# Filip Ilievski

Computer Scientist
University of Southern California, Information Sciences Institute
Marina del Rey, California, USA

Phone: +1 310-448-8712
email: ilievski@isi.edu
WEBSITE: https://usc-isi-i2.github.io/ilievski

## Education

03/2015-02/2019 **PhD in Natural Language Processing, Vrije Universiteit (VU) Amsterdam**

- Title: "Identity of Long-tail Entities in Text"

- Defense date: 6th of September 2019

- Computational Lexicology and Terminology Lab (CLTL) / Knowledge Representation and Reasoning (KRR)

- Promotors: Prof. Piek Vossen (CLTL, VU), Prof. Frank van Harmelen (KRR, VU); co-promotors: Dr. Stefan Schlobach (KRR, VU), Dr. Marieke van Erp (KNAW)

- My dissertation focused on multiple aspects of the under-addressed challenge of interpreting identity of long-tail entities in text, namely: description, analysis, evaluation, access to knowledge, and solution systems. It was part of the project Understanding Language By Machines.

09/2013-02/2015 **MSc in Artificial Intelligence, Vrije Universiteit Amsterdam**

- Specialization: Intelligent Systems Design

- VU Fellowship Program (VUFP) Recipient

- GPA: 8.4/10.00, Graduated Cum Laude

- Thesis title: "NED with two-stage coherence optimization – How i am teaching my Ford automobile not to run on 20th century presidential elections in US"

- Thesis supervisors/advisors: Dr. Marieke van Erp, Dr. Stefan Schlobach, Prof. Piek Vossen, Dr. Wouter Beek

09/2009-09/2013 **BSc in Computer Science and Engineering, Ss. Cyril and Methodius University, Skopje**

- Study Abroad - Vrije Universiteit Amsterdam, The Netherlands (Fall 2012 - Spring 2013, funded by the Erasmus Mundus Euroweb Scholarship

- GPA: 9.56/10.00

## Work Experience

**12/2019-present** Computer Scientist, USC Information Sciences Institute

- Projects: Machine Common Sense (MCS), Knowledge Graph Toolkit (KGTK)

**08/2020-12/2020** Part-time Lecturer, USC School of Engineering

- Taught the course *Building Knowledge Graphs (DSCI-558)* to 31 students.

**03/2019-11/2019** Postdoctoral Researcher, Vrije Universiteit Amsterdam

- Teaching, supervision, and research

- Within the Dutch FrameNet project, I set up an infrastructure that supports: 1. automatic acquisition of paired, multilingual data on events from Wikidata and Wikipedia; 2. iterative development and evaluation of semantic role labeling methods on this data; 3. development of corpora and the Dutch FrameNet.

- Within the Clariah project, I developed an unsupervised entity recognition and linking system for (historical) documents in Dutch.

- Employer: Prof. Piek Vossen

**02/2017-08/2017** Research Visitor, Language Technologies Institute, Carnegie Mellon University

- Project: The Profiling Machine - Active Generalization over Knowledge

- We built neural profiling machines that can provide human-like world expectations, implicit in communication and in knowledge bases. The models were trained on Wikidata, and evaluated against Wikidata entities, human judgments, as well as on a NIL clustering dataset. This work was published in the Journal of Web Semantics.

- Supervisor: Prof. Eduard Hovy

**11/2014-02/2015** Research Assistant, Vrije Universiteit Amsterdam

- Project: NewsReader

- During this period, I developed a contextually adapted Entity Linking module in Python. I also created a streaming architecture for inferring event instances and representing them in RDF format, in the Java programming language.

