

01:10-0.1  
02:10-231

# UNLOCK YOUR POTENTIAL WITH OPEN SOURCE

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231

01:10-0.1  
02:10-231



## GITHUB

INTRODUCTION TO OPEN SOURCE

Introduction to open source software development and the role of GitHub in the ecosystem.

CODING, DECODE, MODEL

Exploring the process of coding, decoding, and modeling in the context of open source.

## HUGGING FACE

AI AND MODELS FACE

Introduction to AI models and the role of Hugging Face in the field.



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG



# Unlocking Potential Through Open Sourcing: Open Code, Open Data, and Open Opportunity

Dr. Seid Muhie Yimam  
University of Hamburg  
Hub of Computing and Data Science



# Concept of Open Source

- **Shared Source Code:** Open source software (OSS) is distributed with its source code, allowing users to **use**, **modify**, and **distribute** it under original rights, fostering collaboration and customization.
- **Historical Origin:** The open source movement began with an ideological drive by **Richard Stallman in 1983**, emphasizing software **freedom** and leading to the development of the **GNU Public License** and the formation of the **Open Source Initiative** in 1998.



Image by storyset on Freepik

[https://cdn.media.amplience.net/i/epamma\\_rk marketplace/open\\_source\\_header?maxW=1200&qI=80&fmt=jpg&bg=rgb\(255,255,255\)&v=1720241841502](https://cdn.media.amplience.net/i/epamma_rk marketplace/open_source_header?maxW=1200&qI=80&fmt=jpg&bg=rgb(255,255,255)&v=1720241841502)



# Free Software vs Open-Source Software

- **Freedom vs. Cost:** "Free software" emphasizes user freedom to modify and share, not necessarily being free of charge.
- **Misinterpretation:** "Free software" often misunderstood as "no cost," leading to the adoption of "open-source" to clarify intent.
- **Terminology:** Terms like FOSS (Free and Open Source Software) and FLOSS (Free/Libre Open Source Software) aim to capture both concepts. -
- **Code Availability:** Both models emphasize access to **source code** for **modification and distribution** rights.

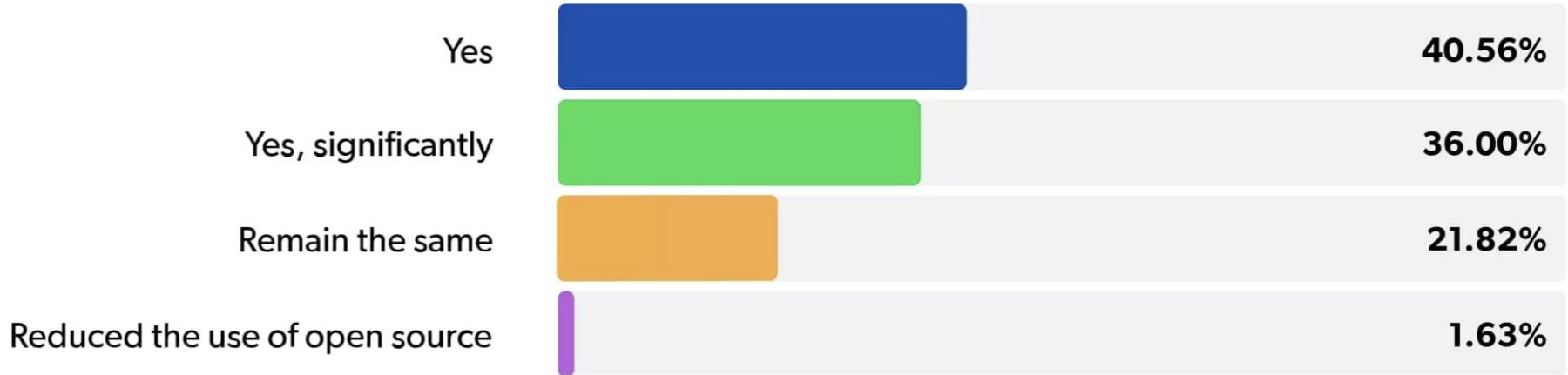


"software should be free – as in **speech**, not beer".

