DeepKnowledge: Deep Multilingual Language Model Technology for Language Understanding

Rodrigo Agerri[1], Eneko Agirre[1], Gorka Azkune[1], Roberto Centeno[2], Anselmo Peñas[2], German Rigau[1], Álvaro Rodrigo[2] and Aitor Soroa[1]

[1]HiTZ Center - Ixa, University of the Basque Country UPV/EHU, Donostia-San Sebastián, Spain
[2]NLP & IR Group, UNED, Spain

Abstract

Being language the most efficient system for exchanging information, Natural Language Processing (NLP) is one of the most important technologies of the current digital transformation. In recent years, the NLP community is contributing to the emergence of powerful new deep learning techniques and tools that are revolutionizing the approach to Language Technology (LT) tasks. NLP is moving from a methodology in which a pipeline of multiple modules was the typical way to implement NLP solutions, to architectures based on complex neural networks trained with vast amounts of text data. Thanks to these recent advancements, the NLP community is currently engaged in a paradigm shift with the production and exploitation of large, pre-trained transformer-based language models. Compared to previous work, results are improving so much that systems are claiming to obtain human-level performance in laboratory benchmarks when testing on some difficult language understanding tasks. Despite their impressive capabilities, large pretrained language models do come with severe drawbacks. Currently we have no clear understanding of how they work, when they fail, or which novel ways of exploiting these models can help to improve state-of-the-art in NLP. It is important to understand the limitations of large pretrained language models. DeepKnowledge will investigate on the pre-training of large language models for the official languages in Spain in a way that could be used by applying novel techniques to extract a more precise and generalizable knowledge.

Keywords

Natural Language Processing, Deep Learning, Transfer learning, Language Models, Text Generation, Multitask Learning, Few-show learning, Multimodality, Multilingualism

1. Introduction

The DeepKnowledge (7418127777-127777-4-21) project is funded by MCIN/AEI/10.13039/501100011033 and by FEDER. DeepKnowledge is a coordinated project, in which the consortium is composed by the HiTZ Center - Ixa[1], and UNED NLP and IR Group[2]. The design of the coordinated project is clearly split into two sub-projects with well-defined complementary goals:

Subproject 1: Deep Language Models for Understanding and Reasoning with Multilingual Content (DeepKnowledge-EHU). This subproject will aim to investigate and develop enabling techniques, methods and novel deep learning tools towards Natural Language Understanding (NLU), with special attention to those related with the semantics and knowledge bases, transformer-based pre-trained language models, and machine-operable representations of textual documents (Reference: PID2021-127777OB-C21).

Subproject 2: Deep Language Models for Understanding Information and Misinformation in Context (DeepInfo-UNED). This subproject will exploit the state of the art techniques, methods and tools developed by the previous subproject to study their potential application to two general scenarios: the assessment of language comprehension and the identification of misinformative messages (Reference: PID2021-127777OB-C22).

In this paper we will provide a the description of both subprojects, which, coordinated by HiTZ, focuses on investigating Large Language Models for Multilingual Content.

In recent years, the Natural Language Processing (NLP) community is contributing to the emergence of powerful new deep learning techniques and tools that are revolutionizing the approach to Language Technology (LT) tasks. We are moving from a methodology in which a pipeline of multiple modules was the typical way to implement NLP solutions, to architectures based on complex neural networks trained with vast amounts of text data. This rapid progress in NLP has been possible because of the confluence of four different research trends: 1) mature deep neural network technology, 2) large amounts of data (and for NLP processing large and diverse multilingual textual data), 3) increase in High Performance Computing (HPC) power in the form of GPUs, and 4) application of simple but effective self-learning and transfer

SEPLN-CEDI-PD 2024: Seminar of the Spanish Society for Natural Language Processing: Projects and System Demonstrations, June 19-20, 2024, A Coruña, Spain
rodrigo.agerri@ehu.eus (R. Agerri)

[1]https://hitz.eus
[2]https://sites.google.com/view/nlp-uned/home

learning approaches using Transformers [1, 2, 3, 4].

