On the Prevalence of Leichte Sprache on the German Web

Hadi Asghari, Freya Hewett, Theresa Züger
firstname.lastname@hiig.de
Alexander von Humboldt Institute for Internet and Society
Berlin, Germany

ABSTRACT

Web accessibility guidelines call for website content to be 'understandable'. In the German public sector, this principle has been interpreted as a specific set of writing rules known as 'Leichte Sprache' (LS). In this paper, we set out to investigate the prevalence of LS on the German web, using both web measurements and qualitative methods. We find that while many of the prerequisites for the creation of content in LS are now in place, such as accessibility monitoring authorities or procedures to translate content into LS, the vast majority of public sector websites are still not accessible in this regard. Based on these findings, we offer four technical and policy recommendations to move towards a more inclusive web.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in accessibility**; • **Computing methodologies** → *Natural language processing*.

KEYWORDS

web accessibility directive, leichte sprache, easy language

ACM Reference Format:
Hadi Asghari, Freya Hewett, Theresa Züger. 2023. On the Prevalence of Leichte Sprache on the German Web. In *15th ACM Web Science Conference 2023 (WebSci '23), April 30–May 01, 2023, Evanston, TX, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3578503.3583599

1 INTRODUCTION

The inventor of the web, Tim Berners-Lee, stated that "the power of the Web is in its universality. Access by everyone, regardless of disability is an essential aspect" [10]. This statement, from the 1990s, rings even more true in 2023, as participation in society depends on being able to access the web for information, services, and connections. For this reason, the European Union has enacted the Web Accessibility Directive [5], which requires public sector websites to be accessible to persons with disabilities. The Directive describes four principles for accessibility, one of which is 'understandability'. In this paper, we look at the implementation of this principle in Germany, as it has been interpreted in a stricter manner than most countries, with formal rules prescribing '*Leichte Sprache*' (henceforth: LS) or easy to understand German.

WebSci '23, April 30–May 01, 2023, Evanston, TX, USA

ACM ISBN 979-8-4007-0089-7/23/04.
https://doi.org/10.1145/3578503.3583599

While many aspects of accessibility are more broadly acknowledged, for instance as set out in the W3C's WCAG guidelines [15], and have been studied, LS remains a German (language) peculiarity. And while there are benefits to having clear rules as to what defines understandable language, legally not all of a website's content needs to be offered in LS, and just a few pages that convey the gist of the website are enough, as long as a contact option is given for users in case they need more information.

This brings up the empirical question: what percentage of the web content is available in LS? This is an important question, as it impacts how accessible the websites are in the end. If the percentage is low, this might be due to the difficulty and costs of creating LS pages—a task which might benefit from some NLP automation. But it could also be due to the fact that adults with cognitive disabilities, who are the primary target of LS, do not use it as much—which might itself be due to the limited number of LS pages.

We thus set out to measure the prevalence of LS across the German web. We tried to answer this question initially using the 'Common Crawl' dataset—one of the best open archives of the web—and built classifiers to detect such pages. But we found that unfortunately, the Common Crawl has not archived many LS pages. In the process we did, however, find a number of other clusters of easier German language that are typically used outside of the bureaucratic context. We describe these findings in more detail in section 4.

Our next attempt was to crawl German public sector websites (which the Directive applies to) ourselves, but also discovered that a comprehensive list of such websites does not exist. We thus contacted (by email and phone) the offices responsible for the monitoring of the accessibility Directive in 16 German states, plus the federal government, to ask for the list of public sector sites. These contacts allowed us to also ask a series of contextual questions, for instance about how much the LS pages are used, any feedback given by the community, and the costs involved in creating them. The results were quite interesting, with some states responding in a very transparent manner, while others preferred not to share such lists.

In sum, the mix of web measurements and qualitative information allowed us to reflect on defining understandability in terms of formal LS rules, and more broadly, the EU Web Accessibility Directive's success, with some suggestions for improvement.

