

---

**CURRICULUM DELL'ATTIVITÀ SCIENTIFICA E DIDATTICA**

**DR. STEFANO MARCHESIN**

SETTEMBRE 2021

---

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



## Indice

<b>1</b>	<b>Dati di sintesi</b>	<b>1</b>
<b>2</b>	<b>Percorso accademico</b>	<b>2</b>
2.1	Assegni di ricerca . . . . .	2
2.2	Borse di studio . . . . .	2
2.3	Titoli di studio . . . . .	2
2.4	Partecipazione alle attività di gruppi di ricerca nazionali o internazionali . . . . .	2
2.5	Attività formative e di ricerca a livello internazionale . . . . .	3
<b>3</b>	<b>Attività scientifica</b>	<b>3</b>
3.1	Tematiche di ricerca . . . . .	3
3.2	Partecipazione a progetti di ricerca a livello europeo . . . . .	5
3.3	Presentazioni a congressi internazionali . . . . .	6
3.4	Presentazioni a congressi nazionali . . . . .	6
3.5	Attività di peer-reviewing . . . . .	7
3.6	Partecipazione all'organizzazione di eventi scientifici . . . . .	7
3.7	Esperienze professionali . . . . .	8
3.8	Premi . . . . .	8
<b>4</b>	<b>Attività didattica</b>	<b>8</b>
4.1	Didattica titolarità . . . . .	9
4.2	Didattica integrativa e servizio agli studenti . . . . .	9
4.3	Seminari . . . . .	9
<b>5</b>	<b>Attività istituzionali, organizzative e di servizio</b>	<b>9</b>
<b>6</b>	<b>Pubblicazioni</b>	<b>10</b>
	Riviste Internazionali . . . . .	10
	Atti di conferenze internazionali . . . . .	10
	Atti di workshop . . . . .	11
	Atti di conferenze nazionali . . . . .	11
	Tesi di dottorato . . . . .	12

## 1 Dati di sintesi

Nome: Stefano  
Cognome: Marchesin  
Data di nascita: 29 ottobre 1992  
Stato civile: Celibe  
Cittadinanza: Italiana  
Codice fiscale: MRCSFN92R29F241W  
Cellulare: 347 649 77 42  
E-mail: stefano.marchesin@unipd.it, marches1@dei.unipd.it

Posizione: Assegnista di ricerca  
Affiliazione: Dipartimento di Ingegneria dell'Informazione  
Università degli Studi di Padova  
Via Gradenigo, 6/a  
35131 Padova

Pagina Web: <http://www.dei.unipd.it/~marches1/>

## 2 Percorso accademico

### 2.1 Assegni di ricerca

- *Febbraio 2021 - in corso*: **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.

### 2.2 Borse di studio

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

### 2.3 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.

### 2.4 Partecipazione alle attività di gruppi di ricerca nazionali o 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 retrieval, entity linking e relation extraction.
- *Aprile 2017 - luglio 2017, marzo 2021 - agosto 2021*: **Visiting researcher** presso il MedGIFT group coordinato dal Prof. Henning Müller della University of Applied Sciences Western Switzerland, Sierre, Switzerland. Tematiche di ricerca: Biomedical 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.

## 2.5 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; Concurrency, Parallelism and Distributed Systems; E-Business.

## 3 Attività scientifica

### 3.1 Tematiche di ricerca

L'obiettivo principale della mia ricerca è affrontare il gap semantico, un problema fondamentale per il reperimento delle informazioni (IR). Il gap semantico può essere descritto come la mancata corrispondenza tra le query degli utenti e il modo in cui i modelli di reperimento rispondono a tali query. Tradizionalmente, i modelli di IR si basano su segnali di corrispondenza lessicale per eseguire il recupero. Avvalendosi delle rappresentazioni "bag-of-words" di query e documenti, i modelli lessicali calcolano quindi uno score di rilevanza utilizzando euristiche definite sulla sovrapposizione lessicale tra query e documenti. Pertanto, quando una query e un documento utilizzano parole diverse per esprimere lo stesso concetto, i modelli lessicali non riescono a combinarli. Per colmare tale lacuna, negli ultimi anni sono emerse due linee principali di lavoro: (i) l'uso di risorse di conoscenza esterne per migliorare le rappresentazioni bag-of-words di query e documenti e (ii) l'uso di modelli semantici per eseguire una corrispondenza semantica tra le rappresentazioni latenti di query e documenti.

