Renato Vukovic

Heinrich Heine University Düsseldorf
Universitätsstr.1
40225 Düsseldorf
NRW, Germany

renato.vukovic@hhu.de
https://renatovukovic.github.io

1 Research interests

My research interests lie generally in **dialogue ontology construction**, that uses techniques from *information extraction* to extract relevant terms from task-oriented dialogue data and order them by finding hierarchical relations between them.

1.1 Dialogue Ontology Construction

Building ontologies for dialogue systems manually is expensive and time consuming (Budzianowski et al., 2018). The ontology is mainly needed by the dialogue state tracking module of task-oriented dialogue (TOD) systems (Heck et al., 2020; van Niekerk et al., 2021), but also important for user simulation Lin et al. (2023). The ontology covers structured information about domains (e.g. hotel), slots (e.g. price range) and values (e.g. expensive), while other conversational aspects such as user emotion can also be part of the dialogue state (Feng et al., 2023). However, generally TOD systems are limited to the knowledge in the ontology, limiting their application to new domains (Feng et al., 2024). Automating parts of the ontology annotation process can thus increase the scalability of TOD systems by making them easier to apply to new domains and unseen data. This is especially interesting when aiming to continually update a TOD system with new knowledge automatically in continual learning setups (Geishauser et al., 2024). Another possibility might be to improve a automatically constructed ontology using label validation approaches from active learning (van Niekerk et al., 2023), since the constructed ontologies come with a significant amount of noise.

Information extraction (IE) aims at structuring information from text data and there are normally two main steps, named entity recognition (NER) and relation extraction (RE) (Genest et al., 2022). Automatic ontology construction can be considered a form of IE for task-oriented dialogue. Ontology construction is about automating the process of building ontologies, rendering manual annotation unnecessary while saving time and making larger portions of unstructured data usable, so the system is able to include new domains, slots and values to talk about dynamically. The process can be separated into three steps, although you can split them up further into more fine-grained steps as well (Toledo-Alvarado et al., 2012; Cimiano et al., 2006):

1. **Term extraction:** extracting relevant domain, slot and value terminology in the textual data and finding concepts
2. **Relation extraction:** predicting hierarchical relations between the concepts, organising them into domains, slots and values
3. **Disambiguation:** ordering the found concepts based on their context such that words with similar meaning or domain end up in the same group, e.g. “expensive” and “high-end”

There are approaches focussing on different steps of the construction process, such as term extractors relying on frequency (Sclano and Velardi, 2007) or induce slots in a data-driven fashion (Qiu et al., 2022). Others extract relevant slots and inducing an ontology hierarchy (Hudeček et al., 2021), which then can be directly used to train a model on a downstream-task, like dialogue state tracking and response generation (Yu et al., 2022) based on the induced slot-schemas.
Apart from that there are also approaches that aim at making state tracking models more versatile and able to handle unseen data (Heck et al., 2022). Furthermore the advent of large language models (LLMs) enables even better generalisation to unseen tasks and domains, such as dialogue state tracking in a zero-shot fashion (Heck et al., 2023).

1.2 Term Extraction

The first step of ontology construction, term extraction aims at capturing all regions of interest in textual data, maximising recall (Nakagawa and Mori, 2002; Wermter and Hahn, 2006). The problem in this first step is that the precision is quite low, which makes additional processing or filtering of non-relevant terms necessary before proceeding with the next step (Frantzi and Ananiadou, 1999). In my research so far I mainly focus on improving the extraction process in terms of precision while keeping the recall at a high level so that less filtering is necessary.

The 20th Annual Meeting of the Young Researchers’ Roundtable on Spoken Dialogue Systems, pages 53–56
September 16–17, 2024.

Furthermore my goal is to develop a term extractor with better generalisation capability to use it on different kinds of datasets, which cover a lot of different domains.

For this goal my group investigates potentially domain-agnostic features of the word embedding space that capture the meaning and the relevance of potential terms. This term extractor model should get terms in a way such that the follow-up steps are as easily feasible as possible to be able to improve the whole ontology construction process in the long-term. Investigated features include features obtained from pre-trained masked language models and ones obtained by applying mathematical tools like topological data analysis on the word embedding space to find meaningful structures. The models trained on these features show good zero-shot transferability to the much larger schema-guided dialogue (SGD) dataset (Rastogi et al., 2020) on the term extraction task (Vukovic et al., 2022) when trained on MultiWOZ (Budzianowski et al., 2018) as seed dataset. The performance of the term extractor can be further improved by computing them on a contextual level rather than on a global static level (Ruppik et al., 2024).