- Report on the state of open source in 2022 made in collaboration between OpenLogic by Perforce and the Open Source Initiative (OSI)



**Has Your Organization Increased the Use of Open Source Software Over the Last Year?**



# Pros of Open Source Software

- **Cost:** Open source software is available without charge.
- **Flexibility:** Allows developers to modify code to suit their needs.
- **Stability:** Publicly distributed source code ensures long-term reliability.
- **Innovation:** Encourages enhancements and new developments from existing code.
- **Community Support:** A collaborative community continuously improves the software.
- **Educational Value:** Offers valuable learning opportunities for new programmers.

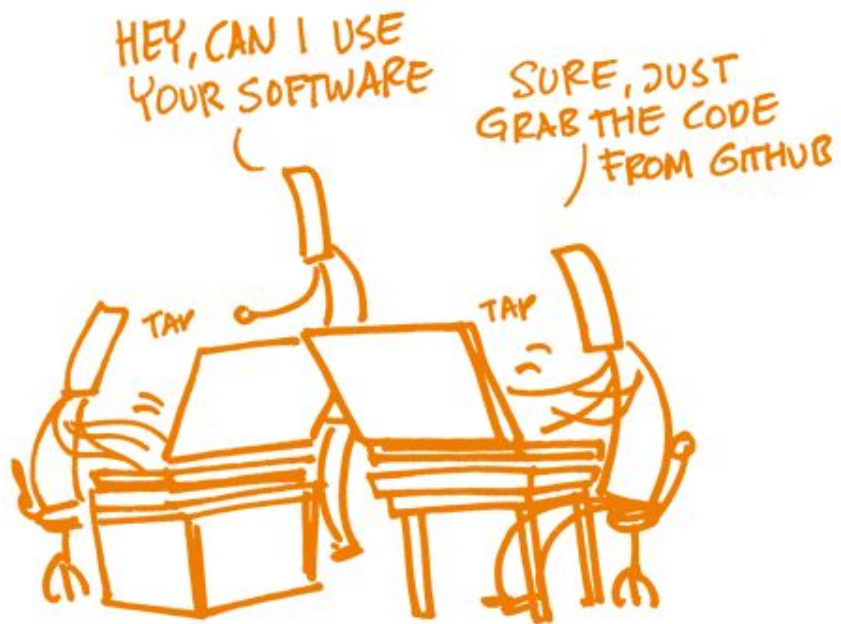


Security

OPEN SOURCE

vs

CLOSED SOURCE



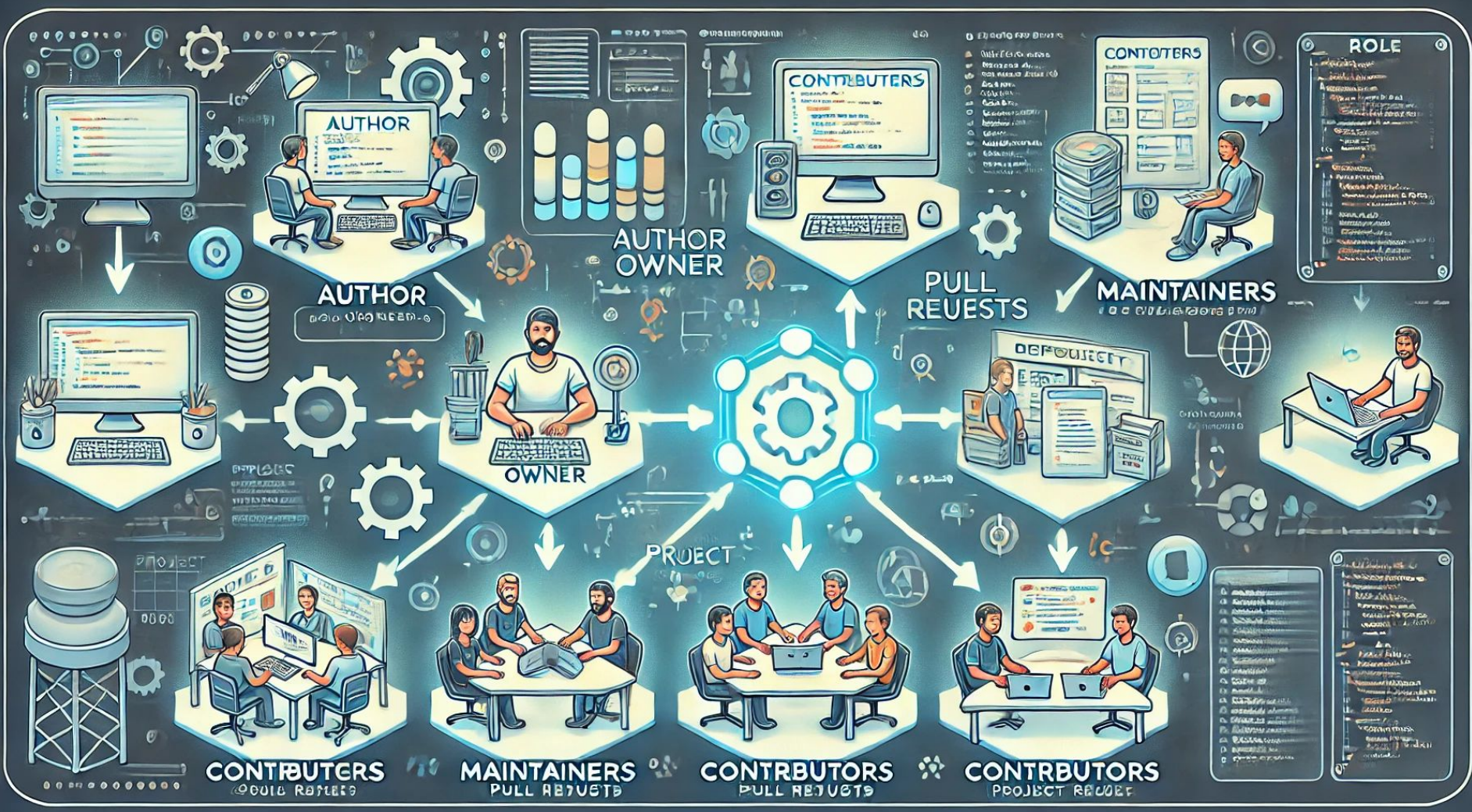
# Benefits of Open Source Contributions/Participation

- **Practical Experience:** Work on real-world projects enhances programming skills and knowledge.
- **Skill Development:** Gain new skills and deepen existing ones through hands-on practice.
- **Mentorship:** Access guidance and support through open-source mentorship programs.
- **Feedback Loop:** Receive immediate feedback to improve development techniques.
- **Networking:** Build connections with a global community of like-minded developers.
