
Personal Information

Gender	Male
Nationality	Ethiopian
Mobile	+4915775598722
Email	seidymam@gmail.com
Address	Hörgensweg 61, 22523, Hamburg, Germany
Home Page	https://seyyaw.github.io/
Current	Postdoctoral researcher at LT Group, Area of research: Annotation and Automation, Semantic technology, NLP for low-resource language, Social NLP (hate speech detection, fake news analysis), Big data
Ph.D. Tittle	Adaptive Approaches to Natural Language Processing in Annotation and Application, Universität Hamburg, Hamburg, Germany, July 3, 2019
From April 2014 – 2019	Doctoral researcher , LT lab, Universität Hamburg Integration of interactive machine learning into a semantic writing aid tool Big data analysis and visualization

Education

Oct 2010-Sep 2011	Advanced master in Human language technologies and interfaces (MPhil) , <i>Trento University, Italy</i> http://www.unitn.it/ Internship – Sentence based sentiment analysis from Telecom Italia texts
Oct 2007- Aug 2009	M.Sc. in Computer Science , <i>Addis Ababa University, http://www.aau.edu.et/</i> . Department of Computer Science, Graduation with Distinction Master Thesis: Question Answering for Amharic
Sep 199-Aug2004	B.Sc in Computer Science , Addis Ababa University, http://www.aau.edu.et/ Bachelor Project: Allocations of students to higher educations in Ethiopian universities, Ministry of Education, Ethiopia.

Work Experience

Sep 2012-Sep 2016	Scientific software engineer , Germany, <i>LT Lab, Universität Hamburg</i> . Main developer of the semantic writing aid tool (SEMSCH): A tool which supports paraphrasing or semantic expansion of words/phrases/sentences based on different lexical and semantic resource. In the future, an adaptive machine learning model will be integrated into the tool to interactively learn and adapt from user experience.
-------------------	---

Main developer of WebAnno (<http://webanno.github.io>). I am the main developer of WebAnno, a web-based and distributed annotation tool. The tool supports customizable annotations for any type annotation problem, curation of multiple annotations, correction of machine predicted annotations, and monitoring projects. Besides, WebAnno incorporates internal machine learning capability to speed up annotations.

Main developer of GermaNER (<http://github.com/tudarmstadt-lt/GermaNER>): an open and free NER for German, which can easily be adapted to other European Languages. The tool includes a German NER model. The tool can be easily configured to train on new language or domain (we already used it for Biomedical NER). The features to be used are easily configurable for selection and it easy to extend and include new feature extractors.

Backend and integration developer for network of the day (<http://www.tagesnetzwerk.de/>): Network of the day enables users/journalists to visualize the main important events, actions and actors based on news data and social network data collected in a daily fashion, which is a show case of big data applications.

Backend and integration developer new/s/leak (<http://newsleak.io/>): new/s/leak is a tool developed to help investigative journalists. It consists of latest technologies from NLP and InfoVis.

Sep 2011 - Sep 2012 **Semantic Web software developer**, Italy, *OKKAM srl*. <http://www.okkam.it/index.php/en/>
I joined OKKAM to work on a project about data integration, based on semantic technologies, as a java developer. I learn the semantic technologies at OKKAM which include RDF, SPARQL, and RDF stores. I also develop different extensions and reconciliation services to the then Google Refine (now Open Refine) RTL tool towards the OKKAM naming system. I also implemented the full stack of WebID, an authentication and authorization protocol based on PKI (Public Key Infrastructure) and RDF.

Aug 2009 - Oct 2007 **Assistant lecturer**, *Haramaya University, Ethiopia*. <http://www.haramaya.edu.et/>
Provide basic and advanced course lectures for undergraduate students, and coordinating the activities in Department of Information Technology

Dec 2010 - Oct 2011 **Lecturer at Adama University, Adama Science and Technology University**, <http://www.astu.edu.et/> Teach advanced programming (Java) and advanced networking courses, participate in the development and evaluation of faculty and department curriculum; participate in requirement analysis and evaluation of different campus-based software applications.

