

# 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](mailto:ilievski@isi.edu)  
HOMEPAGE: <http://ilievski.nl>

## Education

- 03/2015-02/2019 PhD in Natural Language Processing, Vrije Universiteit (VU) Amsterdam
- Title: “Identity of Long-tail Entities in Text”
  - Defense ceremony date: 6th of September 2019
  - Computational Lexicology and Terminology Lab (CLTL) / Knowledge Representation & 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 (CLTL, VU/KNAW Humanities Cluster)
  - My research 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

- 03/2019-present    Postdoctoral Researcher, Vrije Universiteit Amsterdam
- Projects: Dutch FrameNet and Clariah; plus teaching and supervision
  - Within the Dutch FrameNet project, I am setting up an infrastructure that supports:
    1. automatic acquisition of paired, multilingual data on events from Wikidata and Wikipedia;
    2. iterative development and evaluation of SRL systems on this data;
    3. development of other resources, including corpora and a Dutch FrameNet.
  - Within the Clariah project, I am developing an entity recognition and linking system for (historical) documents in Dutch. The system is unsupervised, iterative, and based on state-of-the-art embeddings.
  - Employer: Prof. Piek Vossen
- 02/2017-08/2017    Research Visit, Language Technologies Institute, Carnegie Mellon University
- Project: The Profiling Machine - Active Generalization over Knowledge
  - During this visit, I investigated how to build profiling machines that can provide human-like world expectations, implicit in communication and in knowledge bases. For this purpose, two neural architectures were designed (an autoencoder and an embeddings-based model), implemented in Theano, trained on Wikidata, and evaluated against Wikidata entities, human judgments, as well as on a NIL clustering dataset. The resulting paper is currently under review at JWS.
  - 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
  - We developed a full-stack engine for flexible visualization of statistical data on political maps. For comparison of performance, the same functionality was implemented in multiple languages (Python and Node.js), and the data was stored in multiple databases (flat files, MySQL, Redis, Couchbase).
  - Supervisor: Marijn Deurloo
- 07/2011-08/2011    Intern, ATTE, Bitola, Macedonia
- We created a medical information system in PHP, Wordpress, and MySQL.
  - Supervisor: Dr. Aleksandar Stefanovski

## Awards and Scholarships

- 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 & 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, 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.

5. 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

6. (*Under review*) Filip Ilievski, Pedro Szekely, Jingwei Cheng, Fu Zhang, Ehsan Qasemi. Consolidating Common Sense Knowledge. AKBC 2020.

7. (*Under review*) Kexuan Sun, Yixiang Yao, Filip Ilievski, Pedro Szekely (2020). A Frequency-Based Blocking System for Record Linkage. VLDB.

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

9. 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 2018). Santa Fe, New Mexico, USA. [Slides](#). [GitHub](#).

10. 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](#).

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

12. 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) 2017 Conference. [Poster](#).
13. Filip Ilievski, Marten Postma, Piek Vossen (2016). [Semantic overfitting: what ‘world’ do we consider when evaluating disambiguation of text?](#). In Proceedings of COLING 2016. Osaka, Japan. [Poster](#). [GitHub](#).
14. 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). Crete, Greece. [Slides](#). [GitHub1](#), [GitHub2](#).
15. Filip Ilievski, Wouter Beek, Marieke van Erp, Laurens Rietveld, Stefan Schlobach (2016). LOTUS: Adaptive Text Search for Big Linked Data. In Proceedings of BNAIC 2016. Amsterdam, The Netherlands. [Poster](#).
16. Filip Ilievski, Giuseppe Rizzo, Marieke van Erp, Julien Plu, Raphael Troncy (2016). [Context-enhanced Adaptive Entity Linking](#). In proceedings of LREC 2016. Portoroz, Slovenia. [Slides](#). [GitHub](#).
17. 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 2016. Portoroz, Slovenia. [Slides](#). [GitHub](#).
18. 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) 2016. Bucharest, Romania. January 27-30th, 2016. [Slides](#).

## WORKSHOP

19. 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 2018. New Orleans, Louisiana, USA. [Slides](#). [GitHub1](#), [GitHub2](#).
20. 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. Austin, Texas. [Poster](#).

21. 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 (COLLD), ISWC. Bethlehem, Pennsylvania, USA. October 12th, 2015. [Slides](#).
22. 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 & DBpedia Workshop, ISWC. Bethlehem, Pennsylvania, USA. October 12th, 2015. [Slides](#).
23. 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. Passau, Germany. 16 June 2015. [Slides](#).
24. 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](#). Workshop on Multiple Approaches for Multilingual frame Semantics wordnet generative (MAPLEX-2015), Yamagata, Japan.