- **Field Insights:** Discover more about your interests and related fields through collaboration.
- **Career Opportunities:** Contributions can open doors to job offers and serve as an alternative to traditional internships.



# Formal Roles in Open Source Projects

- **Author:** Initiates the project by creating the original codebase.
- **Owner:** Holds administrative control over the project's repository and settings.
- **Maintainers:** Guide the project's vision, manage contributions, and ensure organizational aspects are aligned; can also be authors or owners.
- **Contributors:** Individuals who provide improvements, features, or bug fixes to the project.
- **Community Members:** Engage with the project by using it, participating in discussions, and suggesting enhancements.



**AUTHOR**  
@000 UAG R0000-0

**AUTHOR OWNER**

**PULL REUETS**

**MAINTAINERS**  
I a c 000000000000000000

**OWNER**

**PROJECT**

**CONTRIBUTORS**  
@0000 R00000

**MAINTAINERS**  
PULL AB0UETS

**CONTRIBUTORS**  
PULL AB0UETS

**CONTRIBUTORS**  
PROJECT R0000F

**ROLE**

- 1. ...
- 2. ...
- 3. ...
- 4. ...
- 5. ...
- 6. ...
- 7. ...
- 8. ...
- 9. ...
- 10. ...
- 11. ...
- 12. ...
- 13. ...
- 14. ...
- 15. ...
- 16. ...
- 17. ...
- 18. ...
- 19. ...
- 20. ...

```
1. ...
2. ...
3. ...
4. ...
5. ...
6. ...
7. ...
8. ...
9. ...
10. ...
11. ...
12. ...
13. ...
14. ...
15. ...
16. ...
17. ...
18. ...
19. ...
20. ...
```