The contributions of this paper are as follows:

(1) An overview of the use of LS in the German public sector (along with a legal summary of the rules);
(2) Experimental results on other sources of simple language within the German web;
(3) Automated tools to crawl and classify pages in simple German;

(4) Accessibility recommendations for both the technical community & public administration.

2 BACKGROUND & RELATED WORK

2.1 Leichte Sprache

Leichte Sprache (or 'Easy Language') differs from other forms of simplified language in that it has a clearly defined set of rules and a specific target group (adults with cognitive disabilities) [12]. This set of rules includes aspects such as ensuring that sentences are kept as short as possible, with no subordinate clauses, and the use of words that are common and easy to understand, with words that are more difficult but necessary in the context being explained. LS, the German branch of easy-to-read language, has been around since the early 2000s [7], but has gained more traction in recent years due to various legal requirements.

Our work builds on a study by Bock [3] in which various problematic topics surrounding LS are discussed and several examples of the availability of LS on the internet are evaluated. The current study aims to extend this exemplary analysis and make it more comprehensive. One factor that Bock discusses is the fact that there is a wide discrepancy regarding the scope of the information that is available in LS; for a few websites there are multiple pages in LS and the content seems to be tailored to the target group (i.e. specific information on taxes, for example), whereas other websites simply have a page which explains how to navigate the website in LS. Without quantifying the amount available, Bock lists the following categories as having information in LS online: politics, charities, media, and culture.

In the NLP domain, there have been various attempts to leverage data available on the web for tasks such as automatic simplification or readability assessment. We outline these datasets as they presumably reflect what kind of data is available online. In prior work, [9] built a corpus in 2013 that consisted of five domains of organisations that support people with disabilities; [2] extended this corpus to include websites of "governments, specialised institutions, translation agencies, and non-profit organisations" covering topics such as politics, health and culture; and [14] used eight different online sources from news portals, organisations for people with disabilities, and governmental websites to create a dataset of parallel data.

2.2 The Legal Setting

The European Parliament and Council enacted Directive 2016/2102 on the "accessibility of the websites and mobile applications of public sector bodies" [5], as despite having joined the Convention on the Rights of Persons with Disabilities in 2010, many European websites remained inaccessible to those who needed assistive technology and other adaptations [10]. The Directive had two important purposes: first, to promote and facilitate the development of accessible digital technology; second, to make it unnecessary for individuals to pursue legal action to ensure access, by requiring governments to setup authorities to self-monitor the accessibility of their services [10].

Regarding the technical aspects of web accessibility, the Directive refers to the European standard EN 301549[1], which is based on the W3C's Web Content Accessibility Guidelines (WCAG) version 2.1 [15]. These guidelines are organized under four principles [8], and state that an accessible website needs to be perceivable, operable, understandable, and robust.

EU Directives are transposed into national laws. The accessibility Directive has been transposed into one to three laws in most EU member states, with Germany being an extreme outlier, having transposed it into 54 laws [8]. This high number is due to Germany's unique federal system. At the federal level, the most important laws are the "Disability Equality Act" (in German: *Behindertengleichstellungsgesetz*[2] or BGG) and the "Accessible IT Ordinance" (in German: *Barrierefreie-Informationstechnik-Verordnung*[3] or BITV 2.0). The Disability Equality Act, which was originally passed in April 2002 and has been updated many times since, sets out a number of rules, including on "accessible IT of public bodies" (art. 12) and communicating in "comprehensible and easy language" (art. 11). The BITV ordinance, last updated in May 2019, offers further details, including "guidelines for the provision of information in Easy Language on the internet" (appendix 2a), which describes the rules of *Leichte Sprache*. The BITV specifies that at a minimum, the homepage of a public website needs to link to one or more pages that explain the website's content, navigation, and ways to obtain additional information in LS (and sign language). The rest of the website can optionally be in LS as well, although in practice this is rarely the case.

