

Simon Hengchen

PhD Candidate at Université libre de Bruxelles

Contact

CP 123 - Avenue F.D.
Roosevelt 50
1050 Brussels
Belgium

+32 2 650 44 58
+32 472 42 75 71

shengche@ulb.ac.be
hengchen.net
[Inkdin.me/hengchen](https://in.linkedin.com/in/hengchen)
git.io/v6kFV

Languages

French mother tongue
English & Dutch fluency
(written and spoken)
very basic Bokmål
learning Hebrew

IT Skills

Python ♥, C++
L^AT_EX
SPARQL, SQL
PHP, JavaScript &
HTML5
some Java
linux as main OS
RDF, Semantic Web
technologies

Education

- 2013- **PhD in Information Science / Natural Language Processing**
Dissertation on Information Extraction in a multilingual, historical, non-structured context. Using regression models (LDA) to detect semantic neologisms. I work with historians from different universities on a corpus spanning a century of (badly digitised) scientific congress reports, among other things.
Université libre de Bruxelles
- 2010-2012 **Master's degree in Information Science (MSc)**
Master's thesis on machine translation: French to Dutch translation of ads. Courses included Programming, Natural Language Processing, Architecture of Information Systems, Digital Publishing.
Université libre de Bruxelles
- 2008-2010 **Master's degree in Germanic philology (MA)**
Master's thesis on the impact of WWII on Harry Mulisch's oeuvre
Université libre de Bruxelles

Experience

Full-Time

- 2015 **Research Fellow**
Research Fellow grant awarded by the CENDARI project. During my time there I used a novel Linked Data based approach to extract and disambiguate place names out of a 3.1 million-page corpus.
Trinity Long Room Hub, Trinity College Dublin
- 2013- **Université libre de Bruxelles** Brussels, Belgium
PhD Candidate
My work is focused on the thesis, but I also give lectures and seminars to groups ranging from 10 to 600 students. I take part in the life of the institution as a representative of the scientific community at the Faculty Council and in other committees.
Detailed achievements:
 - Research fellow at the Trinity College Dublin (CENDARI.eu grant)
 - Co-founder of the Digital Humanities *Fonds National de la Recherche Scientifique* (FNRS) contact group
 - Presented at several scientific conferences and workshops abroad and in Belgium :
 - DHBenelux 2014 & 2015
 - Historical Network Research 2014: Ghent
 - Data Science Training 4 Librarians (DST4L) 2015: Copenhagen, coorganised by Foster Open Science, the Harvard-Smithsonian Center for Astrophysics, Harvard Library and Danmarks Tekniske Universitet
 - Topic Modelling in the Library: University of California Santa Barbara, 2016
 - JADH 2016 (Tokyo): Comparing Topic Model Stability across Language and Size
 - IEEE Big Data 2016 (Washington DC): Exploring archives with probabilistic models
 - Teaching Assistant

2012–2013	Université libre de Bruxelles <i>Teaching Assistant</i> Teaching master-level students the rudiments of IT – webhosting, server management, etc.	Brussels, Belgium
2013	GDF Suez <i>Young Knowledge Officer</i> Strategic Watch and Analysis	Brussels, Belgium
2009	Volontariat NGO <i>Volunteer</i> Drafting of contracts and translations	Puducherry, India

Interests

Professional: natural language processing, topic modelling, information extraction, semantic web technologies, named-entity recognition

Personal: robotics, running, web technologies

Projects

Professional: Forked a topic modelling tool to answer my research needs. It's available on GitHub: <https://github.com/ulbstic/topic-modeling-tool-FR>.

Personal: Co-founded in 2010 <http://sandesh-illam.org>, an evening school and kindergarten.

Publications

Articles in peer-reviewed journals

How hot is .brussels? Analysis of the uptake of the .brussels top-level domain name extension

Margot Waty, Seth Van Hooland, Simon Hengchen, Mathias Coeckelbergs, Max De Wilde
arXiv preprint arXiv:1606.04277 (2016). 2016

Semantic Enrichment of a Multilingual Archive with Linked Open Data

Max De Wilde, Simon Hengchen
Digital Humanities Quarterly – accepted for publication (2015). 2015

L'extraction d'entités nommées: une opportunité pour le secteur culturel?

Simon Hengchen, Seth Van Hooland, Ruben Verborgh, Max De Wilde
I2D – Information, données & documents (2015). 2015

Conferences, Proceedings & Workshops

Text Mining for User Query Analysis: A 5-Step Method for Cultural Heritage Institutions

Anne Chardonnens, Simon Hengchen
Proceedings of the 15th International Symposium on Information Science (ISI 2017); Berlin, Germany, 13th–15th March 20172017

Topic Modelling in the Library: A short Introduction

Simon Hengchen
2016, University of California Santa Barbara

Comparing Topic Model Stability across Language and Size

Simon Hengchen, Alexander O'Connor, Gary Munnely, Jennifer Edmond
Proceedings of the Japanese Association for Digital Humanities Conference 20162016

Exploring archives with probabilistic models: Topic Modelling for the valorisation of digitised archives of the European Commission

Simon Hengchen, Mathias Coeckelbergs, Seth Hooland, Ruben Verborgh, Thomas Steiner

First Workshop «Computational Archival Science: digital records in the age of big data», Washington 2016

NER as a gateway drug to the Linked Data cloud: Application of Named-Entity Recognition on cultural heritage metadata

Simon Hengchen, Max De Wilde, Seth Van Hooland

2014, The Hague

Named-Entity Recognition et Linked Data: quelle valeur ajoutée pour les archives?

Max De Wilde, Simon Hengchen

2014, Louvain la Neuve

Semantic Enrichment of a Multilingual Archive with Linked Open Data

Max De Wilde, Simon Hengchen

<http://dhubenelux.org/wp-content/uploads/2015/04/06.pdf>, Antwerp

Books

Méthodes et pratiques numériques en sciences humaines et sociales

Seth Van Hooland, Florence Gillet, Simon Hengchen, Max De Wilde

De Boeck Supérieur, 2016

You can also find the list of my publications on [Google Scholar](#).