# Platforms and Tools for Open Source Software

- **LibreOffice**: Open Source office suite alternative to Microsoft Office.
- **Zotero**: Open Source [citation manager](#) compatible with LibreOffice and other tools.
- **GitHub**: Popular repository for managing, sharing, and collaborating on open-source projects; alternatives include [GitLab](#) and [BitBucket](#).
- **Hugging Face**: Platform for developing and sharing state-of-the-art machine learning models and datasets.
- **LaTeX/Overleaf**: LaTeX is a high-quality **typesetting system** widely used for scientific and technical documents; **Overleaf** is an **online LaTeX editor** that simplifies collaborative writing and compiling of LaTeX documents.



# Why Open Source Software is Preferred in Research

- **Transparency in methods**: Algorithms in analysis software are open, facilitating rigorous **peer review**.
- **Collaborative advantages**: Supports multi-institution and distributed research efforts without proprietary barriers.
- **Cost-effectiveness**: Relies on **volunteer contributions**, reducing budget dependence for scientific research.
- **Flexibility and adaptation**: Allows custom modification and integration, essential for evolving research needs.



# Data and Model Sharing Frameworks

- **TensorFlow Hub**: Repository for sharing pre-trained TensorFlow models.
- **PyTorch Hub**: Discover and reuse pre-trained PyTorch models.
- **Hugging Face**: Hosts models and datasets for NLP, vision, and more.
- **ONNX**: Interoperable model format for different ML frameworks.
- **MLflow Model Registry**: Manages model lifecycle and versioning.
- **ModelDB**: Tracks and logs machine learning model experiments.
- **Zenodo**: Open repository for datasets, software, and models.
- **Kaggle Datasets**: Platform to discover and share datasets and code.
- **DataONE**: Federated repository for sharing earth observation data.
- **Google Dataset Search**: Tool for finding datasets across the web.



Hu



Dataset Se

Search for Datasets



# Own Experiences on Open Source

# Own Experience - WebAnno - An Open Source Annotation Tool

The screenshot shows a GitHub repository page for 'webanno'. The repository structure is as follows:

Folder Name	Commit Message	Time Ago
webanno-brat	[maven-release-plugin] prepare for n...	3 years ago
webanno-build	[maven-release-plugin] prepare for n...	3 years ago
webanno-constraints	[maven-release-plugin] prepare for n...	3 years ago
webanno-curation	#1895 - Merging relat	3 years ago
webanno-diag	#1893 - Deleting a sp	3 years ago
webanno-doc	Merge branch '3.6.x'	3 years ago
webanno-docker	No issue. Fix versions.	3 years ago
webanno-export	[maven-release-plugin] prepare for n...	3 years ago
webanno-io-conll	[maven-release-plugin] prepare for n...	3 years ago
webanno-io-json	[maven-release-plugin] prepare for n...	3 years ago
webanno-io-tcf	[maven-release-plugin] prepare for n...	3 years ago
webanno-io-tei	[maven-release-plugin] prepare for n...	3 years ago
webanno-io-text	[maven-release-plugin] prepare for n...	3 years ago
webanno-io-tsv	Merge branch '3.6.x'	3 years ago

Repository details: Apache-2.0 license, Activity, Properties, etc.