Projects

- DIVID-DJ: Data Extraction and Interactive Visualization of Unexplored Textual Datasets for Investigative Data-Driven Journalism
- SEMSCH: Semantic Methods for Computer-supported Writing Aids
- NoD: Network of the Day, Hochschulwettbewerb 2014
- CLARIN-D: Implementation of a web-based annotation platform for linguistic annotations (F-AG 7)

Software, Dataset and demo

Software

- GermaNER: German named entity recognition
- WebAnno: Web-based, distributive, and generic annotation tool
- G3G2 – Complex word identification for multiple languages
- Learning Paraphrasing for Multi-word Expressions
- Par4Sim: Paraphrasing for simplification tasks

Area of Experience

Programming Language	Java - proficient, Python- proficient, Scala - basic, HTML, JavaScript and CSS - basic, , SQL - proficient
Version Controls	Git and SVN
Deployment/ Build	Docker, Maven, Ant
Machine learning tools & NLP	Deep learning (PyTorch, Keras and Deeplearning4J), spaCy, NLTK, scikit-learn, WordNet and MultiWordNet, TextPro, TreeTager, Brat Annotation Tool, WebAnno
Database	MySQL, SPARQL, PostgreSQL, Jena
Frameworks & Tools	Anaconda, Apache Wicket, Elasticsearch , Spark , Hibernate, JPA, Serialization of XML or Json- SAX, DOM, Spring, Protégé, Lucene, Octave, Hadoop, Shell script
OS & Tools	Mac OS X , Linux (Debian), Windows ,

Language Skill

English	Advanced
Amharic	Native

Publications

You can also find my articles on [my Google Scholar profile](#).

Monographs

- **Yimam, S.M.** (2019): Adaptive Approaches to Natural Language Processing in Annotation and Application. Ph.D Thesis, University of Hamburg ([pdf](#))
- **Yimam, S.M.** (2009): TETEYEQ (ተጠየቅ): AMHARIC QUESTION ANSWERING SYSTEM FOR FACTOID QUESTIONS. Master Thesis, Addis Ababa University [pdf](#)

Book Chapters

- Biemann, C., Bontcheva, K., Eckart de Castilho, R., Gurevych, I., **Yimam, S.M.** (2017): Collaborative Web-based Tools for Multi-layer Text Annotation. In: N. Ide and J. Pustejovsky (Eds.): Handbook of Linguistic Annotation, Springer ([preprint pdf](#))

Journal Publications

- **Yimam, S.M.**, Biemann, C., Majnarić, L., Šabanović, Š., Holzinger, A. (2016): An adaptive annotation approach for biomedical entity and relation recognition. Brain Informatics, (online first), 10.1007/s40708-016-0036-4 ([pdf](#))

Conference Proceedings

- von Boguszewski, N., Moin, S., Bhowmick, A., **Yimam, S.M.**, Biemann, C. (2021): How Hateful are Movies? A Study and Prediction on Movie Subtitles. Proceedings of KONVENS, Düsseldorf, Germany ([pdf](#))
- Wiechmann, M., **Yimam S. M.**, Biemann, C. (2021): ActiveAnno: General-Purpose Document-Level Annotation Tool with Active Learning Integration. The 2021 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies - System Demonstrations, Mexico City, Mexico (online) ([pdf](#))
- Gooding, S., Kochmar, E., **Yimam S. M.**, Biemann, C. (2021): Word Complexity is in the Eye of the Beholder. The 2021 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT), Mexico City, Mexico. ([pdf](#))
- Mathew, B., Saha, P., **Yimam S. M.**, Biemann, C., Goyal, P., Mukherjee, A. (2021): HateXplain: A Benchmark Dataset for Explainable Hate Speech Detection. Proceedings of AACL-21, Virtual Conference. ([pdf](#))
- Haase, C., Anwar, S., **Yimam S. M.**, Friedrich, A., Biemann, C. (2021): SCoT: Sense Clustering over Time: a tool for the analysis of lexical change. The 2021 Conference of the European Chapter of the Association for Computational Linguistics - System Demonstrations. Kyiv, Ukraine (Online) ([pdf](#))

