and Opportunities. *Proceedings of the 13th International Conference on the Theory of Information Retrieval (ICTIR 2023)*, Taipei, Taiwan, July 23, 2023, pp. 51-63.

[32] F. Giachelle, S. Marchesin, G. Silvello, and O. Alonso. Searching for Reliable Facts over a Medical Knowledge Base. *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2023)*, Taipei, Taiwan, July 23-27, 2023, pp. 3205-3209. (CORE 2023: A*)

[33] B. Cardenas Guevara, N. Marini, S. Marchesin, W. Aswolinskiy, R. J. Schlimbach, D. Podareanu, and F. Ciompi. Caption generation from histopathology whole-slide images using pre-trained transformers. Accepted at *Medical Imaging with Deep Learning (MIDL 2023)*, Nashville, USA, July 10-12, 2023, pp. 5.

[34] G. Faggioli, T. Formal, S. Marchesin, S. Clinchant, N. Ferro, and B. Piwowarski. Query Performance Prediction for Neural IR: Are We There Yet? *Proceedings of the 45th European Conference on Information Retrieval (ECIR 2023)*, Dublin, Ireland, April 2-6, 2023, pp. 232-248. (CORE 2023: A)

[35] S. Marchesin, G.M. Di Nunzio, and G. Silvello. Terminology Extraction in Electronic Health Records. The ExaMode Project. *Proceedings of the 1st International Conference on Multilingual Digital Terminology Today (MDTT 2022)*, Padua, Italy, June 16-17, 2022, pp. 3.

[36] G.M. Di Nunzio, G. Faggioli, and S. Marchesin. Filter, Transform, Expand, and Fuse. The IMS Unipd at TREC 2021 Clinical Trials. *Proceedings of the 30th Text REtrieval Conference (TREC 2021)*, Gaithersburg, Maryland, USA, November 15-19, 2021, pp. 7.

[37] G. Faggioli and S. Marchesin. What Makes a Query Semantically Hard? *Proceedings of the Second International Conference on Design of Experimental Search & Information REtrieval Systems (DESIRES 2021)*, Padova, Italy, September 15-18, 2021, pp. 61-69.

[38] G.M. Di Nunzio and S. Marchesin. A Study on Query Expansion and Rank Fusion for Precision Medicine. The IMS Unipd at TREC 2020 Precision Medicine. *Proceedings of the 29th Text REtrieval Conference (TREC 2020)*, Gaithersburg, Maryland, USA, November 18-20, 2020, pp. 8.

[39] G.M. Di Nunzio, S. Marchesin, and F. Vezzani. A Study on Reciprocal Ranking Fusion in Consumer Health Search. IMS UniPD at CLEF eHealth 2020 Task 2. *Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum (CLEF 2020)*, Thessaloniki, Greece, September 22-25, 2020, pp. 7.

[40] G.M. Di Nunzio, S. Marchesin, and M. Agosti. Exploring how to Combine Query Reformulations for Precision Medicine. *Proceedings of the 28th Text REtrieval Conference (TREC 2019)*, Gaithersburg, Maryland, USA, November 13-15, 2019, pp. 14.

[41] M. Agosti, G.M. Di Nunzio, and S. Marchesin. An Analysis of Query Reformulation Techniques for Precision Medicine. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019)*, Paris, France, July 21-25, 2019, pp. 973-976. (CORE 2023: A*)

[42] G.M. Di Nunzio, and S. Marchesin. The University of Padua IMS Research Group at CENTRE@ TREC 2018. *Proceedings of the 27th Text REtrieval Conference (TREC 2018)*, Gaithersburg, Maryland, USA, November 14-16, 2018, pp. 7.

[43] M. Agosti, G.M. Di Nunzio, and S. Marchesin. The University of Padua IMS Research Group at TREC 2018 Precision Medicine Track. *Proceedings of the 27th Text REtrieval Conference (TREC 2018)*, Gaithersburg, Maryland, USA, November 14-16, 2018, pp. 10.

[44] S. Marchesin. Implicit-Explicit Representations for Case-Based Retrieval. *Proceedings of the Design of Experimental Search & Information REtrieval Systems (DESIRES 2018)*, Bertinoro, Italy, August 28-31, 2018, p. 109.

[45] S. Marchesin. Case-Based Retrieval Using Document-Level Semantic Networks. *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018)*, Ann Arbor, Michigan (USA), July 8-12, 2018, p. 1451. (CORE 2023: A*)

Atti di workshop internazionali (peer-reviewed)

[46] N. Marini, S. Otálora, F. Ciompi, G. Silvello, S. Marchesin, S. Vatrano, G. Buttafuoco, M. Atzori, and H. Müller. Multi-Scale Task Multiple Instance Learning for the Classification of Digital Pathology Images with Global Annotations. *Proceedings of the MICCAI Computational Pathology (COMPAY) Workshop*, Strasbourg, France, September 27, 2021, pp. 170-181.