We would like to remind the reader that defining 'understandable' language as '*Leichte Sprache*' is a unique interpretation of the principles and guidelines set out by the web accessibility Directive and the WCAG. Even in Germany, when we move to a state level, there appears to be disagreement about this interpretation. Ten states, or around two thirds, have accessibility laws that refer to the BITV 2.0 ordinance, which means they mandate their state and municipal websites to support LS. Six states (Brandenburg, Hesse, Lower Saxony, Mecklenburg-Western Pomerania, North Rhine-Westphalia, and Saxony-Anhalt), on the other hand, point directly to the European standard (EN 301549), thus bypassing the specific rules regarding LS.

As a final point, the web accessibility Directive mandates regular monitoring of public sector websites and apps by the EU member states, with public reporting of the results[4]. In Germany, *'BFIT-Bund'* is the federal office that collects the reports centrally, but each state has its own office responsible for monitoring the state of IT accessibility[5].

3 METHODS

As mentioned in the Introduction, we used a combination of computational (web measurements) and qualitative methods to investigate the prevalence of LS and other forms of simplified German. The

[1] Available at https://www.etsi.org/deliver/etsi_en/301500_301599/301549/03.02.01_60/en_301549v030201p.pdf

[2] https://www.gesetze-im-internet.de/bgg/BGG.pdf

[3] https://www.gesetze-im-internet.de/bitv_2_0/BJNR184300011.html

[4] These reports are accessible at https://digital-strategy.ec.europa.eu/en/library/web-accessibility-directive-monitoring-reports

[5] These offices are listed at https://www.bfit-bund.de/DE/Kontakt/Ueberwachungsstellen-der-Laender/ueberwachungsstelle_laender_node.html

code that accompanies the paper and is described in this section can be accessed at https://github.com/hadiasghari/lsweb23.

For the web measurements, we built a classifier to detect pages that are in simplified German. We started with a simple 2-layer feed-forward neural classifier with a bag-of-words model. We trained this classifier on approximately 10,000 German texts balanced between two complexity levels[6], consisting of newspaper articles, web sites, and online lexica for children [6, 13]. Our test set consisted of 2,500 texts from the same data sources.[7] This classifier had an accuracy of 0.99 on the test set, which made it unnecessary to experiment with more complex neural architectures.

Our intention was to run this classifier on the January 2022 Oscar [1] dataset, which is a version of the Common Crawl which has been filtered according to language, and in this manner curate a comprehensive dataset of all webpages that are in LS. We did so, but in the process discovered that the Common Crawl is far from comprehensive, and many LS pages have never been crawled. We nonetheless used it to draw insights about clusters of easy-to-read German. For the clustering, we needed to know the category of the websites, for which we used Curlie, the "largest human-edited directory of the web". [8]

We decided to expand our web survey—given the limits of the Common Crawl—by contacting the offices responsible for monitoring the accessibility Directive in the 16 German states, plus the federal Government, and asking them directly for the list of public sector sites they oversee. Only five states shared their monitoring lists with us, for reasons that we shall dive into later in the paper.

We also added a selection of websites and services that a citizen typically interacts with from Berlin and the federal government, bringing the size of our curated list to approximately 2,000 public sector websites. Amongst these services are, for instance, the central city services, the city recycling, the water supply, the civil registry office, the tax office, and the employment agency.

We downloaded and controlled the quality of the LS on the public sector websites in an automated manner, as follows. First, we used a crawler based on Scrapy[9] to find all links on the homepage matching the regular expression `'leicht.+sprach'` (in their URL or description).[10] Once we found such a page, we extracted the readable content without the HTML markup, and recursively searched for further pages. We then passed the extracted texts to the rule-based LS checker from LanguageTool[11] for a quality check. This open-source tool automatically checks for linguistic characteristics, such as difficult or abstract words, the use of the genitive case, subordinate clauses, and other LS rules.