Per affrontare questo problema, si è per prima cosa eseguita una valutazione approfondita dei modelli lessicali e semantici [3]. L'obiettivo di questa valutazione è stato capire se i modelli lessicali e semantici condividano caratteristiche comuni, se i loro segnali siano complementari e come possano essere combinati per affrontare efficacemente il gap semantico. In particolare, sono stati valutati gli aspetti critici dei modelli neurali semantici attraverso svariate analisi. Ogni analisi ha portato una prospettiva differente nella comprensione dei modelli semantici e della loro relazione con i modelli lessicali. I risultati di questa valutazione hanno evidenziato la natura complementare dei segnali di corrispondenza lessicale e semantica e la necessità di combinarli nelle prime fasi della pipeline di IR per affrontare efficacemente il gap semantico. Pertanto i modelli neurali semantici in grado di reperire documenti dall'intera collezione sono più adatti a questo compito rispetto ai modelli neurali che eseguono la ri-classificazione di una serie di documenti inizialmente reperiti da un modello lessicale. Infatti, il reperimento di documenti dall'intera collezione consente ai modelli neurali semantici di identificare quei documenti rilevanti che sono maggiormente affetti dal gap semantico, come ad esempio i documenti rilevanti che non contengono alcun termine della query.

Basandosi su quanto emerso da questa valutazione sono stati quindi sviluppati modelli lessicali e semantici in grado di affrontare il gap semantico. A questo proposito sono stati sviluppati modelli che integrano la conoscenza da risorse esterne e sono stati valutati nel dominio medico. Il dominio medico è un dominio con un forte impatto sociale, in cui il semantic gap è prominente, e dove la grande presenza di risorse di conoscenza specialistiche – curate manualmente da professionisti – consente di esplorare metodi efficaci per sfruttare la conoscenza esterna e ridurre il gap semantico. Per quanto riguarda i modelli lessicali, sono stati proposti diversi metodi e tecniche per sfruttare la conoscenza contenuta all'interno di risorse esterne. L'utilizzo di risorse di conoscenza esterne per migliorare le rappresentazioni di query e/o documenti ha una lunga tradizione in IR ed è uno degli approcci più popolari per affrontare il gap semantico. All'interno di tale contesto, si è studiato come, e in che misura, concetti