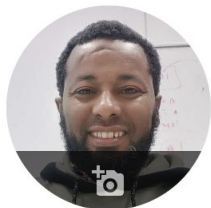
Version: 3.6.11 (Latest) 2021

Contributors: 31

**seyyaw** Seid Muhie Yimam · he/him  
I am Seid, a Technical Lead at the Hub of Computing and Data Science (HCDS) and a Research Associate at LT Group, Universität Hamburg, Germany.  
📍 Germany  
🕒 04:36 (UTC +01:00)  
🔗 Committed to this repository  
👤 Member of [webanno/webanno-admin](#) and [webanno/...](#)  
👤 Member of [WebAnno](#), and 20 more

+ 17 contributors

# WebAnno - Web-based distributive annotation tool



Seid Muhie Yimam (ሰይፍ ሙህ ደማም) ✎

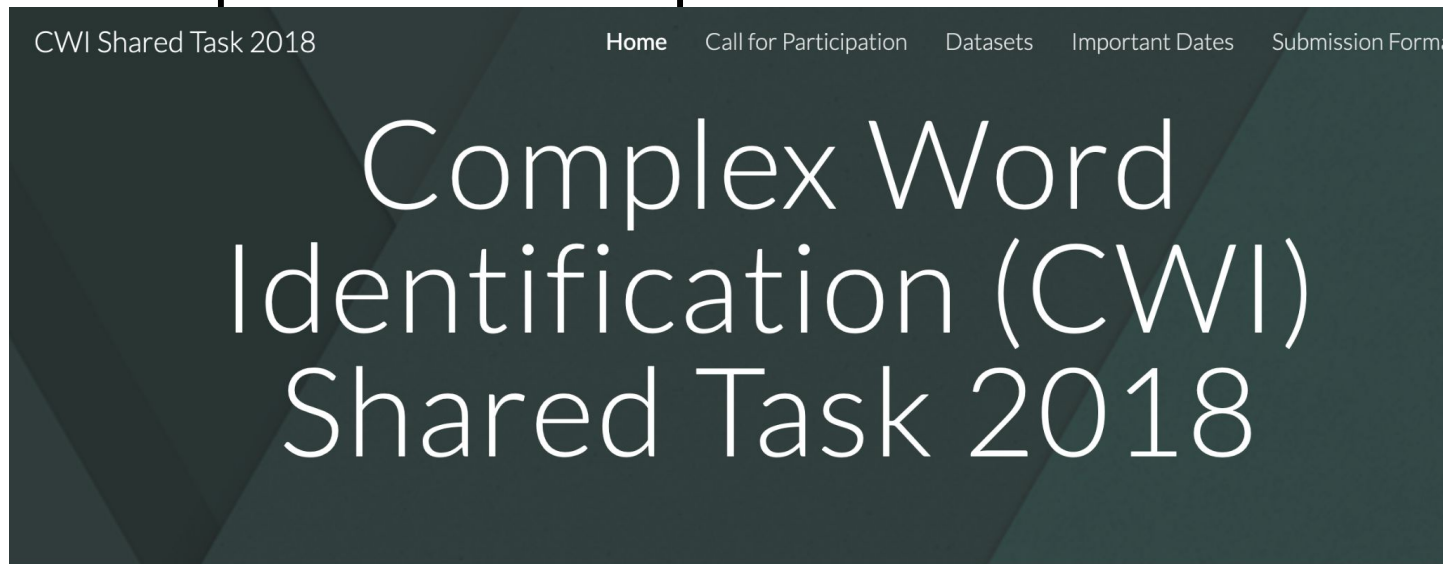
FOLLOW

Technical lead at HCDS and research associate at LT Group, [University of Hamburg](#)  
Verified email at uni-hamburg.de - [Homepage](#)

[Data Science](#) [Digitalization](#) [Interactive and adaptive ma...](#) [Social media and NLP](#)  
[Less resourced language](#)