- Supervisors: Prof. Piek Vossen, Dr. Marieke van Erp

**02/2013-10/2014** Junior Software Developer, i-mapping, Amsterdam

- Project: Studio for Geographical visualization of statistical data

- Supervisor: Marijn Deurloo

**07/2011-08/2011** Intern, ATTE, Bitola, Macedonia

- Project: Web-based medical information system

- Supervisor: Dr. Aleksandar Stefanovski

## Awards and Scholarships

2020        SWSA Dissertation Award Nominee

09/2016     *Second place at the European Linked Open Data Project contest*, Semantics conference

- Awarded to our tool *LOTUS: A Search Engine for the Web of Data*

09/2013-02/2015     *VU Fellowship Programme (VUFP)* Scholarship for Academic Excellence

- VUFP funds (tuition fees + 3k euros/year) the Master's degree studies at Vrije Universiteit Amsterdam for a selection of talented prospective international students.
- Sponsor: Vrije Universiteit Amsterdam, The Netherlands
- Years: 2013-2014 and 2014-2015

09/2012-06/2013     *Erasmus Mundus Euroweb* Scholarship for Student Mobility

- Erasmus Mundus Euroweb is a scholarship programme for Western Balkans students on undergraduate, master, doctoral, and post-doctoral level who want to study/work in EU countries, and vice versa.
- Sponsor: European Union
- Year: 2012-2013

09/2010-06/2012     *Best 125 IT students in Macedonia* Scholarship

- This scholarship is awarded to the top 125 IT students in the country with an average GPA of 9.0 or higher.
- Sponsor: Ministry of Education and Science of Macedonia
- Years: 2010-2011 and 2011-2012

## Publications and talks

BOOK/CHAPTER

1. Filip Ilievski. Identity of Long-Tail Entities in Text (2019). Studies on the Semantic Web. IOS Press.

2. Filip Ilievski. Identity of Long-Tail Entities in Text (2019). PhD dissertation. Vrije Universiteit Amsterdam.

3. Wouter Beek, Laurens Rietveld, Filip Ilievski, Stefan Schlobach (2017). LOD Lab: Scalable Linked Data Processing. In Reasoning Web: Logical Foundation of Knowledge Graph Construction and Query Answering.

JOURNAL

4. (under review) Filip Ilievski, Alessandro Oltramari, Kaixin Ma, Bin Zhang, Deborah L. McGuinness, Pedro Szekely. Dimensions of Commonsense Knowledge. Knowledge-Based Systems (KBS).

5. Filip Ilievski, Eduard Hovy, Piek Vossen, Stefan Schlobach, Qizhe Xie (2019). The Role of Knowledge in Establishing Identity of Long-Tail Entities. Journal of Web Semantics (JWS), Special Issue on Language Technology and Knowledge Graphs. GitHub1, GitHub2, GitHub3.

- *(Earlier, unpublished version)* Filip Ilievski, Eduard Hovy, Qizhe Xie, Piek Vossen (2018). The Profiling Machine: Active Generalization over Knowledge. arXiv.

6. Wouter Beek, Filip Ilievski, Jeremy Debattista, Stefan Schlobach, Jan Wielemaker (2017). Literally Better: Analyzing and Improving the Quality of Literals. In proceedings of the Semantic Web Journal (SWJ). GitHub.

CONFERENCE

7. (under review) Ehsan Qasemi, Filip Ilievski, Muhao Chen, Pedro Szekely. CoreQuisite: Circumstantial Preconditions of Common Sense Knowledge . EMNLP.

8. (under review) Kaixin Ma, Filip Ilievski, Jonathan Francis, Satoru Ozaki, Eric Nyberg, Alessandro Oltramari. What Do Models Learn For Commonsense Question Answering? EMNLP.

9. (under review) Avijit Thawani, Jay Pujara, Filip Ilievski. Numeracy Enhances the Literacy of Language Models. EMNLP.

10. (under review) Kartik Shenoy, Filip Ilievski, Daniel Garijo, Daniel Schwabe, Pedro Szekely. A Study of the Quality of Wikidata. The International Semantic Web Conference (ISWC).

11. Peifeng Wang, Filip Ilievski, Muhao Chen, Xiang Ren. Do Language Models Perform Generalizable Commonsense Inference?. ACL Findings.

12. Avijit Thawani, Jay Pujara, Pedro Szekely, Filip Ilievski (2021). Representing Numbers in NLP: a Survey and a Vision. NAACL.