Thanks to these recent advancements, the NLP community is currently engaged in a paradigm shift with the production and exploitation of large, pre-trained transformer-based language models [1, 3]. As a result, many in the industry have started deploying large pre-trained neural language models in production. For instance, Google and Microsoft have integrated them in their search engines, their flagship product. Compared to previous work, results are improving so much that systems are claiming to obtain human-level performance in laboratory benchmarks when testing on some difficult language understanding tasks.

Furthermore, recent work has shown that pre-trained language models can robustly perform for NLP tasks in a few-shot or even in zero-shot fashion when given an adequate task description in its natural language prompt [2, 5]. Surprisingly, fine-tuning pre-trained language models on a collection of tasks described via instructions (or prompts) substantially boosts zero-shot performance on unseen tasks [6, 4].

Despite their impressive capabilities, large pre-trained language models do come with severe drawbacks. Currently we have no clear understanding of how they work, when they fail, and what emergent properties they may present, or which novel ways of exploiting these models can help to improve state-of-the-art in NLP. As argued by Bender et al. [7], it is important to understand the limitations of large pre-trained language models, which some have called "stochastic parrots". To tackle these questions, much critical multidisciplinary collaboration and research is needed.

DeepKnowledge will extend the state-of-the-art in natural language processing (NLP) and multilingual knowledge enabling technologies in seven interrelated areas of high potential impact. The main research objective of DeepKnowledge consists in advancing the state-of-the-art towards NLU by (i) generating and exploiting new language models for the official languages of Spain plus English by taking into account a multitask and multimodal objective during the pre-training; (ii) exploring novel ways, such as prompting, of exploiting these language models to improve NLP results on zero-shot and few-shot settings (without or very little training data for the target language or task at hand); (iii) by addressing language understanding tasks by text generation; (iv) by leveraging pre-trained language models and building knowledge bases from scratch, (v) developing new benchmarks and datasets for evaluating and assessing the our progress towards Natural Language Understanding; (vi) to apply the newly developed techniques to improve the state-of-the art in language understanding, especially for settings with few or non-existing training data and (vii) by developing a number of advanced content-based domain applications for the main official languages in Spain (including Spanish, Catalan, Basque, and Galician) and English, in multiple sectors and domains (such as eLearning, eHealth, eHumanities, etc).

2. Related Work

Currently, the NLP field is undergoing a paradigm shift with the rise of Large Language Models (also known as Pre-trained Language Models) that are trained on broad data at scale and are adaptable to a wide range of monolingual and multilingual downstream tasks [1, 2]. Though these models are based on standard self-supervised deep learning and transfer learning, their scale results in new emergent and surprising capabilities.

In **self-supervised learning**, the language model is derived automatically from large volumes of unannotated language data. There has been considerable progress in self-supervised learning since word embeddings [8] associated word vectors with context-independent vectors. Shortly thereafter, self-supervised learning based on autoregressive language modelling (predict the next word given the previous words) became popular [9]. The next wave of developments in self-supervised learning — BERT [1], GPT-3 [2], RoBERTa [10], T5 [6], among others — quickly followed, embracing the Transformer architecture [11], incorporating more powerful deep bidirectional encoders of sentences, and scaling up to larger models and datasets.

The idea of **transfer learning** is to take the *knowledge* learned from one task (e.g., predict the next word given the previous words) and apply it to another task (e.g., summarization). With transfer learning, instead of starting the learning process from scratch, you start from patterns that have been learned when solving a different problem. This way you leverage previous learning and avoid starting from scratch. Within deep learning, pre-training is the dominant approach to transfer learning: the objective is to pre-train a deep transformer model on large amounts of data and then reuse this pre-trained language model by fine-tuning it on small amounts of (usually annotated) task-specific data. Thus, transfer learning formalizes a two-phase learning framework: a pre-training phase to capture knowledge from one or more source tasks, and a fine-tuning stage to transfer the captured knowledge to many target tasks.