e relazioni contenuti all'interno di risorse di conoscenza esterne possano essere integrati nelle rappresentazioni bag-of-words di query e documenti al fine di migliorare l'efficacia dei modelli lessicali. In [16, 22], si è indagato come, e se, le relazioni semantiche – identificate tra concetti estratti da documenti medici e collegati ad una risorsa di conoscenza esterna – possano essere sfruttate per reperire letteratura medica per scopi di supporto decisionale clinico (CDS). Sono stati proposti due metodi per identificare le relazioni all'interno di query e documenti: un metodo basato su regole ed un metodo basato sull'apprendimento automatico. Dall'analisi dei risultati è emerso che le relazioni, se pertinenti al bisogno informativo dell'utente, forniscono un contributo maggiore rispetto ai singoli concetti. Tuttavia, nella maggior parte delle query, l'utilizzo di relazioni semantiche per eseguire il recupero non ottiene prestazioni accettabili. Si è poi condotta una serie di studi ed analisi sui task proposti da TREC Precision Medicine (PM). È stato inizialmente condotto uno studio preliminare [10] per il task di TREC PM 2018 relativo agli studi clinici – dove l'obiettivo è quello di recuperare studi clinici rilevanti per i quali il paziente target è idoneo. In questo contesto, gli studi clinici rilevanti rappresentano il potenziale per collegare i pazienti a trattamenti sperimentali qualora i trattamenti esistenti siano stati inefficaci. A questo proposito, è stata proposta una procedura per: 1) espandere le query in modo iterativo, facendo affidamento su risorse di conoscenza mediche, aggiungendo alla query originale le varianti terminologiche di termini legati a neoplasie e geni; 2) filtrare gli studi clinici per i quali il paziente non è idoneo in base ai suoi dati personali. Lo scopo dello studio è stato duplice: (i) valutare come un approccio orientato al recall e basato su un metodo di espansione delle query iterativo e incrementalmente aggressivo influisce sulla precisione; (ii) indagare se l'efficacia dei modelli di IR possa essere correlata alla qualità delle relazioni contenute all'interno delle risorse di conoscenza utilizzate nel processo di espansione. I risultati sperimentali hanno mostrato che i modelli di reperimento funzionano meglio quando nessuna delle tecniche di espansione viene utilizzata. Le motivazioni dietro l'effetto negativo delle espansioni proposte sono due: (i) la mancanza di uno schema di pesatura appropriato per i termini delle query, (ii) l'uso di molteplici risorse di conoscenza indipendentemente dalla loro rilevanza per il task considerato. Tali motivazioni sono state affrontate ed approfondite in [8], dove l'analisi è stata estesa ad entrambi i task proposti da TREC PM – ossia reperimento di letteratura scientifica e di studi clinici. In tal senso sono state proposte e valutate diverse tecniche di espansione e riduzione delle query basate su risorse di conoscenza, al fine di valutare se un particolare approccio possa essere utile sia nel reperimento di letteratura scientifica che di studi clinici. I risultati sperimentali hanno mostrato che non emerge alcun pattern evidente per entrambi i task. Nel complesso, un approccio di espansione delle query favorisce il reperimento della letteratura scientifica, mentre un approccio di riduzione delle query migliora le prestazioni nel recupero di studi clinici. Ciononostante, la maggior parte delle riformulazioni proposte risulta efficace per entrambi i task. Inoltre, è stato possibile individuare una particolare combinazione in grado di ottenere prestazioni tra le migliori sia per TREC PM 2017 che 2018. Pertanto, l'analisi approfondita – derivante dallo studio preliminare – ha evidenziato l'efficacia di applicare uno schema di pesatura sui termini della query e di selezionare risorse di conoscenza appropriate per espandere e/o ridurre le query. Alla luce dei risultati ottenuti nell'analisi approfondita, è stato condotto uno studio di validazione sui dati di TREC PM 2019 in [7]. Entrambi i task sono stati presi in considerazione, con particolare attenzione al reperimento degli studi clinici. L'obiettivo di questo studio è stato duplice: (i) valutare come le diverse riformulazioni – testate sulle precedenti collezioni di TREC PM – influenzano i risultati e se i risultati ottenuti nell'analisi precedente rimangono validi; (ii) verificare se la fusione di differenti riformulazioni, basate su tecniche di espansione e riduzione, risulta efficace in un contesto così specifico. I risultati sperimentali hanno evidenziato l'efficacia delle riformulazioni precedentemente testate nel recupero di studi clinici rilevanti, dimostrando l'abilità delle riformulazioni proposte di rimanere efficaci in più edizioni (e collezioni). D'altro canto, i risultati ottenuti fondendo diverse riformulazioni hanno suggerito la necessità di tecniche più avanzate per combinare efficacemente i differenti segnali forniti da tali riformulazioni. A fronte dei risultati ottenuti, in [19] è stata eseguita un'analisi a posteriori sull'efficacia delle riformulazioni proposte per il reperimento di studi clinici nel corso delle tre edizioni di TREC PM. L'analisi ha confrontato l'approccio proposto con quelli sviluppati dai gruppi di ricerca che hanno partecipato a tutti e tre gli anni di TREC PM. I risultati sperimentali hanno dimostrato l'efficacia delle riformulazioni proposte rispetto ai sistemi partecipanti. Pertanto, tali riformulazioni possono essere utilizzate (e combinate) da sistemi di IR multistadio nelle prime fasi della pipeline di IR, al fine di ottenere un insieme più ricco di documenti rilevanti – riducendo così il gap semantico tra query e documenti.