13. Filip Ilievski, Pedro Szekely, Bin Zhang (2021). CSKG: The CommonSense Knowledge Graph. ESWC Resource Track. GitHub.

14. Kaixin Ma, Filip Ilievski, Jonathan Francis, Yonatan Bisk, Eric Nyberg, Alessandro Oltramari (2021). Knowledge-driven Data Construction for Zero-shot Evaluation in Commonsense Question Answering. AAAI. GitHub.

15. Peifeng Wang, Nanyun Peng, Filip Ilievski, Pedro Szekely, Xiang Ren (2020). Connecting the Dots: A Knowledgeable Path Generator for Commonsense Question Answering. EMNLP Findings. GitHub.

16. Filip Ilievski, Daniel Garijo, Hans Chalupsky, Naren Teja, Yixiang Yao, Craig Rogers, Ronpeng Li, et al. (2020). KGTK: A Toolkit for High-Volume Knowledge Graph Manipulation. International Semantic Web Conference (ISWC), Resource Track. GitHub.

17. Piek Vossen, Filip Ilievski, Marten Postma, Antske Fokkens, Gosse Minnema, Levi Remijnse (2020). Large-scale Cross-lingual Language Resources for Referencing and Framing. LREC.

18. Filip Ilievski, Piek Vossen, Stefan Schlobach (2018). Systematic Study of Long Tail Phenomena in Entity Linking. In Proceedings of the 27th International Conference on Computational Linguistics (COLING). Slides. GitHub.

19. Piek Vossen, Filip Ilievski, Marten Postma, Roxane Segers (2018). Don't Annotate, but Validate: a Data-to-Text Method for Capturing Event Data. In proceedings of LREC 2018. GitHub1, GitHub2.

20. Piek Vossen, Marten Postma, Filip Ilievski (2018). ReferenceNet: a semantic-pragmatic network for capturing reference relations. In proceedings of the Global WordNet Conference (GWC). Slides.

21. Filip Ilievski, Piek Vossen, Marieke van Erp (2017). Hunger for Contextual Knowledge and a Road Map to Intelligent Entity Linking. In proceedings of Language, Data and Knowledge (LDK). Poster.

22. Filip Ilievski, Marten Postma, Piek Vossen (2016). Semantic overfitting: what 'world' do we consider when evaluating disambiguation of text?. In Proceedings of COLING. Poster. GitHub.

23. Filip Ilievski, Wouter Beek, Marieke van Erp, Laurens Rietveld, Stefan Schlobach (2016). LOTUS: Adaptive Text Search for Big Linked Data. In The Extended Semantic Web Conference (ESWC). Slides. GitHub1, GitHub2.

24. Filip Ilievski, Wouter Beek, Marieke van Erp, Laurens Rietveld, Stefan Schlobach (2016). LOTUS: Adaptive Text Search for Big Linked Data. In Proceedings of BNAIC. Poster.

25. Filip Ilievski, Giuseppe Rizzo, Marieke van Erp, Julien Plu, Raphael Troncy (2016). Context-enhanced Adaptive Entity Linking. In proceedings of LREC. Slides. GitHub.

26. Marieke van Erp, Pablo Mendes, Heiko Paulheim, Filip Ilievski, Julien Plu, Giuseppe Rizzo and Joerg Waitelonis (2016). Evaluating Entity Linking: An Analysis of Current Benchmark Datasets and a Roadmap for Doing a Better Job. In proceedings of LREC. Slides. GitHub.

27. Roxane Segers, Egoitz Laparra, Marco Rospocher, Piek Vossen, German Rigau, Filip Ilievski (2016). The Predicate Matrix and the Event and Implied Situation Ontology: Making More of Events. In proceedings of The Global Wordnet Conference (GWC). Slides.

WORKSHOP

28. Mark Mann, Filip Ilievski, Mohammad Rostami, Aastha Aastha, Basel Shbita. Open Drug Knowledge Graph . Knowledge Graph Construction (KGC), ESWC.

29. Ziping Hu, Zepei Zhao, Mohammad Rostami, Filip Ilievski, Basel Shbita. Knowledge Graph-Based Housing Market Analysis. Knowledge Graph Construction (KGC), ESWC. GitHub.