- **Yimam S. M.**, Alemayehu H. M., Ayele A. A. and Biemann C. (2020): Exploring Amharic Sentiment Analysis from Social Media Texts: Building Annotation Tools and Classification Models. The 28th International Conference on Computational Linguistics (COLING 2020), Barcelona, Spain ([pdf](#)) ([poster](#))
- **Yimam S. M.**, Venkatesh, G., Lee, J. Biemann, C. (2020): Automatic Compilation of Resources for Academic Writing and Evaluating with Informal Word Identification and Paraphrasing System, The International Conference on Language Resources and Evaluation (LREC 2020), Marseille, France. ([pdf](#))
- Wiedemann G., **Yimam S.M.**, and Biemann C. (2018) : A Multilingual Information Extraction Pipeline for Investigative Journalism. In Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018). Brussels, Belgium ([pdf](#))
- **Yimam S. M.**, Biemann C. (2018): Demonstrating Par4Sem - A Semantic Writing Aid with Adaptive Paraphrasing. In Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018). Brussels, Belgium ([pdf](#)).
- Wiedemann G., **Yimam S.M.**, and Biemann C. (2018) : New/s/leak 2.0 – Multilingual Information Extraction and Visualization for Investigative Journalism. In: Proceedings of the 10th International Conference on Social Informatics (SocInfo 2018). St.Petersburg, Russia ([pdf](#))
- **Yimam S.M.**, Biemann C. (2018): Par4Sim – Adaptive Paraphrasing for Text Simplification. In Proceedings of The 27th International Conference on Computational Linguistics (COLING 2018). Santa Fe, New-Mexico, USA ([pdf](#)).
- **Yimam S.M.**, Štajner S., Riedl M., Biemann C. (2017): CWIG3G2 - Complex Word Identification Task across Three Text Genres and Two User Groups. In Proceedings of The 8th International Joint Conference on Natural Language Processing (IJCNLP 2017). Taipei, Taiwan ([pdf](#))
- **Yimam S.M.**, Štajner S., Riedl Martin, Biemann C. (2017): Multilingual and Cross-Lingual Complex Word Identification. In Proceedings of The 2017 International Conference on Recent Advances in Natural Language Processing (RANLP). Varna, Bulgaria ([pdf](#))
- **Yimam, S.M.**, Ulrich, H., von Landesberger, T., Rosenbach, M., Regneri, M., Panchenko, A., Lehmann, F., Fahrer, U., Biemann, C. and Ballweg, K. (2016): new/s/leak – Information Extraction and Visualization for Investigative Data Journalists. ACL 2016 Demo Session, Berlin, Germany ([pdf](#))
- **Yimam, S.M.**, Biemann, C., Majnaric, L., Šabanović, Š., Holzinger, A. (2015): Interactive and Iterative Annotation for Biomedical Entity Recognition, International Conference on Brain Informatics and Health (BIH'15), London, UK ([pdf](#))
- Benikova, D., **Yimam, S.M.**, Biemann C. (2015). GermaNER: Free Open German Named Entity Recognition Tool. In: Proceedings of the GSCL 2015. Essen, Germany ([pdf](#))
- **Yimam, S.M.**, Eckart de Castilho, R., Gurevych, I., Biemann C. (2014): Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno. Proceedings of ACL 2014 System Demonstrations, Baltimore, MD, USA ([pdf](#))
- **Yimam, S.M.**, Gurevych, I., Eckart de Castilho, R., and Biemann C. (2013): WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations. Proceedings of ACL-2013, demo session, Sofia, Bulgaria ([pdf](#))

Workshop Proceedings

- David Ifeoluwa Adelani and Jade Abbott and Graham Neubig and Daniel D'souza and Julia Kreutzer and Constantine Lignos and Chester Palen-Michel and Happy Buzaaba and Shruti Rijhwani and Sebastian Ruder and Stephen Mayhew and Israel Abebe Azime and Shamsuddeen Muhammad and Chris Chinenye Emezue and Joyce Nakatumba-Nabende and Perez Ogayo and Anuoluwapo Aremu and Catherine Gitau and Derguene Mbaye and Jesujoba Alabi and **Seid Muhie Yimam** and Tajuddeen Gwadabe and Ignatius Ezeani and Rubungo Andre Niyongabo and Jonathan Mukiibi and Verrah Otiende and Iroro Orife and Davis David and Samba Ngom and Tosin Adewumi and Paul Rayson and Mofetoluwa Adeyemi and Gerald Muriuki and Emmanuel

Anebi and Chiamaka Chukwuneke and Nkiruka Odu and Eric Peter Wairagala and Samuel Oyerinde and Clemencia Siro and Tobius Saul Bateesa and Temilola Oloyede and Yvonne Wambui and Victor Akinode and Deborah Nabagereka and Maurice Katusiime and Ayodele Awokoya and Mouhamadane MBOUP and Dibora Gebreyohannes and Henok Tilaye and Kelechi Nwaike and Degaga Wolde and Abdoulaye Faye and Blessing Sibanda and Orevaoghene Ahia and Bonaventure F. P. Dossou and Kelechi Ogueji and Thierno Ibrahima DIOP and Abdoulaye Diallo and Adewale Akinfaderin and Tendai Marengereke and Salomey Osei (2021): MasakhaNER: Named Entity Recognition for African Languages. Proceedings of AfricaNLP Workshop, Strengthening African NLP. EACL 2021. ([pdf](#)).