Per quanto riguarda i modelli semantici, si è proposto in [11] un framework di reperimento per combinare rappresentazioni implicite di query/documenti – ottenute attraverso l'apprendimento distribuzionale – e rappresentazioni esplicite – derivate da risorse di conoscenza esterne – per ridurre il gap semantico nelle attività di recupero a fini di supporto decisionale clinico. La combinazione di rappresentazioni basate sull'apprendimento distribuzionale con rappresentazioni semantiche formali derivate da risorse di conoscenza mira ad arricchire la comprensione semantica di query e documenti, riducendo così il gap semantico. Infatti, anche se i modelli di apprendimento distribuzionale catturano relazioni latenti tra unità testuali – siano esse parole, frasi, o persino documenti – basandosi solo sui documenti, sono ostacolati da due limitazioni principali che l'uso di risorse di conoscenza esterne può alleviare. In primo luogo, i modelli di apprendimento distribuzionale non riescono a discriminare le parole polisemiche, poiché imparano rappresentazioni univoche per le parole indipendentemente dal contesto. In secondo luogo, i modelli di apprendimento distribuzionale non riescono ad apprendere rappresentazioni simili per sinonimi che occorrono in contesti differenti, poiché mancano della conoscenza relazionale necessaria per identificare le relazioni di sinonimia tra le parole. Pertanto, sulla base dei risultati della valutazione eseguita in [3], è stato perfezionato il framework proposto in [11] ed è stato presentato in [2] il Semantic-Aware Neural Framework for IR (SAFIR) – un framework neurale di apprendimento automatico non supervisionato e ottimizzato per il reperimento delle informazioni. SAFIR apprende rappresentazioni di parole, concetti e documenti. Le rappresentazioni apprese sono ottimizzate per il reperimento delle informazioni e codificano sia la polisemia che la sinonimia per affrontare il gap semantico tra query e documenti. SAFIR può essere applicato a qualsiasi dominio in cui siano disponibili risorse di conoscenza esterne e non richiede dati etichettati per la fase di apprendimento. Una valutazione sperimentale è stata condotta per confrontare SAFIR con dei language model neurali potenziati da risorse di conoscenza esterne sul reperimento di letteratura medica, un task fortemente legato al CDS. Sono state considerate due strategie di reperimento negli esperimenti: reperimento di documenti ed espansione di query. Il reperimento dei documenti ha offerto l'opportunità di studiare l'efficacia dell'integrazione della conoscenza esterna nei modelli neurali per il tipico scenario di IR, in cui i sistemi recuperano una serie di documenti data una query. L'espansione delle query ha consentito invece di indagare l'efficacia dei modelli neurali potenziati dalla conoscenza – progettati specificamente per affrontare il gap semantico – nel recupero di documenti per approcci basati su Pseudo Relevance Feedback (PRF). In altre parole, si è valutato se i modelli neurali potenziati dalla conoscenza forniscono termini di espansione efficaci nel ridurre il gap semantico per i modelli lessicali. I risultati sperimentali hanno mostrato che SAFIR recupera documenti più esclusivi rispetto ai language model neurali potenziati dalla conoscenza nella maggior parte delle query. Inoltre, la precisione di SAFIR, insieme alla sua esclusività in termini di documenti rilevanti recuperati, lo rende adatto a metodi basati su PRF. SAFIR, infatti, coglie segnali differenti rispetto ai modelli lessicali e recupera documenti più efficaci nel fornire termini di espansione per i metodi PRF. Pertanto, la valutazione ha evidenziato l'efficacia dell'utilizzo di SAFIR nelle prime fasi della pipeline di IR, dove i diversi segnali che fornisce possono essere utilizzati da sistemi IR multistadio per ottenere un pool più ricco di documenti rilevanti, portando così a risposte migliori per query semanticamente difficili.

La ricerca che si sta conducendo, insieme ai modelli e ai metodi proposti, si inserisce in un progetto di ricerca a lungo termine presentato per la prima volta in [12]. La ricerca ha l'obiettivo finale di ridurre il gap semantico e aumentare l'efficacia dei metodi di supporto decisionale clinico in scenari reali. Pertanto, si sta continuando a lavorare sulle tematiche descritte al fine di sviluppare sistemi all'avanguardia per CDS. In particolare, la ricerca si sta focalizzando sullo studio e lo sviluppo di metodi ed approcci di Entity Linking (EL) e di Relation Extraction (RE), dove l'obiettivo è quello di identificare ed estrarre concetti e relazioni contenute all'interno di documenti testuali per arricchire risorse di conoscenza che verranno poi utilizzate, ad esempio, dai metodi presentati sopra per supporto decisionale clinico.