30. Filip Ilievski, Pedro Szekely, Daniel Schwabe (2020). Commonsense Knowledge in Wikidata. Wikidata workshop, ISWC.

31. Marten Postma, Levi Remijnse, Filip Ilievski, Antske Fokkens, Sam Titarsolej, Piek Vossen (2020). Combining Conceptual and Referential Annotation to Study Variation in Framing. International FrameNet Workshop, LREC.

32. Marten Postma, Filip Ilievski, Piek Vossen (2018). SemEval-2018 Task 5: Counting Events and Participants in the Long Tail. In proceedings of the SemEval-2018 workshop, NAACL. Slides. GitHub1, GitHub2.

33. Marten Postma, Filip Ilievski, Piek Vossen, Marieke van Erp (2016). Moving away from semantic overfitting in disambiguation datasets. In Proceedings of EMNLP 2016's UBLP (Uphill Battles in Language Processing) workshop. Poster.

34. Filip Ilievski, Wouter Beek, Marieke van Erp, Laurens Rietveld, Stefan Schlobach (2015). LOTUS: Linked Open Text UnleaShed. In The 6th International Workshop on Consuming Linked Data (COLD), ISWC. Slides.

35. Marieke van Erp, Filip Ilievski, Marco Rospocher, Piek Vossen (2015). Missing Mr. Brown and buying an Abraham Lincoln – Dark Entities and DBpedia. In The NLP and DBpedia Workshop, ISWC. Slides.

36. Tommaso Caselli, Piek Vossen, Marieke van Erp, Antske Fokkens, Filip Ilievski, Ruben Izquierdo Bevia, Minh Lê, Roser Morante, and Marten Postma (2015). When it's all piling up: investigating error propagation in an NLP pipeline. In: NLP Applications: completing the puzzle. Slides.

37. Piek Vossen, Tommaso Caselli, Filip Ilievski, Ruben Izquierdo, Alessandro Lopopolo, Emiel Van Miltenburg, Roser Morante, Ngoc Minh, Marten Postma (2015). Words in context: a reference perspective on the lexicon. MAPLEX workshop.

Abstract

38. Filip Ilievski, Daniel Garijo, Pedro Szekely (2021). KGTK: A Knowledge Graph Toolkit for Exploiting Large Knowledge Graphs. The Web Conference 2021, Developers Track. Slides.

39. Avijit Thawani, Jay Pujara, Filip Ilievski, Pedro Szekely (2020). Representing Numbers in Language Models. WeCNLP.

40. Ngoc Minh, Filip Ilievski (2015). A Look inside Babelfy: Examining the Bubble. Computational Linguistics in the Netherlands (CLiN). Slides. GitHub.

41. <u>Filip Ilievski</u>, Marieke van Erp, Piek Vossen, Stefan Schlobach, Wouter Beek (2015). Named Entity Disambiguation with two-stage coherence optimization. Computational Linguistics in the Netherlands (CLiN). Slides. GitHub.

### Deliverable

42. Marieke van Erp, Roxane Segers, <u>Filip Ilievski</u>, Antske Fokkens, Rodrigo Agerri, Marco Rospocher, Itziar Aldabe, Egoitz Laparra, German Rigau, Sanne Vrijenhoek (2016). D5. 2.2 Domain model for financial and economic events, version 3 Deliverable D5. 2.2.

43. Piek Vossen, Tommaso Caselli, Agata Cybulska, Antske Fokkens, <u>Filip Ilievski</u>, Anne-Lyse Minard, Paramita Mirza, Itziar Aldabe, Egoitz Laparra, German Rigau (2016). Event Narrative Module, version 3 Deliverable D5.1.3.. GitHub.