#### ABSTRACT

25. Ngoc Minh, Filip Ilievski (2015). [A Look inside Babelify: Examining the Bubble](#). Presented at: CLiN 26. [Slides](#). [GitHub](#).
26. Filip Ilievski, Marieke van Erp, Piek Vossen, Stefan Schlobach, Wouter Beek (2015). [Named Entity Disambiguation with two-stage coherence optimization](#). Presented at: CLiN 25. [Slides](#). [GitHub](#).

#### DELIVERABLE

27. Marieke van Erp, Roxane Segers, Filip Ilievski, 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](#).
28. Piek Vossen, Tommaso Caselli, Agata Cybulska, Antske Fokkens, Filip Ilievski, Anne-Lyse Minard, Paramita Mirza, Itziar Aldabe, Egoitz Laparra, German Rigau (2016). [Event Narrative Module, version 3 Deliverable D5.1.3.. GitHub](#).

## Technical skills

- Daily experience with Python's scientific stack: Scikit-learn, Matplotlib/Seaborn, Notebooks, Networkx, Pandas, Theano, TensorFlow ...
- Some experience with many other languages: Node.js, HTML/Javascript/CSS, C#, Matlab, Java, R, C, C++
- Databases: (My/MS)SQL, Neo4j, ElasticSearch, Redis, Couchbase, MongoDB
- Experience with Semantic web: RDF, SPARQL, LDF, ontologies
- Formats: RDF, XML, JSON, CSV/TSV
- Knowledge of algorithms, machine learning, and neural networks
- Statistics and linear algebra
- Crowdsourcing with Figure Eight
- Linux, Mac OS X
- Unified Modeling Language (UML)

## Teaching

|           |   |
|-----------|---|
| 2018-2019 | Text Mining, Central Bureau for Statistics (CBS) Netherlands, Lab session coordinator. <a href="#">Material</a> .                                     |
| 2018-2019 | Knowledge Representation on the Web, Master course, Vrije Universiteit Amsterdam, Guest Lecturer, session "KR and NLP: The curious case of entities". |
| 2018-2019 | Text Mining, Bachelor course, Vrije Universiteit Amsterdam, Lecturer. <a href="#">Material</a> .  |
| 2017-2018 | Python for Text Analysis, Master course, Vrije Universiteit Amsterdam, Lecturer. <a href="#">Material</a> .   |
| 2017-2018 | Programming for Humanities and Social Sciences, Bachelor course, Vrije Universiteit Amsterdam, Lecturer.  |
| 2017-2018 | Digital and Environmental Humanities, Master course, Vrije Universiteit Amsterdam, Lecturer.  |
| 2017-2018 | Databases, Master course, Rijksuniversiteit Groningen, Guest Lecturer, session "NoSQL Databases". <a href="#">Slides</a> .                            |
| 2015-2016 | Information processing (Informatieverwerking), Bachelor course, Vrije Universiteit Amsterdam, Lecturer.   |

## Supervision

- 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

### COMPETITION

Marten Postma, Filip Ilievski, Piek Vossen (2016). [SemEval-2018 Task: Counting Events and Participants within Highly Ambiguous Data covering a very long tail](#), SemEval-2018 workshop, co-located with NAACL 2018

### WORKSHOP

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

### TUTORIAL

Filip Ilievski, Pedro Szekely, Mayank Keyriwal (2020). Commonsense Knowledge Graphs. ISWC 2020.

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) 2016.

## Program Committee and Reviewing

### JOURNAL

[Language Technology and Knowledge Graphs](#), Special issue, Journal of Web Semantics (JWS), 2019



[Journal of Natural Language Engineering](#), Cambridge University Press, 2019

[Data Science: Methods, Infrastructure, and Applications](#), iOS, 2017

[Information Processing and Management](#), Elsevier, 2016

#### CONFERENCE

[COLING 2020](#)

[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 2015 \(CLiN26\)](#)

#### WORKSHOP

[International Workshop on Semantic Evaluation 2020 \(SemEval-2020\)](#)

[International Workshop on Semantic Evaluation 2018 \(SemEval-2018\)](#), co-located with NAACL 2018

[Joint Conference on Lexical and Computational Semantics \(\\*SEM 2018\)](#), co-located with NAACL 2018

[Open Knowledge Base Question Answering \(OKBQA\) 2016](#), co-located with COLING 2016

[NEEL \(Named Entity rEcognition and Linking\) 2016](#) challenge, workshop #Microposts2016:  
Making Sense of Microposts

[NLP Applications: completing the puzzle workshop 2015](#), co-located with NLDB2015

[European Summer School in Logic, Language and Information \(ESSLLI\) 2017](#) (student session)

## Other Experience

- Volunteer at [HackYourFuture](#), teaching refugees to code
- Participation in various civil and student NGOs

## Languages (CEFR levels)

- **Macedonian** Native speaker
- **English** C2
- **Serbocroatian** C1
- **Dutch** B1
- **Spanish** A2
- **German** A1