### **3.2 Partecipazione a progetti di ricerca a livello europeo**

**EXA-MODE (Contratto n. 825292, durata del progetto: 2019 - 2022)** EXA-MODE (EXtreme-scale Analytics via Multimodal Ontology Discovery & Enhancement), è un progetto di tipo ICT co-finanziato dalla Commissio-



ne 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.** Membro dell'unità di ricerca dell'Università degli Studi di Padova.

### 3.3 Presentazioni a congressi internazionali

- *Exploring how to Combine Query Reformulations for Precision Medicine*, 28th Text REtrieval Conference (TREC 2019), Gaithersburg, Maryland, U.S.A., 13-15 novembre 2019.
- *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.
- *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.
- *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.
- *Implicit-Explicit Representations for Case-Based Retrieval*, 1st International Conference on Design of Experimental Search & Information REtrieval Systems (DESIRE 2018), Bertinoro, Italia, 28-31 agosto 2018.
- *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, U.S.A., 8-12 luglio 2018.
- *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

- *SAFIR: a Semantic-Aware Neural Framework for IR*, 11th Italian Information Retrieval Workshop (IIR 2021), Bari, Italia, 13-15 settembre 2021.
- *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.
- *Reproducibility of the Neural Vector Space Model via Docker*, 16th Italian Research Conference on Digital Libraries (IRCDL 2020), Bari, Italia, 30-31 gennaio 2020.
- *Knowledge Enhanced Representations for Clinical Decision Support*, 10th Italian Information Retrieval Workshop (IIR 2019), Padova, Italia, 16-18 settembre 2019.
- *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.

- *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.
- *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 Attività di peer-reviewing

#### Riviste

- Artificial Intelligence in Medicine
- IEEE Access
- Transactions on Information Systems (TOIS)
- The International Journal of Digital Libraries (IJDL)
- International Journal of Information Technology and Decision Making (IJITDM)

Si sono condotte attività di revisione di lavori scientifici con regolarità dal 2017 ad oggi per riviste scientifiche internazionali. Si sono condotte attività di revisione anche per diversi congressi internazionali dal 2017 ad oggi; si riportano i congressi dove si è svolto il ruolo di membro del comitato di programma.

#### Membro di comitati di programma di congressi internazionali

- **Full & short paper track.** The 44th European Conference on Information Retrieval (ECIR 2022).
- **Reproducibility paper track.** The 44th European Conference on Information Retrieval (ECIR 2022).
- **Regular paper track.** The 13th biennial conference on Recent Advances in Natural Language Processing (RANLP 2021).
- **Full paper track.** The 30th ACM International Conference on Information and Knowledge Management (CIKM 2021).
- **Applied paper track.** The 30th ACM International Conference on Information and Knowledge Management (CIKM 2021).
- **Reproducibility paper track.** The 43rd European Conference on Information Retrieval (ECIR 2021).

### 3.6 Partecipazione all'organizzazione di eventi scientifici

- **Proceedings chair.** The 2nd Design of Experimental Search & Information REtrieval Systems Conference (DESIRES 2021), Padova, Italia, dal 15 settembre al 18 settembre 2021.  
URL: <http://desires.dei.unipd.it/>
- **Publicity chair.** The 11th Italian Information Retrieval Workshop (IIR 2021), Bari, Italia, dal 13 settembre al 15 settembre 2021.  
URL: <https://sisinflab.github.io/iir2021/>

- **Organization chair.** The 17th Italian Research Conference on Digital Libraries (IRCDL 2021), Padova, Italia, dal 18 febbraio al 19 febbraio 2021.  
URL: <http://ircdl2021.dei.unipd.it/>
- **Student volunteer.** The 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019), Parigi, Francia, dal 21 Luglio al 25 Luglio 2019.  
URL: <https://sigir.org/sigir2019/>
- **Student volunteer.** The 38th European Conference on Information Retrieval (ECIR 2016), Padova, Italia, dal 20 Marzo al 23 Marzo 2016.  
URL: <http://ecir2016.dei.unipd.it/>