### Invited talks

| | |
|---|---|
| 05/2021 | Building Agents with Common Sense. Knowledge-infused Learning workshop, Knowledge Graph Conference. Invited keynote. Held virtually, May 3rd, 2021. Slides. |
| 04/2021 | Building Agents with Common Sense. Distributed Artificial Intelligence group, King's College London, United Kingdom. Invited talk. Slides. |
| 12/2020 | Reasoning with Commonsense Knowledge Graphs. Turing NLP, London, United Kingdom. Invited talk. Slides. |
| 04/2020 | Machine Common Sense and Other Quests at USC-ISI. NEC Labs, Germany. Invited talk. |

## Teaching

| | |
|---|---|
| 04/2021 | Building Knowledge Graphs (DSCI-558), Graduate course, University of Southern California, Guest lecturer, session "Commonsense Knowledge Graphs". |
| 10/2020 | Text Analytics (ISE-540), Graduate course, University of Southern California, Guest lecturer, session "Knowledge Graphs". |
| 08/2020-12/2021 | Building Knowledge Graphs (DSCI-558), Graduate course, University of Southern California, Lecturer. |
| 05/2019-07/2019 | Text Mining, Central Bureau for Statistics (CBS) Netherlands, Lab session coordinator. Material. |
| 05/2019 | Knowledge Representation on the Web, Master course, Vrije Universiteit Amsterdam, Guest lecturer, session "KR and NLP: The curious case of entities". |
| 01/2019-03/2019 | Text Mining, Bachelor course, Vrije Universiteit Amsterdam, Lecturer. Material. |
| 10/2017-01/2018 | Python for Text Analysis, Master course, Vrije Universiteit Amsterdam, Lecturer. Material. |
| 10/2017-01/2018 | Programming for Humanities and Social Sciences, Bachelor course, Vrije Universiteit Amsterdam, Lecturer. |
| 10/2017-01/2018 | Digital and Environmental Humanities, Master course, Vrije Universiteit Amsterdam, Lecturer. |
| 10/2017 | Databases, Master course, Rijksuniversiteit Groningen, Guest lecturer, session "NoSQL Databases". Slides. |
| 01/2016-03/2016 | Information processing (Informatieverwerking), Bachelor course, Vrije Universiteit Amsterdam, Lecturer. |

## Supervision

2020-      Ehsan Qasemi, PhD student. University of Southern California. Advisor.
2020-      Avijit Thawani, PhD student. University of Southern California. Advisor.
2020-      Peifeng Wang, PhD student. University of Southern California. Advisor.
2021-      Pushkaraj Jitendra Sarnobat. MSc volunteer. Advisor.
2021-      Kartik Shenoy. MSc student worker. Advisor.
2020-2021  Bin Zhang. MSc student worker. Advisor.
2020-2021  Hanzhi Zhang. MSc student worker. Advisor.
2020-      Nicholas Klein. MSc student worker/USC Rising star intern. Advisor.
2019       Jiri Brummer: Comparing various Machine Learning Methods for Element Extraction in Dutch Contracts. MSc thesis, Business Analytics, KPMG and VU Amsterdam. Supervisor.
2019       Guido Jansen: Low Resource Domain Specific Named Entity Recognition. MSc thesis, Information studies, Data science, University of Amsterdam. Second reader.
2018       Arash Parnia: Mathematical Entity Recognition System MERCS. MSc thesis, Artificial Intelligence, VU Amsterdam. Second reader.
2016       Sai Sreewathsa Kovullari: Entity Typing with Page Ranking. MSc thesis, Computer Science, VU Amsterdam. Second reader.

## Organizing Committee

**JOURNAL**

Filip Ilievski, Antoine Bosselut, Kenneth Forbus, Simon Razniewski, Vered Shwartz (2021). Special Issue on Commonsense Knowledge and Reasoning. Semantic Web Journal (SWJ).

**WORKSHOP**

Filip Ilievski, Alessandro Oltramari, Deborah McGuinness, Pedro Szekely (2021). Commonsense Knowledge Graphs (CSKGs). AAAI.

Filip Ilievski, Marten Postma, Piek Vossen (2016). 2nd Spinoza workshop: Looking at the Long Tail. Vrije Universiteit Amsterdam.

**TUTORIAL**

Filip Ilievski, Daniel Garijo, Hans Chalupsky, Pedro Szekely (2021). KGTK: Tools for Creating and Exploiting Large Knowledge Graphs. International Semantic Web Conference (ISWC).

