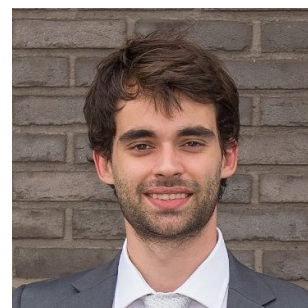


# Lucas Sterckx

✉ [lucassterckx@intec.ugent.be](mailto:lucassterckx@intec.ugent.be) | ☎ +32 (0)9 33 14979  
 🌐 <http://lusterck.github.io> | 🐦 @lusterck  
 📍 Technologiepark-Zwijnaarde 15, 9052 Ghent, Belgium



## RESEARCH STATEMENT

I am a fourth year doctoral student at Ghent University and currently looking for opportunities to collaborate with experienced researchers to work on difficult natural language processing problems. In the past I have worked on several information extraction problems. Since recently, my research focuses on sequence-to-sequence models, neural language models and neural machine translation.

## RESEARCH

### GHENT UNIVERSITY - IMEC | GHENT, BE

INTERNET BASED COMMUNICATION NETWORKS AND SERVICES (IBCN)

DEPARTMENT OF INFORMATION TECHNOLOGY (INTEC)

2013 – ... | DOCTORAL STUDENT

- **Topic Models [1, 2]**

For my master thesis, I did research on topic modeling in large corpora of song lyrics. My research was on better, automatic evaluation of topic models and measuring their interpretability. For this, I invented a new evaluation measure which has higher correlation with human evaluation than standard methods.

- **Knowledge Base Population [3, 4, 5, 6]**

I was the developer of the Ghent University's knowledge base population system for two participations in the 2014 and 2015 English Slot Filling tracks at the Text Analysis Conference benchmarks organized by NIST. In 2015, this system placed 2<sup>nd</sup> out of 12 systems. My technique for relation extraction used low dimensional embeddings of relation patterns to filter distantly supervised training data and train better supervised classifiers.

- **Keyphrase Extraction [7, 8, 9]**

I was researcher for the keyphrase extraction track of STEAMER, a government funded project in collaboration with some of Flanders' biggest media providers. In this context, I was a visiting researcher at the University of North Texas to work with prof. Cornelia Caragea and her group, to research methods to counter noise in training data for keyphrase extraction. The developed techniques were made commercially available via a spin-off company of the research group.

- **Sequence To Sequence Models**

Currently, I am working on models which automatically annotate texts which are difficult to understand or interpret by readers. For this, I use techniques from the field of neural machine translation.

## TEACHING

- Teaching Assistant for the introductory course on programming for first-year engineering students
- Supervising student projects for data mining and engineering courses
- Private tutoring of undergrad students in math and statistics courses

## EDUCATION

### GHENT UNIVERSITY

2011 - 2013 | IR. MSc IN COMPUTER SCIENCE ENGINEERING

Ghent | Graduated with honors

### VRIJE UNIVERSITEIT BRUSSEL

2007 - 2011 | ING. MSc IN ELECTRONICS AND INFORMATION TECHNOLOGY ENGINEERING

Brussels | Graduated with honors

## REVIEWING

Knowledge Based Systems  
Knowledge Extraction Workshop 2014, 2015

## AWARDS

2011 - Award for Outstanding Work on Thesis by VUB  
2015 - FWO Travel Scholarship

## LANGUAGES

### PROGRAMMING

Experience with programming in:  
Python • Shell • Java • C++ • JavaScript • Matlab • R  
Experience with Machine Learning and Deep  
Learning Frameworks:  
Scikit-Learn • Tensorflow • Keras • Lasagne

### SPOKEN & WRITTEN

Fluent  
English, Dutch  
Intermediate:  
French  
Elementary:  
German

## PUBLICATIONS

- [1] L. Sterckx, "Topic detection in a million songs," 2013. Diss. MSc in Computer Science Engineering.
- [2] L. Sterckx, T. Demeester, J. Deleu, L. Mertens, and C. Develder, "Assessing quality of unsupervised topics in song lyrics," in ECIR, vol. 8416 of Lecture Notes in Computer Science, pp. 547–552, Springer, 2014.
- [3] M. Feys, L. Sterckx, L. Mertens, J. Deleu, T. Demeester, and C. Develder, "Ghent University-IBCN participation in TAC-KBP 2014 slot filling and cold start tasks," in 7th Text Analysis Conference (TAC 2014), pp. 1–10, 2014.
- [4] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "Using active learning and semantic clustering for noise reduction in distant supervision," in 4e Workshop on Automated Base Construction at NIPS 2014 (AKBC-2014), pp. 1–6, 2014.
- [5] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "Ghent University-IBCN participation in the TAC KBP 2015 cold start slot filling task," in 8th Text Analysis Conference, 2015.
- [6] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "Knowledge base population using semantic label propagation," Knowledge-Based Systems, vol. 108, pp. 79 – 91, 2016. New Avenues in Knowledge Bases for Natural Language Processing.
- [7] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "Topical word importance for fast keyphrase extraction," in WWW (Companion Volume), pp. 121–122, ACM, 2015.
- [8] L. Sterckx, T. Demeester, J. Deleu, and C. Develder, "When topic models disagree: Keyphrase extraction with multiple topic models," in WWW (Companion Volume), pp. 123–124, ACM, 2015.
- [9] L. Sterckx, C. Caragea, T. Demeester, and C. Develder, "Supervised keyphrase extraction as positive unlabeled learning," in Proc. Conf. Empirical Methods in Natural Lang. Proc. (EMNLP 2016), (Austin, TX, USA), 1–5 Nov. 2016.

## REFERENCES

I was supervised and have worked closely together with the people listed below.

### PROF. CHRIS DEVELDER

Department of Information Technology  
Internet Based Communication Networks and Services  
Ghent University - iMinds  
Technologiepark-Zwijnaarde 15, 9052 Gent, Belgium  
chris.develder@intec.ugent.be

### DR. THOMAS DEMEESTER

Department of Information Technology  
Internet Based Communication Networks and Services  
Ghent University - iMinds  
Technologiepark-Zwijnaarde 15, 9052 Gent, Belgium  
thomas.demeester@intec.ugent.be

### PROF. CORNELIA CARAGEA

Computer Science and Engineering  
University of North Texas  
F228 Discovery Park, Denton, TX 76203, USA  
cornelia.caragea@unt.edu