2.1. Few-shot Learning

Recent work has shown that pre-trained language models can robustly perform classification tasks in a few-shot or even in zero-shot fashion, when given an adequate task description in its natural language prompt [2]. Unlike traditional supervised learning, which trains a model to take in an input and predict an output, prompt-based learning is based on exploiting pre-trained language models to solve a task using text directly [5]. To use these models to perform prediction tasks, the original input is modified using a template into a textual string prompt that has some missing slots, and then the language model is used to probabilistically fill the missing information to obtain a final string, from which the final output for the task can be derived. This framework looks very promising for a number of reasons: it allows the language model to be pre-trained on massive amounts of raw text, and by defining a new prompting function the model is able to perform few-shot or even zero-shot learning, adapting to new scenarios, languages and domains with few or no labeled data. Thus, some NLP tasks can be solved in a fully unsupervised fashion by providing a pre-trained language model with task descriptions in natural language [6]. Surprisingly, fine-tuning pre-trained language models on a collection of tasks described via instructions (or prompts) substantially boosts zero-shot performance on unseen tasks [6, 2, 4].

2.2. Multilingual Language Models

Multilingual Language Models (MLLMs) such as mBERT [1], XLM-RoBERTa [12], mT5 [13], etc. have emerged as a viable option for bringing the power of pre-training to a large number of languages. For example, mBERT is pre-trained with the Multilingual Masked Language Modeling (MMLM) task using non-parallel multilingual Wikipedia corpora in 104 languages. mBERT has the ability to generalize cross-lingual knowledge in zero-shot scenarios. This indicates that even with the same structure of BERT, using multilingual data can enable the model to learn cross-lingual representations. A MLLM is pre-trained using large amounts of unlabeled data from multiple languages with the hope that low-resource languages may benefit from high-resource languages due to a shared vocabulary and latent language properties. The surprisingly good performance of MLLMs in crosslingual transfer as well as bilingual tasks motivates the hypothesis that MLLMs are learning universal patterns [14, 15]. Thus, of particular interest is the ability of MLLMs to facilitate zero-shot crosslingual transfer from a resource-rich language to a resource-deprived language which does not have any task-specific training data, or to fine-tune more robust language models by using annotated training data in multiple languages.

2.3. Text Generation

Natural Language Generation (NLG) has become one of the most common yet challenging tasks in NLP which is currently being addressed by the intense development and release of many Large Language Models (LLMs) such as the popular GPT family, Llama and Mistral models [2, 3, 4]. One of the advantages of these neural models is that they enable end-to-end learning of semantic mappings from input to output in text generation. These decoder models [2, 3, 4] are currently the standard architectures for generating high quality text which in turn generates a crucial need for the evaluation of the generated text.

In DeepKnowledge the progress will be measured by developing new understanding and generation natural language benchmarks and tasks for Basque, Spanish and English, focusing on the truthfulness and reliability of the output generated by the LLMs. Thus, we will provide new benchmarks for popular tasks based on text generation and understanding such as Long Answer Question Answering, Explanatory Argument Generation and Inferential tasks for which annotated data for evaluation exists only for English. By doing so we are aiming at significantly improving the state-of-the-art of AI-based Large Language Models in low resource scenarios for languages such as Basque and Spanish thereby contributing to the improvement of Language Technology Applications and its deployment in the current digital transformation.

2.4. Applications

Current NLP technology allows many *advanced applications* which have been unthinkable only a few years ago. NLP is present in our daily lives, for example, through search engines, recommendation systems, virtual assistants, chatbots, text editors, text predictors, automatic translation systems, automatic summaries, inclusive technology, etc [16]. Its rapid development in recent years predicts even more encouraging and also exciting results in the near future [17]. Currently, our society is developing some fears towards the digital world associated with information distrust of what is published given the growing amount of false content. Our project aims at alleviating these problems by developing new methods and advancing the state of the art in machine reading comprehension of language and misinformation detection.