<input type="checkbox"/>	TITLE		CITED BY	YEAR
<input type="checkbox"/>	<a href="#">Hatexplain: A benchmark dataset for explainable hate speech detection</a>	B Mathew, P Saha, SM Yimam, C Biemann, P Goyal, A Mukherjee Proceedings of the AAAI conference on artificial intelligence 35 (17), 14867 ...	590	2021
<input type="checkbox"/>	<a href="#">Webanno: A flexible, web-based and visually supported system for distributed annotations</a>	SM Yimam, I Gurevych, RE De Castilho, C Biemann Proceedings of the 51st annual meeting of the Association for Computational ...	305	2013
<input type="checkbox"/>	<a href="#">A web-based tool for the integrated annotation of semantic and syntactic structures</a>	RE De Castilho, E Mújdricza-Maydt, SM Yimam, S Hartmann, I Gurevych, ... Proceedings of the workshop on language technology resources and tools for ...	248	2016
<input type="checkbox"/>	<a href="#">A report on the complex word identification shared task 2018</a>	SM Yimam, C Biemann, S Malmasi, GH Paetzold, L Specia, S Štajner, ... arXiv preprint arXiv:1804.09132	201	2018
<input type="checkbox"/>	<a href="#">Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno</a>	CB Seid Muhie Yimam, Richard Eckart de Castilho, Iryna Gurevych	104	2014

# Own Experience - Complex Word Identification Dataset



Text simplification systems aim to facilitate reading comprehension to different target readerships such as foreign language learners, native speakers with low literacy levels or various kinds of reading impairments. Identifying which the words are considered difficult for a given target population is an important step for building better performing lexical simplification systems. This step is known as complex word identification (CWI).

Following the success of the first [CWI shared task at SemEval 2016](#), we are organizing the Second CWI Shared Task co-located with the [BEA Workshop 2018](#) at [NAACL-HLT](#) in New Orleans, USA.

The Second CWI Shared Task features a multilingual dataset and participants can choose to participate in one or more of the following tracks:





## Seid Muhie Yimam (ሰይጅ ሙህ ደማም)

 FOLLOW

Technical lead at HCDS and research associate at LT Group, [University of Hamburg](#)  
Verified email at uni-hamburg.de - [Homepage](#)

Data Science Digitalization Interactive and adaptive ma... Social media and NLP  
Less resourced language

<input type="checkbox"/>	TITLE  	CITED BY	YEAR
<input type="checkbox"/>	<a href="#">Hatexplain: A benchmark dataset for explainable hate speech detection</a> B Mathew, P Saha, SM Yimam, C Biemann, P Goyal, A Mukherjee Proceedings of the AAAI conference on artificial intelligence 35 (17), 14867 ...	590	2021
<input type="checkbox"/>	<a href="#">Webanno: A flexible, web-based and visually supported system for distributed annotations</a> SM Yimam, I Gurevych, RE De Castilho, C Biemann Proceedings of the 51st annual meeting of the Association for Computational ...	305	2013
<input type="checkbox"/>	<a href="#">A web-based tool for the integrated annotation of semantic and syntactic structures</a> RE De Castilho, E Mújdricza-Maydt, SM Yimam, S Hartmann, I Gurevych, ... Proceedings of the workshop on language technology resources and tools for ...	248	2016
<input type="checkbox"/>	<a href="#">A report on the complex word identification shared task 2018</a> SM Yimam, C Biemann, S Malmasi, GH Paetzold, L Specia, S Stajner, ... arXiv preprint arXiv:1804.09132	201	2018
<input type="checkbox"/>	<a href="#">Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno</a> CB Seid Muhie Yimam, Richard Eckart de Castilho, Iryna Gurevych Proceedings of 52nd Annual Meeting of the Association for Computational ...	104	2014



# HateXplain - Explainable hate speech dataset





Seid Muhie Yimam (ሰይጅ ሙሄ ይማም) ✎

 FOLLOW

Technical lead at HCDS and research associate at LT Group, [University of Hamburg](#)  
Verified email at uni-hamburg.de - [Homepage](#)

