

Introducing the NLU Showroom: A NLU demonstrator for the German language



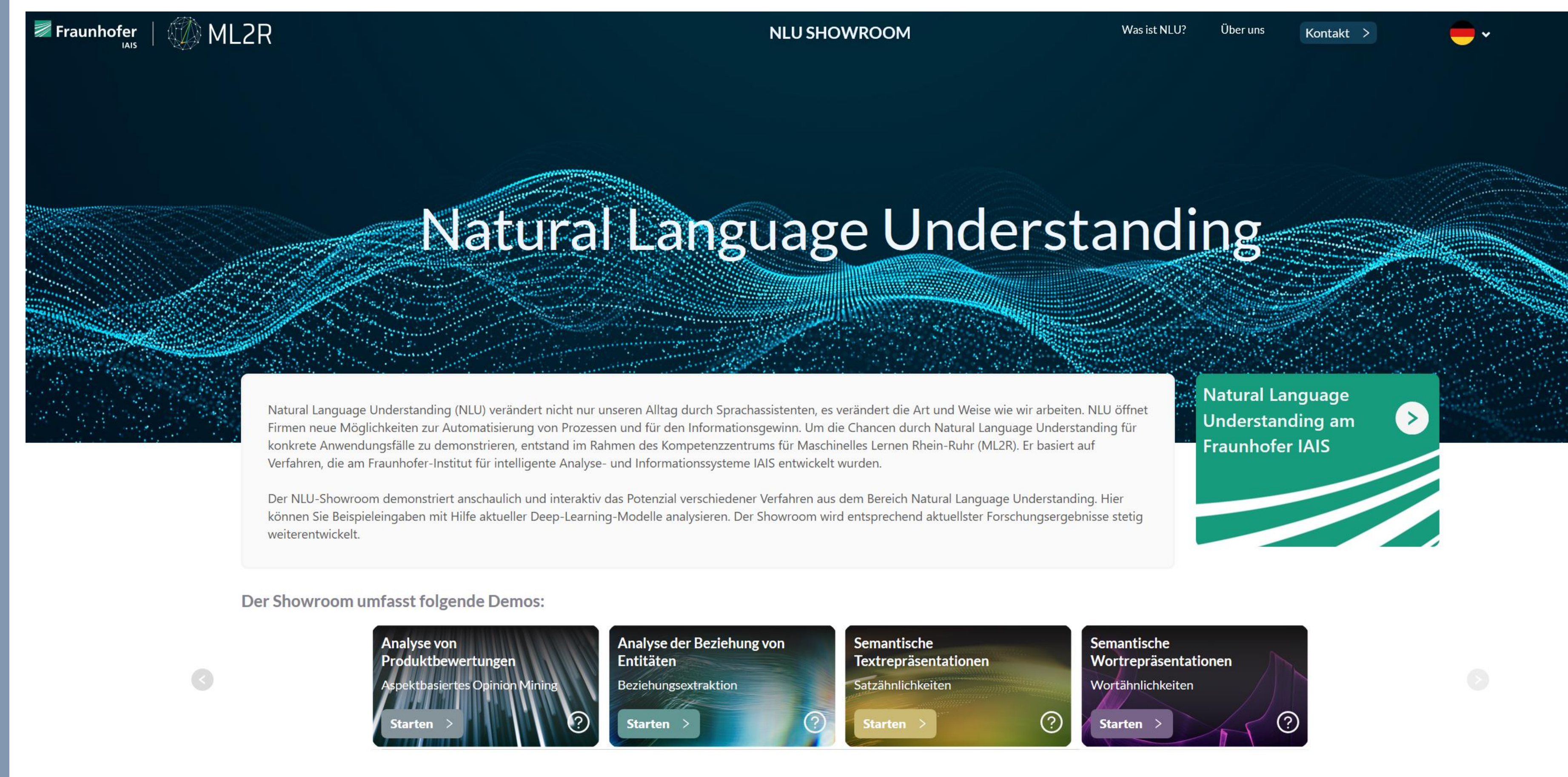
Dennis Wegener^{1,2}, Sven Giesselbach^{1,2}, Niclas Doll^{1,2}, Heike Horstmann^{1,2}
 [1] Fraunhofer IAIS, [2] Competence Center for Machine Learning Rhine-Ruhr