1.3 Relation Extraction and Disambiguation

In relation extraction, we consider three hierarchical relations between domains and slots, domains and values, and slots and values respectively that have to be predicted between terms. Furthermore, we consider an equivalence relation between terms of the same category in order to disambiguate semantically equivalent terms, such as "expensive" and "high-end" as price values. In our initial set-up we predict all the given relations jointly with one model, although experimental results might suggest that more emphasis on disambiguation might be needed. Note that another possibility is to infer the ontology hierarchy via clustering (Yu et al., 2022), which is not in line with most information extraction approaches.

By utilising such general structural relations, our goal is to utilise existing annotated datasets in order to extract semantic information in the form of an ontology on unseen data. In our research on ontology relation extraction, we experiment with updated decoding mechanisms for language models, such as constrained generation and chain-of-thought (CoT) decoding (Wang and Zhou, 2024) in order to improve generalisability of few-shot prompted and fine-tuned language models. In a transfer learning set-up we show that constrained chain-of-thought decoding improves performance of a language model trained on MultiWOZ as seed dataset and SGD as target dataset (Vukovic et al., 2024).

2 Spoken dialogue system (SDS) research

Understanding and acting upon natural language is one of the earliest challenges for artificial intelligence (AI), as it is part of the Turing test (Turing, 1950). Spoken dialogue system research evolved from the goal of solving the Turing test to solving more specific problems related to language, which in my opinion is one of the most important means for human communication and learning. I think SDS will become more and more incorporated in everyday life as you can already see in personal assistants, such as Siri or Amazon's Alexa. As long as they add value to the life of their user by making it more comfortable or make a personal secretary affordable to broader parts of the society, as they are much less expensive than paying real humans for the more and more tasks the systems are capable of.

Language is one of the main means human learning after mastering their mother tongue, since even movements to learn are normally accompanied by descriptions in language. This observation makes me assume that AI capable of understanding and acting upon language in a human-like manner might learn from the same sources humans learn from (Lynn and Bassett, 2020). This accomplishment would make large amounts of textual data usable for training large models with general knowledge and abilities. Altogether this assumption shows the great potential which lies in SDS research.

In my opinion, it is really hard to forecast what will happen in 10 years time, as there can be large jumps of progress if certain milestones are reached.

3 Suggested topics for discussion

- What degree of supervision is needed to extract semantic information from text and build an ontology?
- How can we leverage existing annotated data to structure information on unseen data?
- How to update the ontology of a model dynamically while interacting with users?
- What can you take from human learning and interaction from and with speech to adapt in spoken dialogue systems?
- Which architectures and training approaches are best suited for ontology relation extraction?

Biographical sketch

Renato Vukovic is currently a second year PhD student working on dialogue ontology construction for task-oriented dialogue under the supervision of Prof. Milica Gašić at Heinrich Heine University Düsseldorf. He holds a bachelor's and master's degree in computer science from the HHU Düsseldorf.

References

Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. MultiWOZ - a large-scale multi-domain Wizard-of-Oz dataset for task-oriented dialogue modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Brussels, Belgium, pages 5016–5026. https://doi.org/10.18653/v1/D18-1547.

Philipp Cimiano, Johanna Völker, and Rudi Studer. 2006. Ontologies on demand? A description of the state-of-the-art, applications, challenges and trends for ontology learning from text. *Information, Wissenschaft und Praxis* Vol. 57, No. 6-7:315–320.

Shutong Feng, Hsien chin Lin, Christian Geishauser, Nurul Lubis, Carel van Niekerk, Michael Heck, Benjamin Ruppik, Renato Vukovic, and Milica Gašić. 2024. Infusing Emotions into Task-oriented Dialogue Systems: Understanding, Management, and Generation. *arXiv preprint arXiv:2408.02417* https://arxiv.org/abs/2408.02417.

Shutong Feng, Nurul Lubis, Benjamin Ruppik, Christian Geishauser, Michael Heck, Hsien-chin Lin, Carel van Niekerk, Renato Vukovic, and Milica Gašić. 2023. From chatter to matter: Addressing critical steps of emotion recognition learning in task-oriented dialogue. In Svetlana Stoyanchev, Shafiq Joty, David Schlangen, Ondrej Dusek, Casey Kennington, and Malihe Alikhani, editors, *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, Prague, Czechia, pages 85–103. https://doi.org/10.18653/v1/2023.sigdial-1.8.

Katerina T Frantzi and Sophia Ananiadou. 1999. The C-value/NC-value domain-independent method for multi-word term extraction. *Journal of Natural Language Processing* 6(3):145–179. https://doi.org/10.5715/jnlp.6.3_145.