In this project we target five application scenarios, namely, eLearning, Question Answering and Machine Comprehension, Misinformation, Biomedical Text Analysis and Conversational Agents. In all these application areas we will apply the latest neural language model technology developed within the project.

Recent progress in NLP has been driven by advances in both language model architecture and model pre-training. Transformer architectures have facilitated the building of higher-capacity language models for a wide variety of tasks. Open-source libraries such as Transformers [18] may open up these advances to a wider NLP community. The library consists of carefully engineered state-of-the art Transformer architectures under a unified API and a curated collection of pre-trained models. Unfortunately, the resources necessary to create the best-performing neural language models are found almost exclusively at US and China technology giants. Moreover, this transformative technology poses problems from a research advancement, environmental, and ethical perspective. For example, models such as GPT-3 or GPT-4 are private, anglo-centric, and inaccessible to academic organisations [19]. There are also worrying shortcomings in the text corpora used to train these models, ranging from a lack of representation of populations, to a predominance of harmful stereotypes, and to the inclusion of personal information.

3. Methodology and Work Plan

In this context of paradigm shift within the NLP community, DeepKnowledge will aim to develop new language models (i) with multitask and multimodal training objective (ii) for specific domains, (iii) and to explore novel methods of exploiting such language models such as the use of prompts or text generation, which we believe will help these pre-trained models to ground their knowledge improving understanding and generalization skills.

3.1. Objectives DeepKnowledge-EHU

DeepKnowledge will build models that are capable to deal with text generation tasks, as well as models that are trained in a multi-task fashion, which have shown to generalize better and yield good results work in zero-shot and few-shot scenarios. We will also work towards filling the current gap on language models in these languages for specific domains, such as Health, Education and Social media. Regarding text processing applications, the research team has ample experience developing NLP tools, both basic NLP modules [20] as well as advanced semantic processing tools in many languages [15, 21, 22]. Following this, we list the specific objectives for DeepKnowledge-EHU:

1. To develop large scale datasets and corpora to pretrain new text generation models for the official languages of Spain (WP2).
2. To come up with novel strategies, such as prompting, to exploit language models for text generation to perform better in zero-shot and few-shot scenarios in cross-lingual and multimodal settings. This will be crucial to improve results on common tasks but also to mitigate the lack of training data for a given language or specific domain (WP3).
3. To apply domain-specific language models to improve state-of-the-art results on applications related with medical text processing, social media or educational and e-learning applications (WP3).
4. To explore how large language models can productively interact with existing semantic networks and ontologies (WP4).
5. To leverage the generated language models to develop state-of-the-art, ready-to-use, deep-learning linguistic processors for many NLP tasks, such as lemmatization, NER, SRL, POS tagging or Coreference Resolution, among others (WP2).
6. To improve qualitative and quantitative evaluation of text generation-based tasks such as text simplification or argument generation; organize a shared task to motivate work on this topic (WP5).
7. To leverage the generated models and new techniques of exploiting them for elearning, Question Answering, Medical Text Processing, Misinformation detection and Conversational Agents (WP6).

3.2. Objectives DeepInfo-UNED

DeepInfo-UNED collaborates with two institutions: (i) Instituto Cervantes and (ii) president Carter Foundation (USA). One of the goals of Instituto Cervantes is the certification of human proficiency in the use of Spanish language. The collaboration between our project and Instituto Cervantes is focused on: (i) creating a dataset in Spanish for the evaluation of machine reading and comprehension capabilities which will address the lack of training and evaluation resources for other languages different to English, (ii) developing automatic assisting methods to help evaluators to prepare and check the exams.