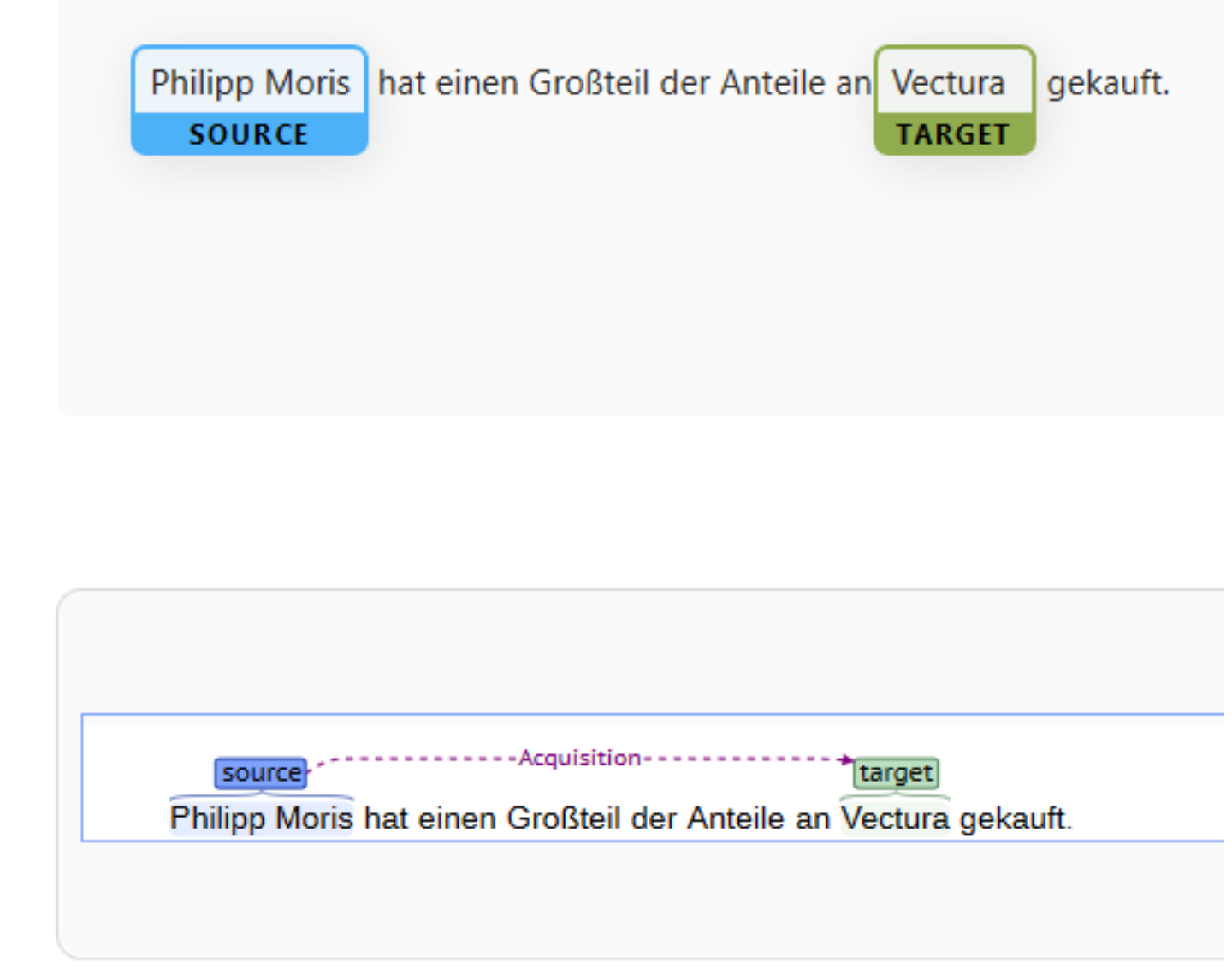
Motivation

- It is becoming more common to use NLP and NLU approaches in software to automate text based processes
- Most resources are available in the English language, there is great unused economic potential in using NLP and NLU in other languages, such as German
- The NLU Showroom should serve as platform to demonstrate NLP and NLU approaches in the German language and to show its potential
- It consists of demos for different tasks, in which state-of-the-art models are demonstrated, so that users can get a realistic assessment of the capabilities of NLP/NLU
- The architecture allows the Showroom to be easily extended with further demos and models (German GPT-2 text generation demo next)

NLU Showroom Website



Relation Extraction & Classification



Bert[2] trained on German Smartdata Corpus (industry related relations)

Semantic Similarity (word level)

w2v	glove	fasttext
Pamplona 0.84	Valencia 0.65	Zaragossa 0.96
