Filip Ilievski

Computational Lexicology & Terminology Lab/Knowledge Representation & Reasoning
Vrije Universiteit Amsterdam
The Netherlands

Phone: +31 63 6489042
email: f.ilievski@vu.nl
HOMEPAGE: http://ilievski.nl

Education

03/2015-present PhD in Natural Language Processing, Vrije Universiteit Amsterdam

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

- Project: Understanding Language By Machines
  (http://www.understandinglanguagebymachines.org/)

- Supervisors: Prof. Dr. Piek Vossen, Dr. Marieke van Erp (CLTL, VU), Prof. Dr. Frank van Harmelen, Dr. Stefan Schlobach (KRR, VU)

- My research focuses on multiple aspects of the underaddressed challenge of interpreting long-tail identities in text, namely: description, analysis, evaluation, access to knowledge, and solution systems.

- Expected graduation date: March 2019

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

- Specialization: Intelligent Systems Design

- GPA: 8.4/10.00, Graduated Cum Laude

- VU Fellowship Program (VUFP) Recipient

- 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. Dr. 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)
  Erasmus Mundus Euroweb Scholarship Recipient

- GPA: 9.56/10.00

Experience

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

- Project: The Profiling Machine - Active Generalization over Knowledge

- Supervisor: Prof. Dr. Eduard Hovy

- 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, and validated against human judgments. The resulting paper is currently under anonymous review at AAAI'19.

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

- Project: NewsReader

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

- 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.

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

- Project: Studio for Geographical visualization of statistical data

- Supervisor: Marijn Deurloo

- 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).

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

- Supervisor: Dr. Aleksandar Stefanovski

- We created a medical information system in PHP, Wordpress, and MySQL.

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 Program (VUFP)* Scholarship for Academic Excellence

- Sponsor: Vrije Universiteit Amsterdam, The Netherlands

- Years: 2013-2014 and 2014-2015

2

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

- Sponsor: European Union

- Year: 2012-2013

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

- Sponsor: Ministry of Education and Science of Macedonia

- Years: 2010-2011 and 2011-2012

Technical skills

- Daily experience with Python's scientific stack: Theano, Scikit-learn, Matplotlib/Seaborn, Notebooks, Networkx, Pandas, ...

- 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

- UML

Software

- codalab

- https://github.com/cltl/Profiling Code from the paper 'The Profiling Machine: Active Generalization over Knowledge', where we build two neural architectures for generating expectations for implicit/missing values of people profiles.

- https://github.com/cltl/AnnotatingEntityProfiles All code from the annotation tool that was used to mark evidence for entity properties in text, but also complement the structured data with new knowledge found in text.

- https://github.com/cltl/LongTailAnnotation The entire code from the first version of this tool, used to annotate mentions of events in text based on structured knowledge.

- https://github.com/cltl/EL-long-tail-phenomena Analysis of various long-tail properties of entity linking datasets in relation to system performance.

- https://github.com/filipdbrsk/LOTUS_Indexer All code that was used to create the LOTUS index, the largest public text search index to the LOD cloud.

- https://github.com/filipdbrsk/LOTUS_Search All code that powers the search API and website of LOTUS (the largest public text search index to the LOD cloud), once the index is already created.

- https://github.com/D2KLab/relink Entity Linking system based on several human-inspired heuristics for domain adaptation, written in Python.

- https://github.com/cltl/HumanLikeEL Entity linking system inspired by human classification of knowledge types for framing.

- https://github.com/cltl/News2RDF A pipeline for conversion of news documents to RDF knowledge.

- https://github.com/cltl/SemEval2018-5_Postprocessing Postprocessing scripts for SemEval-2018 task 5.

- https://github.com/cltl/GunViolenceCorpus The Gun Violence Corpus, mostly annotated during the organization of the SemEval-2018 task 5.

- https://github.com/cltl/LongTailQATask The entire code for the SemEval-2018 task 5.

- https://github.com/cltl/SemanticOverfitting The code used to analyze datasets from five disambiguation and reference tasks using a set of statistics.

- https://github.com/globalwordnet/ili RDF backend of the Wordnet InterLingual Index.

- https://github.com/dbpedia-spotlight/evaluation-datasets Analysis of evaluation datasets for Entity Linking

- https://github.com/filipdbrsk/LODLanguages Analysis of the state of the language literals in the LOD cloud.

- https://github.com/filipdbrsk/EventCoreference Implementation of the event coreference module used in the NewsReader project pipeline.

- https://github.com/filipdbrsk/ned-twophaseoptimization Master thesis code of the entity linking modules.

- https://github.com/cltl/python-for-text-analysis Material of the course 'Python for text analysis'

Publications & talks

JOURNAL

1. 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). LINK

CHAPTER

2. 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. LINK

CONFERENCE

3. Filip Ilievski, Piek Vossen, Stefan Schlobach (2018). Systematic Study of Long Tail Phenomena in Entity Linking. Accepted at COLING 2018. Santa Fe, New Mexico, USA. PDF

4. 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. PDF

5. 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. PDF

6. 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. PDF

7. 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. PDF

8. 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 (EWSC). Crete, Greece. PDF

9. Filip Ilievski, Giuseppe Rizzo, Marieke van Erp, Julien Plu, Raphael Troncy (2016). Context-enhanced Adaptive Entity Linking. In proceedings of LREC 2016. Portoroz, Slovenia. PDF

10. 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. PDF

11. 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. PDF

WORKSHOP

12. 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. PDF

13. 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. PDF

14. 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. Bethlehem, Pennsylvania, USA. October 12th, 2015. PDF

15. 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 11th or 12th, 2015. PDF

16. 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. PDF

17. 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 Mul-

tilingual frame Semantics wordnet generative (MAPLEX-2015), Yamagata, Japan. PDF

POSTER &
ABSTRACT

18. 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.

19. Ngoc Minh, Filip Ilievski (2015). A Look inside Babelfy: Examining the Bubble. Presented at: CLiN 26. LINK

20. Filip Ilievski, Marieke van Erp, Piek Vossen, Stefan Schlobach, Wouter Beek (2015). Named Entity Disambiguation with two-stage coherence optimization. Presented at: CLiN 2015. LINK

DELIVERABLE

21. 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. PDF

22. 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. PDF

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

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

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

Information Processing and Management, Elsevier, 2016

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), 2018 (listed as 'outstanding reviewer')

SEMANTiCS 2016, 2018 ('Posters and Demos' track)

Computational Linguistics in the Netherlands 2015 (CLiN26)

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)

Teaching

11/2017-01/2018 Python for Text Analysis, Vrije Universiteit Amsterdam, Co-instructor

11/2017-01/2018 Programming for Humanities and Social Sciences, Vrije Universiteit Amsterdam, Co-instructor

11/2017 Digital and Environmental Humanities, Vrije Universiteit Amsterdam, Guest Lecturer

10/2017 Databases, Rijksuniversiteit Groningen, Guest Lecturer

02/2016 Information processing (Informatieverwerking), Vrije Universiteit Amsterdam, Guest Lecturer

Other Experience

- Volunteer at HackYourFuture, teaching refugees to code
- Participation in various civil and student NGOs

Language Skills

- **Macedonian** Native language
- **English** Highly fluently spoken and written
- **Serbocroatian** Highly fluently spoken and written
- **Dutch** Intermediate level of speaking and writing
- **Spanish** Basic level of speaking and writing