### 3.7 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.

### 3.8 Premi

- **Vincitore del Premio di Studio 2016 per Laureati Magistrali in Ingegneria Informatica** assegnato dalla Fondazione Luciano Iglesias.
- **Vincitore del SIGIR Student Travel Grant** finanziato da ACM SIGIR (ACM Special Interest Group on Information Retrieval) per partecipare a "The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval" (SIGIR 2018) tenutasi ad Ann Arbor, Michigan (U.S.A.) dall'8 al 12 luglio 2018.
- **Vincitore della ESSIR 2017 Scholarship** finanziata da ACM SIGIR (ACM Special Interest Group on Information Retrieval) per partecipare a "The 11th European Summer School in Information Retrieval" (ESSIR 2017) tenutasi a Barcellona (SP) dal 4 all'8 settembre 2017.

## 4 Attività didattica

Le attività didattiche sono state condotte negli anni accademici 2018/2019 e 2019/2020 per il numero di ore di didattica consentito dalla Scuola di Dottorato in Ingegneria dell'Informazione. Nell'anno accademico 2020/2021 si è ricominciata l'attività di didattica nei termini previsti dal contratto di assegnista di ricerca dell'Università degli Studi di Padova.

## 4.1 Didattica titolarità

### Anno accademico 2021/2022

- “Basi di Dati 1”, **3 CFU** per il corso di Laurea Triennale in Ingegneria Informatica dell'Università degli Studi di Padova.
- “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.

### 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.

## 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 Seminari

- *Query Reformulation Techniques for Precision Medicine*, seminario per il corso di Information Retrieval del Master in Computer Science dell'Università di Copenhagen, tenuto online, 13 maggio 2020.

## 5 Attività istituzionali, organizzative e di servizio

- **Coordinatore tutor junior** per il Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova. Anno accademico 2018/2019.
- *Novembre 2017 - Novembre 2019*: **Rappresentante dei dottorandi** nel Consiglio di Dipartimento di Ingegneria dell'Informazione dell'Università degli Studi di Padova.

## 6 Pubblicazioni

### Riviste internazionali (peer-reviewed)

- [1] S. Marchesin, G. M. Di Nunzio, and M. Agosti. 2021. Simple but Effective Knowledge-Based Query Reformulations for Precision Medicine Retrieval. *MDPI Information*, accepted for publication, September 2021, 28 pages, 2021.
- [2] M. Agosti, S. Marchesin, and G. Silvello. 2020. 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
- [3] S. Marchesin, A. Purpura and G. Silvello. 2020. 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

### Riviste internazionali

- [4] S. Marchesin. 2021. 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

### Atti di conferenze internazionali

- [5] G. Faggioli, and S. Marchesin. 2021. What Makes a Query Semantically Hard? in *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.
- [6] G.M. Di Nunzio, S. Marchesin, and F. Vezzani. 2020. A Study on Reciprocal Ranking Fusion in Consumer Health Search. IMS UniPD at CLEF eHealth 2020 Task 2 in *Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum (CLEF 2020)*, Thessaloniki, Greece, September 22-25, 2020, pp. 7.
- [7] G.M. Di Nunzio, S. Marchesin, and M. Agosti. 2019. Exploring how to Combine Query Reformulations for Precision Medicine in *Proceedings of the 28th Text REtrieval Conference (TREC 2019)*, Gaithersburg, Maryland, USA, November 13-15, 2019, pp. 14.
- [8] M. Agosti, G.M. Di Nunzio, and S. Marchesin. 2019. An Analysis of Query Reformulation Techniques for Precision Medicine in *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.
- [9] G.M. Di Nunzio, and S. Marchesin. 2018. The University of Padua IMS Research Group at CENTRE@TREC 2018 in *Proceedings of the 27th Text REtrieval Conference (TREC 2018)*, Gaithersburg, Maryland, USA, November 14-16, 2018, pp. 7.
- [10] M. Agosti, G.M. Di Nunzio, and S. Marchesin. 2018. The University of Padua IMS Research Group at TREC 2018 Precision Medicine Track in *Proceedings of the 27th Text REtrieval Conference (TREC 2018)*, Gaithersburg, Maryland, USA, November 14-16, 2018, pp. 10.
- [11] S. Marchesin. 2018. Implicit-Explicit Representations for Case-Based Retrieval in *Proceedings of the Design of Experimental Search & Information REtrieval Systems (DESIRES 2018)*, Bertinoro, Italy, August 28-31, 2018, p. 109.