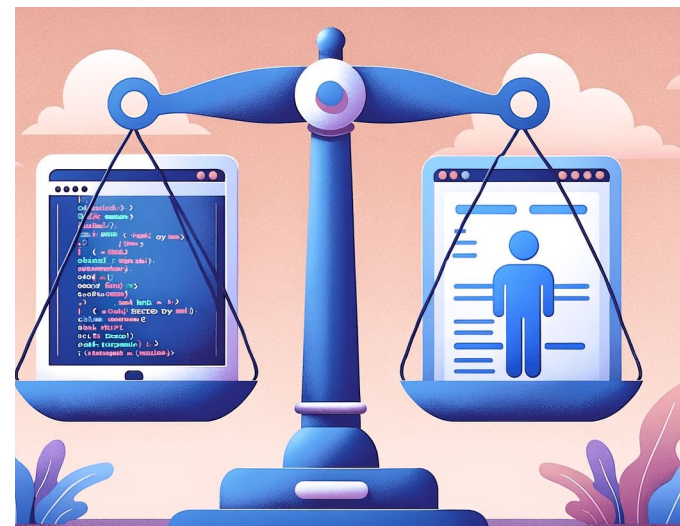
[Data Science](#) [Digitalization](#) [Interactive and adaptive ma...](#) [Social media and NLP](#)  
[Less resourced language](#)

<input type="checkbox"/>	TITLE  	CITED BY	YEAR
<input type="checkbox"/>	<a href="#">HateXplain: A benchmark dataset for explainable hate speech detection</a> B Mathew, P Saha, SM Yimam, C Biemann, P Goyal, A Mukherjee Proceedings of the AAAI conference on artificial intelligence 35 (17), 14867 ...	590	2021
<input type="checkbox"/>	<a href="#">Webanno: A flexible, web-based and visually supported system for distributed annotations</a> SM Yimam, I Gurevych, RE De Castilho, C Biemann Proceedings of the 51st annual meeting of the Association for Computational ...	305	2013
<input type="checkbox"/>	<a href="#">A web-based tool for the integrated annotation of semantic and syntactic structures</a> RE De Castilho, E Mújdricza-Maydt, SM Yimam, S Hartmann, I Gurevych, ... Proceedings of the workshop on language technology resources and tools for ...	248	2016
<input type="checkbox"/>	<a href="#">A report on the complex word identification shared task 2018</a> SM Yimam, C Biemann, S Malmasi, GH Paetzold, L Specia, S Štajner, ... arXiv preprint arXiv:1804.09132	201	2018
<input type="checkbox"/>	<a href="#">Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno</a> CB Seid Muhie Yimam, Richard Eckart de Castilho, Iryna Gurevych	104	2014

# Ethics and License

# Open Source Ethics

- **Respect:** Awareness and adherence to the roles, expectations, and **intellectual property** involved in open source.
- **Moral:** Understanding and acceptance of the central motivations and ethical considerations in open source interactions.
- **Compliance:** Ensuring adherence to **legal and licensing requirements** in use, redistribution, and contribution to open source software.



<https://blog.openreplay.com/images/ethical-considerations-in-software-development/images/hero.png>

# What is a software license?

- **Defines Usage Rights and Obligations:** A software license specifies **how software can be used, modified, and distributed**, including any obligations users must follow.
- **Governs Third-Party Code:** It covers reused code components, such as libraries and frameworks, ensuring compliance and preventing legal issues.
- **Ensures Legal Protection:** Proper licensing helps companies avoid **legal risks**, such as lawsuits or injunctions, by clearly outlining permissions and restrictions.



<https://hypertecsp.com/wp-content/uploads/software-licensing-640-kb.jpg>

<https://www.blackduck.com/blog/5-types-of-software>



How do you know if you can use a software?

If you open source your code that include proprietary library, how do you deal?