Murcia 0.83	Pamplona 0.61	Saragossas 0.94
Badajoz 0.83	Huesca 0.61	Saragossa-Huesca 0.93
Granada 0.83	Madrid 0.6	Murcia 0.9

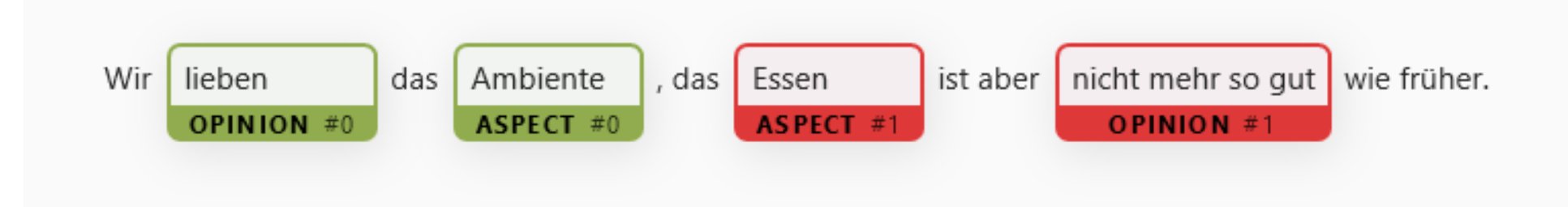
word2vec[6], glove[5], fastText[1] trained on German Wikipedia Corpus

Semantic Similarity (sentence level)

doc2vec_CosSim	w2vec_wmSim	glove_wmSim	fastText_wmSim
Wenn ich nach Bayern fahre, werde ich auf jeden Fall ein Spiel in München schauen. -0.07	Wenn ich nach Bayern fahre, werde ich auf jeden Fall ein Spiel in München schauen. 0.04	Der FCB hat die neunte deutsche Meisterschaft in Folge gewonnen. 0.11	Der FCB hat die neunte deutsche Meisterschaft in Folge gewonnen. 0.19
Der FCB hat die neunte deutsche Meisterschaft in Folge gewonnen. -0.08	Der FCB hat die neunte deutsche Meisterschaft in Folge gewonnen. 0.03	Wenn ich nach Bayern fahre, werde ich auf jeden Fall ein Spiel in München schauen. 0.11	Wenn ich nach Bayern fahre, werde ich auf jeden Fall ein Spiel in München schauen. 0.14