Finally, to better contextualize and understand the prevalence of LS, we asked the state authorities in over fifty emails and five in-depth phone interviews a series of questions. The questions spanned the process of 'translating' pages into LS, the costs involved, feedback they might have received from website users, and any other information regarding usage. We report these richer but anecdotal findings as well.

[6]These are simplified German texts, at the A2/B1 level and spanning various forms of simplified language including LS, and regular German texts.
[7]For more details about the classifier and training process, please see the accompanying repository code.
[8]We used the Curlie dataset curated by the Homepage2Vec project [11] along with a series of heuristics to collapse the multiple categories of a website to one main category.
[9]https://scrapy.org/
[10]As explained in the legal background, public sector websites need to have a link to their LS pages on their homepage. This regular expression was chosen after manually inspecting how this link had been implemented on dozens of federal and state websites.
[11]https://languagetool.org/de/leichte-sprache

4 FINDINGS

4.1 Science & governmental websites use difficult language

We processed around 17 million pages from the Oscar dataset that have a .de domain and are also present in Curlie. Our classifier identifies 15.5% of pages as using easy-to-read language (not necessarily LS). Table 1 breaks this statistic down according to the website categories. Our analysis shows that pages that belong to the categories science and governmental/regional have the lowest proportion of easy-to-read pages. At the other end of the scale, pages from the categories news, games, and kids use on average simpler language.

Category	URLs	Domains	Easy Pages
Science	2,859,859	7,483	4.3%
Gov. / Region	2,617,151	24,467	7.2%
Business	1,738,920	36,666	9.5%
Shopping	642,924	4,812	18.2%
Health	700,874	12,138	18.7%
Society	1,471,080	15,748	19.0%
Computers	763,282	3,841	19.0%
Recreation	2,805,087	28,243	19.9%
Travel	470,097	9,290	21.6%
Kids	182,729	864	24.9%
Games	267,319	1,413	29.4%
News	2,139,275	1,236	30.5%

Table 1: Categories of websites in our subset of Oscar which have been classified as easy-to-read by our classifier.

A second finding of this analysis was that the Common Crawl, and therefore also the Oscar dataset, do not offer a full picture of the web. We manually collected 319 URLs which definitely contain LS, and found just nine of these pages in the Common Crawl. This low ratio is not ideal for research that aims to offer a comprehensive picture of the web. Moreover, since the sampling process used for the Crawl is not clear to us, we do not know how representative the sample is. Other limitations we observed were duplicate pages and that the text for some of the pages appear to be incomplete.[12] We highlight these limitations to underscore that the results of Table 1 are experimental in nature, and that being aware of these limitations is valuable for the web science community.

4.2 Quality of LS is good, but its coverage not so

Quite strikingly, and perhaps in contrast to the extensive efforts around the implementation of accessibility more generally, the legal requirements for LS do not consider the quality or breadth of the webpages that need to be covered.

[12]These limitations can potentially be solved, but they are clearly not trivial, as otherwise the makers of the Common Crawl would have already done so.

Region	Sites checked	Sites w. LS	Sites w. > 3 LS p.	Rule breaks*
State 1	1443	452	8	0.9
State 2	25	13	6	1.3
State 3	69	43	10	1.5
State 4	45	3	2	2.3
State 5	64	4	2	0.9
State 6	311	130	2	1.9
Federal	16	15	10	2.1
Total	1973	660	40	

Table 2: Pages from the regions. (*The number of rule breaks is calculated per 100 characters, averaged across the LS pages for that region.)

Our conversations with the public monitoring offices confirmed that they merely check for the existence of one page in LS.

We used an automated pipeline to analyse the quality of a total of 2,413 LS pages that our crawler found on our list of 1973 public sector websites. The results are presented in Table 2. The table also lists the number of websites that have more than three pages of content in LS, as an indication of broader coverage.