- [12] S. Marchesin. 2018. Case-Based Retrieval Using Document-Level Semantic Networks in *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.

### **Atti di workshop**

- [13] N. Marini, S. Otálora, F. Ciompi, G. Silvello, S. Marchesin, S. Vatrano, G. Buttafuoco, M. Atzori, and H. Müller. 2021. Multi-Scale Task Multiple Instance Learning for the Classification of Digital Pathology Images with Global Annotations in *Proceedings of the MICCAI Computational Pathology (COMPAY) Workshop*, Strasbourg, France, September 27, 2021, pp. 170-181.
- [14] N. Ferro, S. Marchesin, A. Purpura, and G. Silvello. 2019. A Docker-Based Replicability Study of a Neural Information Retrieval Model in *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.
- [15] S. Marchesin. 2019. Knowledge Enhanced Representations to Reduce the Semantic Gap in Clinical Decision Support in *Proceedings of the 9th PhD Symposium on Future Directions in Information Access (FDIA 2019)*, Milan, Italy, July 17, 2019, pp. 4-9.
- [16] M. Agosti, G.M. Di Nunzio, S. Marchesin, and G. Silvello. 2018. A Relation Extraction Approach for Clinical Decision Support in *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.
- [17] 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. 2017. Towards Open-Source Shared Implementations of Keyword-Based Access Systems to Relational Data in *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**

- [18] M. Agosti, S. Marchesin, and G. Silvello. 2021. SAFIR: a Semantic-Aware Neural Framework for IR in *Proceedings of the 11th Italian Information Retrieval Workshop (IIR 2021)*, Bari, Italy, September 13-15, 2021, pp. 4.
- [19] M. Agosti, G.M. Di Nunzio, and S. Marchesin. 2020. A Post-Analysis of Query Reformulation Methods for Clinical Trials Retrieval in *Proceedings of the 28th Italian Symposium on Advanced Database Systems (SEBD 2020)*, Villasimius, Sardinia, Italy, June 21-24, 2020, pp. 152-159.
- [20] N. Ferro, S. Marchesin, A. Purpura, and G. Silvello. 2020. Reproducibility of the Neural Vector Space Model via Docker in *Proceedings of the 16th Italian Research Conference on Digital Libraries (IRCDL 2020)*, Bari, Italy, January 30-31, 2020, pp. 3-8.
- [21] S. Marchesin, and M. Agosti. 2019. Knowledge Enhanced Representations for Clinical Decision Support in *Proceedings of the 10th Italian Information Retrieval Workshop (IIR 2019)*, Padua, Italy, September 16-18, 2019, pp. 17-18.
- [22] M. Agosti, G.M. Di Nunzio, S. Marchesin, and G. Silvello. 2019. Medical Retrieval using Structured Information Extracted from Knowledge Bases in *Proceedings of the 27th Italian Symposium on Advanced Database Systems (SEBD 2019)*, Castiglione della Pescaia, Grosseto, Italy, June 16-19, 2019, pp. 8.

- [23] M. Agosti, G.M. Di Nunzio, N. Ferro, M. Maistro, S. Marchesin, N. Orio, C. Ponchia, and G. Silvello. 2018. Thirty Years of Digital Libraries Research at the University of Padua: The User Side in *Proceedings of the 14th Italian Research Conference on Digital Libraries (IRCDL 2018)*, Udine, Italy, January 25-26, 2018, pp. 42-54.
- [24] 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. 2017. Keyword-based access to relational data: To reproduce, or to not reproduce? in *Proceedings of the 25th Italian Symposium on Advanced Database Systems (SEBD 2017)*, Squillace Lido (Catanzaro), Italy, June 25-29, 2017, p. 166.
- [25] M. Agosti, S. Lawless, S. Marchesin, and V. Wade. 2017. An Adaptive Cross-Site User Modelling Platform for Cultural Heritage Websites in *Proceedings of the 13th Italian Research Conference on Digital Libraries (IRCDL 2017)*, Modena, Italy, January 26-27, 2017, pp. 132-141.

### **Tesi di dottorato**

- [26] S. Marchesin. 2021. 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

settembre 2021