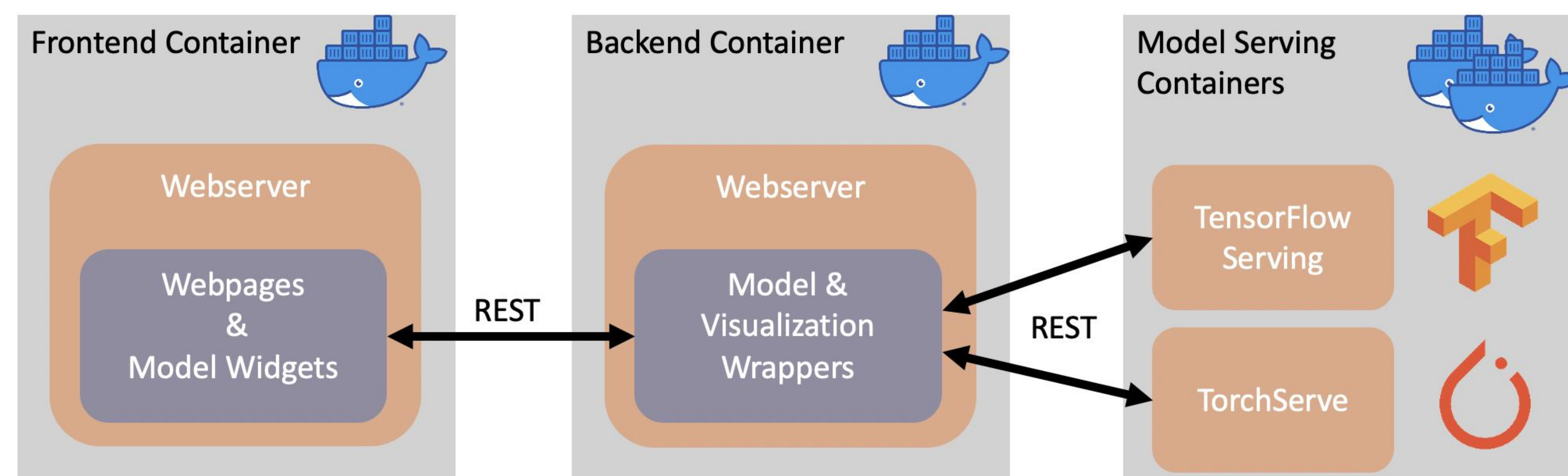
doc2vec[4] and word2vec[6], glove[5], fasttext[1] using wmd[3]. All trained on German Wikipedia Corpus

Aspect based opinion mining



Using Bert[2] for a) Aspect-Opinion Extraction, b) Opinion-based Sentiment Analysis and c) Aspect-Opinion Linking. Trained on manually labeled yelp reviews.

Technical Architecture



References

- [1] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. CoRR, abs/1607.04606, 2016. URL: <http://arxiv.org/225abs/1607.04606>, arXiv:1607.04606.
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, 2019. Association for Computational Linguistics. URL: <https://www.aclweb.org/anthology/N19-1423>, doi:10.18653/v1/N19-1423
- [3] Matt Kusner, Y. Sun, N.I. Kolkin, and Kilian Weinberger. From word embeddings to document distances. Proceedings of the 32nd International Conference on Machine Learning (ICML2015), pages 957–966, 01 2015.
- [4] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. CoRR, abs/1405.4053, 2014. URL: <http://arxiv.org/abs/1405.4053>, arXiv:1405.4053
- [5] Omer Levy and Yoav Goldberg. Neural word embedding as implicit matrix factorization. In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14, page 2177–2185, Cambridge, MA, USA, 2014. MIT Press.
- [6] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. CoRR, abs/1310.4546, 2013. URL: <http://arxiv.org/abs/1310.4546>, arXiv:1310.4546.252
- [7] Christopher Olston, Noah Fiedel, Kiril Gorovoy, Jeremiah Harmsen, Li Lao, Fangwei Li, Vinu Rajashekhar, Sukriti Ramesh, and Jordan Soyke. Tensorflow-serving: Flexible, high-performance ML serving. CoRR, abs/1712.06139, 2017. URL: <http://arxiv.org/abs/1712.25506139>, arXiv:1712.06139.
- [8] Zachary DeVito, Jason Ansel, Will Constable, Michael Suo, Ailing Zhang and Kim Hazelwood, Using Python for Model Inference in Deep Learning, 2021, URL: <https://arxiv.org/abs/2104.00254>, arXiv:2104.00254.