[47] N. Ferro, S. Marchesin, A. Purpura, and G. Silvello. A Docker-Based Replicability Study of a Neural Information Retrieval Model. *Proceedings of the Open-Source IR Replicability Challenge co-located with 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (OSIRRC@SIGIR 2019)*, Paris, France, July 25, 2019, pp. 37-43.

[48] S. Marchesin. Knowledge Enhanced Representations to Reduce the Semantic Gap in Clinical Decision Support. *Proceedings of the 9th PhD Symposium on Future Directions in Information Access (FDIA 2019)*, Milan, Italy, July 17, 2019, pp. 4-9.

[49] M. Agosti, G.M. Di Nunzio, S. Marchesin, and G. Silvello. A Relation Extraction Approach for Clinical Decision Support. *Proceedings of the CIKM 2018 Workshops co-located with 27th ACM International Conference on Information and Knowledge Management (CIKM 2018)*, Torino, Italy, October 22, 2018, pp. 6.

[50] A. Badan, L. Benvegnù, M. Biasetton, G. Bonato, A. Brighente, A. Cenzato, P. Ceron, G. Cogato, S. Marchesin, A. Minetto, L. Pellegrina, A. Purpura, R. Simionato, N. Soleti, M. Tessarotto, A. Tonon, F. Vendramin, and N. Ferro. Towards Open-Source Shared Implementations of Keyword-Based Access Systems to Relational Data. *Proceedings of the Workshops of the EDBT/ICDT 2017 Joint Conference (EDBT/ICDT 2017)*, Venice, Italy, March 21-24, 2017, pp. 5.

Atti di conferenze nazionali (peer-reviewed)

[51] S. Marchesin and G. Silvello. Binomial Confidence Intervals for Knowledge Graph Accuracy Estimation (Extended Abstract). To appear in *Proceedings of the 33rd Italian Symposium on Advanced Database Systems (SEBD 2025)*, Ischia, Italy, June 16-19, 2025, pp. 10.

[52] F. Giachelle, S. Marchesin, L. Menotti, and G. Silvello. Extending Nanopublications with Knowledge Provenance for Multi-Source Scientific Assertions. *Proceedings of the 21st Conference on Information and Research Science Connecting to Digital and Library Science (IRCDL 2025)*, Udine, Italy, February 20-21, 2025, pp. 15. **Best paper.**

[53] S. Marchesin, L. Menotti, F. Giachelle, G. Silvello, and O. Alonso. Bootstrapping Gene Expression-Cancer Knowledge Bases with Limited Human Annotations. *Proceedings of the 32nd Italian Symposium on Advanced Database Systems (SEBD 2024)*, Villasimius, Italy, June 23-26, 2024, pp. 163-173.

[54] F. Giachelle, S. Marchesin, L. Menotti, and G. Silvello. Publishing CoreKB Facts as Nanopublications. *Proceedings of the 20th Conference on Information and Research Science Connecting to Digital and Library Science (IRCDL 2024)*, Bressanone, Italy, February 22-23, 2024, pp. 16-24.

[55] F. Giachelle, S. Marchesin, G. Silvello, and O. Alonso. CoreKB: A Web-based Platform for Searching Reliable Facts over a Medical Knowledge Base. *Proceedings of the 31st Italian Symposium on Advanced Database Systems (SEBD 2023)*, Galzignano Terme, Padova, Italy, July 2-5, 2023, pp. 200-209.

[56] L. Menotti, S. Marchesin, and G. Silvello. An Ontology-Driven Knowledge Extraction Tool for Pathology Record Classification. *Proceedings of the 31st Italian Symposium on Advanced Database Systems (SEBD 2023)*, Galzignano Terme, Padova, Italy, July 2-5, 2023, pp. 228-238.

[57] G. Faggioli, T. Formal, S. Marchesin, S. Clinchant, N. Ferro and B. Piwowarski. On the Limitations of Query Performance Prediction for Neural IR (Discussion paper). *Proceedings of the 31st Italian Symposium on Advanced Database Systems (SEBD 2023)*, Galzignano Terme, Padova, Italy, July 2-5, 2023, pp. 379-390.

[58] G.M. Di Nunzio, G. Faggioli, and S. Marchesin. An Analysis of a Methodology and Experimental Results for the Retrieval of Clinical Trials. *Proceedings of the 13th Italian Information Retrieval Workshop (IIR 2023)*, Pisa, Italy, June 8-9, 2023, pp. 53-57.