Filip Ilievski, Antoine Bosselut, Simon Razniewski, Mayank Kejriwal (2021). Commonsense Knowledge Acquisition and Representation. AAAI.

Filip Ilievski, Pedro Szekely, Mayank Keyriwal (2020). Commonsense Knowledge Graphs. International Semantic Web Conference (ISWC).

Wouter Beek, Javier Fernandez, Paul Groth, Filip Ilievski (2016). Join the LOD Lab! Scale your Linked Data evaluations to the Web. Extended Semantic Web Conference (ESWC).

**COMPETITION**

Marten Postma, Filip Ilievski, Piek Vossen (2018). SemEval-2018 Task: Counting Events and Participants within Highly Ambiguous Data covering a very long tail, SemEval-2018 workshop, co-located with NAACL.

## Program Committee and Reviewing

JOURNAL

Semantic Web Journal (SWJ), various tracks

Machine Learning Journal, Special Track on Learning and Reasoning, 2020

Journal of Web Semantics (JWS), Special Issue on Language Technology and Knowledge Graphs, 2020

Journal of Natural Language Engineering, Cambridge University Press, 2019 and 2020

Data Science: Methods, Infrastructure, and Applications, iOS, 2017

Information Processing and Management, Elsevier, 2016

CONFERENCE

CIKM 2021 (short)

EMNLP 2021

ACL-IJCNLP 2021

ACM Hypertext 2021

ESWC 2021

ACM Symposium On Applied Computing (SAC) 2021

AAAI 2021

International Joint Conferences on Artificial Intelligence (IJCAI) 2021

Global Wordnet Conference (GWC) 2021

AACL-IJCNLP 2020

COLING 2020

CIKM 2020 (short)

Extended Semantic Web Conference (ESWC) 2020

Language Resources and Evaluation Conference (LREC) 2020

Biographical Data in a Digital World 2019 (BD2019)

Global Wordnet Conference (GWC) 2019

COLING 2018 (listed as 'outstanding reviewer')

SEMANTiCS 2018 ('Posters and Demos' track)

Biographical Data in a Digital World (BD) 2017

Language, Data and Knowledge (LDK) 2017

Extended Semantic Web Conference (ESWC) 2017 (secondary reviewer)

COLING 2016 (secondary reviewer)

SEMANTiCS 2016

Computational Linguistics in the Netherlands (CLiN) 2015

WORKSHOP

Wikidata workshop at ISWC 2021

EACL 2021 Student Research Workshop (SRW)

International Workshop on Semantic Evaluation 2021 (SemEval-2021)

International Workshop on Semantic Evaluation 2020 (SemEval-2020)

International Workshop on Semantic Evaluation 2018 (SemEval-2018)

Joint Conference on Lexical and Computational Semantics (*SEM 2018)

Open Knowledge Base Question Answering (OKBQA) 2016

#Microposts2016: Making Sense of Microposts, NEEL 2016 challenge

NLP Applications: completing the puzzle workshop 2015

European Summer School in Logic, Language and Information (ESSLLI) 2017

## Technical skills

- **Specialties:** Commonsense Reasoning, Knowledge Graphs, Information Extraction, Semantic Web, Machine Learning, Language Modeling, Knowledge Integration

- **Languages:** Python, Java, Node.js, HTML/Javascript/CSS, C#, Matlab, R, C, C++

- **Systems, Tools, and Packages:** SciKit Learn, t-SNE, Sentence-transformer, graph-tool, NetworkX, TensorFlow, SpaCy, NLTK, Jupyter Notebooks, Pandas, Matplotlib

- **Databases:** SQL, Neo4j, ElasticSearch, Redis, Couchbase, MongoDB, Semantic Web technologies (RDF, SPARQL, ontologies)

## Other Experience

- Volunteer at HackYourFuture, teaching refugees to code

- Participation in various civil and student NGOs in Europe: EESTEC, IAESTE, Youth Cultural Center (MKC) Bitola

## Languages (CEFR levels)

- **Macedonian** Native speaker

- **English** C2

- **Serbocroatian** C1

- **Dutch** B1

- **Spanish** A2

- **German** A1