The Carter Foundation acts as an international observer in elections all over the world. Traditionally, these observers were a team of persons that moved physically to the country and tracked the process. However, nowadays there is also a need to monitorize political activity in social networks. By taking into account these two use cases, the specific objectives of DeepInfo-UNED are defined as follows:

1. Study new methods for stance Detection based on the user's profile and network (WP6).
2. Characterize the assessment of intentionality in misinformation diffusion processes (WP5, WP6).
3. Develop new methods for difficulty assessment of Question Answering and Reading Comprehension tests in Spanish (WP6).
4. Organize a shared task on Misinformation (WP5).

3.3. Work Plan

The Work Plan is structured in six Work Packages of which three are focused on the scientific contributions of the project.

WP2: Methodology. DeepKnowledge will build state-of-the-art multilingual language models for all languages in Spain as well as English. The models will be based on news technologies, architectures and training paradigms that allow a better generalization between domains and languages. We will build generative models that allow the generation of text in these languages, which is needed in tasks such as summarization, simplification or generation of counter-arguments against misinformation. Besides, the project will also build language models adapted to specific domains of Health, Education, Social media.

WP3: Novel paradigms for the exploitation of language models. Develop novel ways to exploit the full potential of large language models, including prompting, generation and multimodal training. The objective of such exploitation paradigms is two-fold: (i) to improve the overall language understanding capabilities of language models, and (ii) to make them usable for a great variety of applications and languages with minimal preparation effort, through zero-shot and few-shot learning.

WP4: Knowledge Acquisition, Integration and Reasoning. The main objective of this work package is to investigate how large language models can productively interact with existing semantic networks. On the one hand, helping on the development of broad-coverage lexical knowledge bases such as the Multilingual Central Repository [23] in the languages covered by the project and adapted to specific domains such as medicine. On the other hand, using these large-scale knowledge bases to generate lexical semantic, world knowledge and common sense probes for testing the abilities of modern large language models.

WP5: Evaluation. the objective of this work package is to measure the research progress via objective evaluation metrics and relevant open evaluation campaigns. An important component will also be investigating the evaluation of tasks based on text generation (WP3). Datasets for Machine Comprehension and Question Answering in Spanish will be generated. Furthermore, we will organize a workshop on misinformation.

WP6: Applications and Use Cases. This work package aims at demonstrating the scientific advances of DeepKnowledge in different scenarios. It will include applications in elearning, recommender systems for education and research, question answering, reading comprehension, and misinformation.

4. Concluding Remarks

This paper outlines the DeepKnowledge project, which is focused on researching and incorporating the latest insights in deep learning technology, such as large pretrained language models, transfer learning, few-shot and zero-shot capabilities, multimodal and multi-task processing, prompting, etc. DeepKnowledge will leverage deep learning techniques and large pre-trained language models and carefully designed datasets and knowledge bases to advance the state of the art towards natural language understanding to English, Spanish, Catalan, Basque and Galician in several domains and digital sectors. DeepKnowledge will also investigate new text generation approaches for applications such as argument generation, text simplification or abstractive summarization. Additionally, DeepKnowledge will apply the new language models in novel ways for tasks and applications such as misinformation detection, Question Answering or elearning.

Ongoing work can be checked in the project's website: http://ixa2.si.ehu.eus/deepknowledge/. Future work includes further experimentation training LLMs for low-resource languages and on the evaluation of text generation, a crucial topic to understand the performance of our models.

Acknowledgments

We acknowledge the support of DeepKnowledge (PID2021-127777OB-C21) and DeepInfo (PID2021-127777OB-C22), projects funded by MCIN/AEI/10.13039/501100011033 and by FEDER. Rodrigo Agerri was also funded by the RYC-2017-23647 fellowship (MCIN/AEI/10.13039/501100011033 and by ESF Investing in your future).

References

[1] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: North American Chapter of the Association for Computational Linguistics, 2019.