The pipeline checked whether various rules for LS have been broken, and counted and normalized this count by the length of the page text. The rule breaks per 100 characters ranged from .9 to 2.3. The rules most commonly broken were the use of abstract words and difficult words. As has been argued in the literature [4], 'difficult' words should not necessarily be avoided, depending on the context, goal, and intended reader of a text. Although rules can be a useful way to orient the writing of texts, we acknowledge that they should be interpreted only loosely as an indicator for quality. That being said, based on the relatively low amount of rule-breaks, we would characterize most of these LS pages to be of high quality. This is perhaps not surprising, as the public offices responsible for accessibility told us that the vast majority of LS texts are written by external specialist translation agencies (and often verified by users of LS). However, regarding the breadth of the coverage, as the table shows, very few websites went above and beyond the minimum requirement.

4.3 The list of which public sites are monitored is (inadvertently) political

Although there is a federal office in Germany that is responsible for collating the results that the states submit on their accessibility inspections, the implementation of accessibility is realized on a state level, as explained in section 2.2. We requested information from all 16 state-level offices, asking them to describe their inspection process in detail, since the implementation of accessibility standards is checked in an extensive manner.

Twelve of the state offices replied to our request. Their answers varied quite drastically in regards to openness concerning their workflow and data. Five states shared their list of (all or inspected) public sector websites with us. Other states replied that such a list either does not exist, or that it cannot be shared with us for "legal reasons". One state explicitly replied that they do not publish which websites are checked as they aim to build a relationship of trust between the public administration and the monitoring body; and thus do not want to inadvertently "name and shame" websites which may have done poorly in the accessibility inspection.

The method for assessing websites differs amongst the states, but it is always based on a random selection of their public sector websites, with one state adding that public services can also request to have their website inspected. The inspection is done partly manually, and partly with the aid of some tools, which do not appear to be shared among states.

In our opinion, not publicly sharing the list of checked websites is a failed opportunity, as it would allow communities interested in accessibility to give feedback on what has more priority for them, and it would also be a step towards having a harmonized testing suite. The fact that only a percentage of the public sites will be chosen for testing in itself is not controversial, given available resources, and also common practice in other countries [10].

4.4 Costs of doing or not-doing LS

In general, the feedback we received from the public offices is that there are some workflows in place which support the process of creating content in LS. The costs, which some offices were willing to share with us—although most did not or felt it was hard to estimate—ranged between 140-250 Euro for checking and certifying already translated content in LS by an external service, to 4,900 Euro for the whole process of analyzing and translating a site.

As the offices report, the implementation of LS sometimes requires many organizational steps. The process of implementation according to the offices' estimates varied in length, between six weeks, in which case the office received support from their state government, and six months, in which case the implementation was done by the office on top of their usual duties. The competence center for digital accessibility in Berlin, which at the same time functions as the state's office to monitor the state of accessibility, writes in its newsletter that "all services holding a berlin.de domain are eligible to receive assistance with the implementation of accessibility measures".[13]

One state office explained in an interview that they see it as a shortcoming that they have no power to ensure compliance. Even though they report missing accessibility features to the federal and European authorities, they have no means to ensure their implementation in the future, by for instance imposing sanctions for non-compliance after a set time. The state offices therefore face the problem that the compliance and realization of accessibility is up to the good efforts of the public administrators in each office.

Overall, the process to implement LS seems to be doable for public offices. Institutional support exists: processes can be managed in-house or with the help of existing and professional external providers for translation in LS. Only anecdotal remarks in the interviews suggest that the implementation of other more technical aspects of accessibility in practice are perceived as confusing, since the myriad of laws leave space for multiple technical interpretations. LS, as specific as it is, does not seem to pose an undue burden on the public sector.