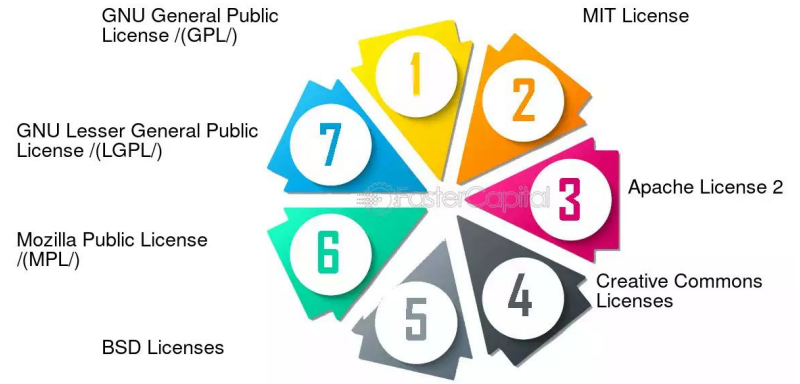
# Different Types of Software Licenses

- **Permissive:** Minimal restrictions, allowing broad use and modification, e.g., **MIT**, **Apache**.
- **Weak Copyleft:** Allows linkage with fewer obligations, e.g., **LGPL**.
- **Copyleft:** Requires derivative works to be open-sourced under the same license, e.g., **GPL**.
- **Commercial/Proprietary:** **Restricts use** and modification, typically for **profit**.
- **Dual Licensing:** Offers multiple license options, balancing open-source and commercial use.
- **Public Domain:** Free of copyright, allowing **unrestricted use**.
- **Unlicensed:** Requires **explicit permission**, as no license implies copyright protection.



Which license do you think is most popular?  
Which one is less attractive for research?

<https://fastercapital.com/i/Open-source-software--How-to-use-open-source-software-and-benefit-from-its-advantages--Exploring-Popular-Open-Source-Licenses.webp>





# Open Source Participation

# Recognition and Contribution in Open Source Projects

- Build a **public portfolio** showcasing your skills and contributions.
- **Gain visibility** and **recognition** from peers and industry leaders.
- Contribute to **widely-used projects**, impacting global users.
- Receive **endorsements** and **acknowledgments** from project maintainers.
- Establish **credibility** and a **reputation** within the tech community.

# Leveraging Open Source for Professional Growth

- Use contributions as a portfolio to attract **potential employers**.
- Demonstrate **initiative and technical expertise** through public work.
- **Network** with industry professionals and open-source communities.
- **Secure job** opportunities and freelance contracts through visibility.
- Enhance **problem-solving** and collaboration skills in a professional setting.

# Important skills

- **Programming:** Python and R for analysis.
- **Data Manipulation:** Use Pandas and NumPy.
- **Machine Learning:** Apply TensorFlow/PyTorch/SciKitLearn.
- **Visualization:** Create charts with Matplotlib/Plotly.
- **Big Data:** Work with Spark/Hadoop.
- **NLP:** Leverage Hugging Face tools.
- **Version Control:** Use Git and GitHub.
- **Command Line:** Navigate using CLI tools.
- **Data management:** SQL, NoSQL, JSON, CSV.
- **Cloud Computing:** Utilize AWS or Google Cloud.
- **Reproducibility:** Ensure with Jupyter and Docker.
- **Communication:** Present research effectively.

# Open source framework for your publicity

- **Contribute:** Engage with open-source projects on platforms like Hugging Face.
- **Create Projects:** Develop and share open-source projects using frameworks; host documentation on GitHub Pages, Host your **academic page** for free.
- **Write Content:** Publish tutorials or articles on frameworks used.
- **Social Media:** Share work on **X** and **LinkedIn** with relevant hashtags.
- **Speak:** Present at conferences, **Meetups**, and in webinars.
- **Collaborate:** Build a community around your projects on platforms like Hugging Face.
- **Showcase:** Highlight success stories and case studies.
- **Train:** Offer workshops or webinars on using frameworks.
- **Network:** Engage with framework maintainers and contributors.
- **License and Guide:** Clearly define project licenses and contribution guidelines.



# Opportunities, Venues, Calls, communities

- Deep learning Indaba
- IndabaX
- WiNLP
- Black in AI
- EthioNLP
- Masakhane
- ACL/EMNLP/NAACL



## Join the WU-KIOT Public repositories

- GitHub:

<https://github.com/wu-kiot>

- Hugging Face:

<https://huggingface.co/wu-kiot>



# Thank you!

Dr. Seid Muhie Yimam

University of Hamburg, HCDS



[seid.muhie.yimam@uni-hamburg.de](mailto:seid.muhie.yimam@uni-hamburg.de)



@seyyaw



@seyyaw



@seyaw



@seyyaw