[59] F. Giachelle, S. Marchesin, and G. Silvello. SKET X: A Visual Analytics Tool for Explaining Knowledge Extraction Results. *Proceedings of the 13th Italian Information Retrieval Workshop (IIR 2023)*, Pisa, Italy, June 8-9, 2023, pp. 119-124.

[60] G.M. Di Nunzio, N. Ferro, F. Giachelle, O. Irrera, S. Marchesin, and G. Silvello. SKET: an Unsupervised Knowledge Extraction Tool to Empower Digital Pathology Applications. *Proceedings of the 19th Conference on Information and Research science Connecting to Digital and Library science (IRCDL 2023)*, Bari, Italy, February 23-24, 2023, pp. 144-152.

[61] G. Faggioli and S. Marchesin. Assessing the Semantic Difficulty of Queries. *Proceedings of the 12th Italian Information Retrieval Workshop (IIR 2022)*, Milan, Italy, June 29-30, 2022, pp. 6.

[62] S. Marchesin and G. Silvello. Exploiting Curated Databases to Train Relation Extraction Models for Gene-Disease Associations. *Proceedings of the 30th Symposium on Advanced Database Systems (SEBD 2022)*, Tirrenia (Pisa), Italy, June 19-22, 2022, pp. 133-140.

[63] M. Agosti, S. Marchesin, and G. Silvello. SAFIR: a Semantic-Aware Neural Framework for IR. *Proceedings of the 11th Italian Information Retrieval Workshop (IIR 2021)*, Bari, Italy, September 13-15, 2021, pp. 4.

[64] M. Agosti, G.M. Di Nunzio, and S. Marchesin. A Post-Analysis of Query Reformulation Methods for Clinical Trials Retrieval. *Proceedings of the 28th Italian Symposium on Advanced Database Systems (SEBD 2020)*, Villasimius, Sardinia, Italy, June 21-24, 2020, pp. 152-159.

[65] N. Ferro, S. Marchesin, A. Purpura, and G. Silvello. Reproducibility of the Neural Vector Space Model via Docker. *Proceedings of the 16th Italian Research Conference on Digital Libraries (IRCDL 2020)*, Bari, Italy, January 30-31, 2020, pp. 3-8.

[66] S. Marchesin and M. Agosti. Knowledge Enhanced Representations for Clinical Decision Support. *Proceedings of the 10th Italian Information Retrieval Workshop (IIR 2019)*, Padua, Italy, September 16-18, 2019, pp. 17-18.

[67] M. Agosti, G.M. Di Nunzio, S. Marchesin, and G. Silvello. Medical Retrieval using Structured Information Extracted from Knowledge Bases. *Proceedings of the 27th Italian Symposium on Advanced Database Systems (SEBD 2019)*, Castiglione della Pescaia, Grosseto, Italy, June 16-19, 2019, pp. 8.

[68] M. Agosti, G.M. Di Nunzio, N. Ferro, M. Maistro, S. Marchesin, N. Orio, C. Ponchia, and G. Silvello. Thirty Years of Digital Libraries Research at the University of Padua: The User Side. *Proceedings of the 14th Italian Research Conference on Digital Libraries (IRCDL 2018)*, Udine, Italy, January 25-26, 2018, pp. 42-54.

[69] A. Badan, L. Benvegnù, M. Biasetton, G. Bonato, A. Brighente, S. Marchesin, A. Minetto, L. Pellegrina, A. Purpura, R. Simionato, M. Tessarotto, A. Tonon, and N. Ferro. Keyword-based access to relational data: To reproduce, or to not reproduce? *Proceedings of the 25th Italian Symposium on Advanced Database Systems (SEBD 2017)*, Squillace Lido (Catanzaro), Italy, June 25-29, 2017, p. 166.

[70] M. Agosti, S. Lawless, S. Marchesin, and V. Wade. An Adaptive Cross-Site User Modelling Platform for Cultural Heritage Websites. *Proceedings of the 13th Italian Research Conference on Digital Libraries (IRCDL 2017)*, Modena, Italy, January 26-27, 2017, pp. 132-141.

Tesi di dottorato

[71] S. Marchesin. Developing Unsupervised Knowledge-Enhanced Models to Reduce the Semantic Gap in Information Retrieval. *PhD Thesis*, Doctoral School in Information Engineering, Department of Information Engineering, University of Padua, Italy, March 24, 2021, pp. 240.

Ai sensi degli articoli 19, 46, 47 e 76 D.P.R. 445/2000, dichiaro che le informazioni contenute nel presente curriculum sono vere e corrette.

Stefano Marchesin giugno 2025