- Wiedemann G, **Yimam, S.M.**, Biemann C. (2020): UHH-LT at SemEval-2020 Task 12: Fine-Tuning of Pre-Trained Transformer Networks for Offensive Language Detection. Proceedings of The 14th International Workshop on Semantic Evaluation (SemEval), Barcelona, Spain. ([pdf](#)) ([poster](#)).
- **Yimam, S.M.**, Ayele, A. A., Biemann C. (2019): Analysis of the Ethiopic Twitter Dataset for Abusive Speech in Amharic. In Proceedings of International Conference On Language Technologies For All: Enabling Linguistic Diversity And Multilingualism Worldwide (LT4ALL 2019). Paris, France p. 210-214 ([pdf](#)).
- **Yimam, S.M.**, Biemann, C., Malmasi, S., Paetzold, G.H., Speica, L., Štajner, S., Tack, A., Zampieri, M., (2018): A Report on the Complex Word Identification Shared Task 2018. Proceedings of the 13th Workshop on Innovative Use of NLP for Building Educational Applications, New Orleans, LA, USA ([pdf](#))
- **Yimam S.M.**, Remus S., Panchenko A., Holzinger A., Biemann C. (2017): Entity-Centric Information Access with the Human-in-the-Loop for the Biomedical Domains. Biomedical NLP Workshop associated with RANLP 2017. Varna, Bulgaria ([pdf](#))
- Müller, M., Ballweg, K. von Landesberger, T., **Yimam, S.M.**, Fahrer, U., Biemann, C., Rosenbach, M., Regneri, M., Ulrich, H. (2017). Guidance for Multi-Type Entity Graphs from Text Collections. EuroVis Workshop on Visual Analytics 2017, Barcelona, Spain ([pdf](#))
- Nandi, T., Biemann, C., **Yimam, SM.**, Gupta, Deepak., Kohail, S., Ekbal, A., Bhattacharyya, Pushpak. (2017): IT-UHH at SemEval-2017 Task 3: Exploring Multiple Features for Community Question Answering and Implicit Dialogue Identification, In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval 2017), Vancouver, Canada. ([pdf](#))
- Eckart de Castilho, R. Mújdricza-Maydt, E., **Yimam, S.M.**, Hartmann, S., Gurevych, I., Frank, A. and Biemann, C. (2016): A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures. Proceedings of the COLING workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH), Osaka, Japan ([pdf](#))
- Ballweg K., Zouhar F., Wilhelmi-Dworski P., von Landesberger T., Fahrer U., Panchenko A., **Yimam S.M.** Biemann C., Regneri M., Ulrich H. (2016) new/s/leak – A Tool for Visual Exploration of Large Text Document Collections in the Journalistic Domain, Baltimore, MD, USA, ([pdf](#))
- **Yimam, S.M.**, Martínez Alonso, H., Riedl M. and Biemann, C. (2016): Learning Paraphrasing for Multiword Expressions. The 12th Workshop on Multiword Expressions (MWE 2016), co-located with ACL 2016, Berlin, Germany ([pdf](#))
- **Yimam, S.M.** (2015): Narrowing the Loop: Integration of Resources and Linguistic Dataset Development with Interactive Machine Learning. NAACL 2015 Student Research Workshop, p. 88–95, Denver, Colorado ([pdf](#))
- Eckart de Castilho, R., Biemann, C., Gurevych, I., **Yimam, S.M.** (2014): WebAnno: a flexible, web-based annotation tool for CLARIN. CLARIN Annual Conference 2014, Soesterberg, The Netherlands ([pdf](#))
- Benikova, D., Fahrer, U., Gabriel, A., Kaufmann, M., **Yimam, S.M.**, von Landesberger, T., Biemann, C. (2014): Network of the Day: Aggregating and Visualizing Entity Networks from Online Sources. KONVENS 2014 Workshop proceedings: NLP4CMC, pp. 48-52, Hildesheim, Germany ([pdf](#))
- **Yimam, S.M.**, Libse, M. (2009): TETEYEQ: Amharic Question Answering For Factoid Questions, SEPLN09. SALT MIL workshop - Information Retrieval and Information Extraction for Less Resourced Languages (IE-IR-LRL), p. 17-25 ([pdf](#))

Posters

- [Yimam, S.M](#), Biemann C.,. (2019): Current Status, Issues, and Future Directions for Ethiopian Natural Language Processing (NLP) Research. International Conference Language Technologies for All (LT4All), Paris, France ([pdf](#))