[2] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. J. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems, volume 33, Curran Associates, Inc., 2020, pp. 1877–1901.

[3] H. Touvron, L. Martin, K. R. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. M. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. S. Hartshorn, S. Hosseini, R. Hou, H. Inan, M. Kardas, V. Kerkez, M. Khabsa,

I. M. Kloumann, A. V. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovich, Y. Lu, Y. Mao, X. Martinet, T. Mihaylov, P. Mishra, I. Molybog, Y. Nie, A. Poulton, J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, T. Scialom, Llama 2: Open foundation and fine-tuned chat models, ArXiv abs/2307.09288 (2023).

[4] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de Las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, W. E. Sayed, Mistral 7b, ArXiv abs/2310.06825 (2023).

[5] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, G. Neubig, Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing, ACM Computing Surveys 55 (2021) 1 – 35.

[6] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, Journal of Machine Learning Research 21 (2020) 1–67.

[7] E. M. Bender, T. Gebru, A. McMillan-Major, M. Mitchell, On the dangers of stochastic parrots: Can language models be too big?, Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (2021).

[8] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, A. Joulin, Advances in pre-training distributed word representations, in: Proceedings of the 11th Language Resources and Evaluation Conference, Miyazaki, Japan, 2018.

[9] A. M. Dai, Q. V. Le, Semi-supervised sequence learning, NeurIps (2015).

[10] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, RoBERTa: A robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692 (2019).

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in neural information processing systems, 2017, pp. 5998–6008.

[12] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Annual Meeting of the Association for Computational Linguistics, 2019.

[13] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, C. Raffel, mT5: A massively multilingual pre-trained text-to-text transformer, in: NAACL, Association for Computational Linguistics, 2021.

[14] T. Pires, E. Schlinger, D. Garrette, How multilingual is multilingual BERT?, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2019, pp. 4996–5001.

[15] R. Agerri, E. Agirre, Lessons learned from the evaluation of Spanish Language Models, Proces. del Leng. Natural 70 (2023) 157–170.

[16] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heinz, D. Roth, Recent advances in natural language processing via large pre-trained language models: A survey, ACM Computing Surveys 56 (2021) 1 – 40.

[17] X. Han, Z. Zhang, N. Ding, Y. Gu, X. Liu, Y. Huo, J. Qiu, Y. Yao, A. Zhang, L. Zhang, et al., Pre-trained models: Past, present and future, AI Open 2 (2021) 225–250.

[18] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, A. Rush, Transformers: State-of-the-art natural language processing, in: Q. Liu, D. Schlangen (Eds.), Proceedings of EMNLP: System Demonstrations, 2020, pp. 38–45.

[19] L. Floridi, M. Chiriatti, Gpt-3: Its nature, scope, limits, and consequences, Minds and Machines 30 (2020) 681–694.

[20] O. Toporkov, R. Agerri, On the Role of Morphological Information for Contextual Lemmatization, Computational Linguistics (2024) 1–35. doi:10.1162/coli_a_00497.

[21] O. Sainz, I. García-Ferrero, R. Agerri, O. de Lacalle, G. Rigau, E. Agirre, Gollie: Annotation guidelines improve zero-shot information-extraction, Twelfth International Conference on Learning Representations (ICLR 2024), 2024.

[22] I. García-Ferrero, R. Agerri, A. A. Salazar, E. Cabrio, I. de la Iglesia, A. Lavelli, B. Magnini, B. Molinet, J. Ramirez-Romero, G. Rigau, J. M. Villa-Gonzalez, S. Villata, A. Zaninello, Medical mT5: An Open-Source Multilingual Text-to-Text LLM for The Medical Domain, Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING), 2024.

[23] A. G. Agirre, E. Laparra, G. Rigau, Multilingual Central Repository version 3.0: upgrading a very large lexical knowledge base, in: GWC 2012 6th International Global Wordnet Conference, 2012, p. 118.