[13] https://www.berlin.de/moderne-verwaltung/barrierefreie-it/anlaufstellen/kompetenzstelle/artikel.988002.php

4.5 Estimating LS usage is difficult

We additionally asked the state monitoring bodies, as well as the accessibility officers of all German ministries, about the number of impressions on their LS pages. Our goal was to evaluate how extensively the LS pages are visited and used.

The replies again gave a varying range, from 2,000-3,000 visits per month being the highest (reported by one of the federal ministries), to only 1 click per month for a specific state's parliamentary website. A state office for political education reported an average of 115 clicks by 77 different people per month, over the last twelve months, on their pages in LS.

The main challenge, reported by multiple parties, was that good estimates require an opt-in of users to allow the tracking of visits, which due to the 'cookie rules' are not possible.

Ultimately, when creating content in LS which has the goal of making content more accessible, more information on the usage of the pages would be ideal. Information on page impressions is the bare minimum, as it is not a reliable indicator of usage, and even this information appears hard to gather. For the statistics that we did receive, the large differences in page impressions indicates that some content might just not be a priority for the target groups of LS; the waste collection services and public transportation in Berlin, for example, both reported fairly high page impressions, indicating their relevance.

5 DISCUSSION

Our findings portray a glass half full, half empty situation.

On the one hand, even though the goal of a truly accessible web still appears far away—with varying levels of implementation in the public sector—we conclude that the EU Web Accessibility Directive has succeeded in at least two ways: (i) all German states have monitoring offices set up, with personnel that are understanding and competent with regards to web accessibility requirements; (ii) processes and suppliers are in place that can create LS pages at costs that are reasonable for the public sector. In other words, the Directive's objectives of shifting the burden of monitoring accessibility to the public authorities, and using their procurement powers to incentivize suppliers, seem to be working, although perhaps at a slower pace than one would hope.

There are, nonetheless, several aspects that can in our opinion be improved, both technically and administratively.

First, we recommend that all monitoring bodies publicly release the list of websites that are in their jurisdiction, as well as those that they inspect for accessibility. This would allow the various communities serving accessibility to help prioritize the content and services that needs to be enhanced.[14] With regards to LS specifically, we also recommend running web experiments and surveys to better understand the use of LS among target users.

Second, and in parallel, we recommend that the web technical community crowdsource a list of public sector websites across Europe (and even the world), and crawl them periodically and report on their state of accessibility. This will help reach a fair and inclusive web that works for all persons, and is necessary since existing archives such as the Common Crawl currently miss much of the public sector sites.

Third, we recommend that the W3C adds new HTML tags to classify 'easy language', be it in German (LS) or any other language. Such tags would make it easier for crawlers to identify simplified texts, which could in turn help build datasets to train models for automatic text simplification, thus allowing for a larger amount of web content to be available in easy language.

Lastly, we suggest that using LS to improve understandability should be combined with efforts to simplify the bureaucratic process in general, and striving to make all government communications and content more understandable—instead of offering only navigation and key information in LS. While we acknowledge this will not be easy, some other countries seem to have adopted this strategy towards a more inclusive society.[15]

[14] In the United Kingdom, for example, members of the public can request pages to be monitored via a number of mechanisms [10].

[15] See, for instance, the American Plain Language Act and guidelines at https://www.plainlanguage.gov, and the Dutch Government's inclusive communication guidelines https://www.communicatierijk.nl/vakkennis/inclusieve-communicatie/tips-voor-inclusieve-communicatie

ACKNOWLEDGMENTS

We would like to thank Ferdinand Bálint Dorendorf, Irina Kühnlein, Daniel Pothmann, and Christopher Richter for their help in collecting research data. Furthermore, we would like to thank the interviewees from the federal, state, and local governments across Germany who shared their information and valuable insights with us. This research was funded by a grant from the German Ministry of Education and Research (BMBF).

REFERENCES

[1] Julien Abadji, Pedro Ortiz Suarez, Laurent Romary, and Benoît Sagot. 2022. Towards a Cleaner Document-Oriented Multilingual Crawled Corpus. *arXiv e-prints*, Article arXiv:2201.06642 (Jan. 2022), arXiv:2201.06642 pages. arXiv:2201.06642 [cs.CL]

[2] Alessia Battisti, Dominik Pfütze, Andreas Säuberli, Marek Kostrzewa, and Sarah Ebling. 2020. A Corpus for Automatic Readability Assessment and Text Simplification of German. In *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, 3302–3311. http://arxiv.org/abs/1909.09067 arXiv: 1909.09067.

[3] Bettina M Bock. 2015. Zur Wirksamkeit von Regellisten Leichter Sprache in verschiedenen Kommunikationsbereichen und im World Wide Web. *trans-kom* 8, 1 (2015), 79–102. http://www.trans-kom.eu/bd08nr01/trans-kom_08_01_04_Bock_Leichte_Texte.20150717.pdf

[4] Bettina M Bock. 2019. „Leichte Sprache" – Kein Regelwerk. In *Kommunikation - Partizipation - Inklusion; Band 5*. Frank & Timme GmbH, 1–99.

[5] European Parliament and Council. 2016. Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies (Text with EEA relevance). https://eur-lex.europa.eu/eli/dir/2016/2102/oj

[6] Freya Hewett and Manfred Stede. 2021. Automatically evaluating the conceptual complexity of German texts. In *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*. KONVENS 2021 Organizers, Düsseldorf, Germany, 228–234. https://aclanthology.org/2021.konvens-1.23

[7] Gudrun Kellermann. 2014. Leichte und Einfache Sprache – Versuch einer Definition. *bpb: Aus Politik und Zeitgeschichte* 64, 9–11 (2014), 7–10. https://www.bpb.de/shop/zeitschriften/apuz/179351/leichte-und-einfache-sprache

[8] Ana Keselj, P. Bego, Krunoslav Zubrinic, and M. Milicevic. 2021. Comparison of Accessibility in EU Public Sector Websites. *undefined* (2021). https://www.semanticscholar.org/paper/Comparison-of-Accessibility-in-EU-Public-Sector-Keselj-Bego/adbafd179b36600064d70f0b98cba155ff92e6d6

[9] David Klaper, Sarah Ebling, and Martin Volk. 2013. Building a German/Simple German Parallel Corpus for Automatic Text Simplification. In *Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*. Association for Computational Linguistics, Sofia, Bulgaria, 11–19. https://aclanthology.org/W13-2902

[10] Sarah Lewthwaite and Abi James. 2020. Accessible at last?: what do new European digital accessibility laws mean for disabled people in the UK? *Disability & Society* 35, 8 (Sept. 2020), 1360–1365. https://doi.org/10.1080/09687599.2020.1717446

[11] Sylvain Lugeon, Tiziano Piccardi, and Robert West. 2022. Homepage2Vec: Language-Agnostic Website Embedding and Classification. In *Proceedings of the International AAAI Conference on Web and Social Media*.

[12] Christiane Maaß. 2020. *Easy Language – Plain Language – Easy Language Plus : Balancing Comprehensibility and Acceptability*. Frank & Timme. https://doi.org/10.26530/20.500.12657/42089

[13] Andreas Säuberli, Sarah Ebling, and Martin Volk. 2020. Benchmarking Data-driven Automatic Text Simplification for German. In *1st Workshop on Tools and Resources to Empower People with REAding DIfficulties*. European Language Resources Association (ELRA), Marseille, France, 41–48. https://doi.org/10.5167/UZH-188569 Publisher: European Language Resources Association.

[14] Vanessa Toborek, Moritz Busch, Malte Boßert, Christian Bauckhage, and Pascal Welke. 2022. A New Aligned Simple German Corpus. https://doi.org/10.48550/ARXIV.2209.01106

[15] W3C. 2018. Web Content Accessibility Guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/

Received 30 November 2022