Christian Geishauser, Carel van Niekerk, Nurul Lubis, Hsien-chin Lin, Michael Heck, Shutong Feng, Benjamin Ruppik, Renato Vukovic, and Milica Gašić. 2024. Learning with an open horizon in ever-changing dialogue circumstances. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 32:2352–2366. https://doi.org/10.1109/TASLP.2024.3385289.

Pierre-Yves Genest, Pierre-Edouard Portier, Elöd Egyed-Zsigmond, and Laurent-Walter Goix. 2022. Promptore-a novel approach towards fully unsupervised relation extraction. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. pages 561–571.

Michael Heck, Nurul Lubis, Benjamin Ruppik, Renato Vukovic, Shutong Feng, Christian Geishauser, Hsien-chin Lin, Carel van Niekerk, and Milica Gašić. 2023. ChatGPT for zero-shot dialogue state tracking: A solution or an opportunity? In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Toronto, Canada, pages 936–950. https://doi.org/10.18653/v1/2023.acl-short.81.

Michael Heck, Nurul Lubis, Carel van Niekerk, Shutong Feng, Christian Geishauser, Hsien-Chin Lin, and Milica Gašić. 2022. Robust dialogue state tracking with weak supervision and sparse data. *Transactions of the Association for Computational Linguistics* 10:1175–1192. https://doi.org/10.1162/tacl_a_00513.

Michael Heck, Carel van Niekerk, Nurul Lubis, Christian Geishauser, Hsien-Chin Lin, Marco Moresi, and Milica Gašić. 2020. TripPy: A triple copy strategy for value independent neural dialog state tracking. In Olivier Pietquin, Smaranda Muresan, Vivian Chen, Casey Kennington, David Vandyke, Nina Dethlefs, Koji Inoue, Erik Ekstedt, and Stefan Ultes, editors, *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, 1st virtual meeting, pages 35–44. https://doi.org/10.18653/v1/2020.sigdial-1.4.

Vojtěch Hudeček, Ondřej Dušek, and Zhou Yu. 2021. Discovering dialogue slots with weak supervision. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online, pages 2430–2442. https://doi.org/10.18653/v1/2021.acl-long.189.

Hsien-Chin Lin, Shutong Feng, Christian Geishauser, Nurul Lubis, Carel van Niekerk, Michael Heck, Benjamin Ruppik, Renato Vukovic, and Milica

Gasić. 2023. Emous: Simulating user emotions in task-oriented dialogues. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, SIGIR ’23, page 2526–2531. https://doi.org/10.1145/3539618.3592092.

Christopher W. Lynn and Danielle S. Bassett. 2020. How humans learn and represent networks. *Proceedings of the National Academy of Sciences* 117(47):29407–29415. https://doi.org/10.1073/pnas.1912328117.

Hiroshi Nakagawa and Tatsunori Mori. 2002. A simple but powerful automatic term extraction method. In *COLING-02: COMPUTERM 2002: Second International Workshop on Computational Terminology*. https://aclanthology.org/W02-1407.

Liang Qiu, Chien-Sheng Wu, Wenhao Liu, and Caiming Xiong. 2022. Structure extraction in task-oriented dialogues with slot clustering. https://doi.org/10.48550/ARXIV.2203.00073.

Abhinav Rastogi, Xiaoxue Zang, Srinivas Sunkara, Raghav Gupta, and Pranav Khaitan. 2020. Towards scalable multi-domain conversational agents: The schema-guided dialogue dataset. In *Proceedings of the AAAI Conference on Artificial Intelligence*. volume 34, pages 8689–8696.

Benjamin Matthias Ruppik, Michael Heck, Carel van Niekerk, Renato Vukovic, Hsien chin Lin, Shutong Feng, Marcus Zibrowius, and Milica Gašić. 2024. Local Topology Measures of Contextual Language Model Latent Spaces With Applications to Dialogue Term Extraction. *arXiv preprint arXiv:2408.03706* https://arxiv.org/abs/2408.03706.

Francesco Sclano and Paola Velardi. 2007. TermExtractor: A web application to learn the shared terminology of emergent web communities. In Ricardo J. Gonçalves, Jörg P. Müller, Kai Mertins, and Martin Zelm, editors, *Enterprise Interoperability II*. Springer London, London, pages 287–290. https://doi.org/10.1007/978-1-84628-858-6_32.

J. Toledo-Alvarado, Adolfo Guzman-Arenas, and G. Martínez-Luna. 2012. Automatic building of an ontology from a corpus of text documents using data mining tools. *Journal of applied research and technology* 10:398–404. https://doi.org/10.22201/icat.16656423.2012.10.3.395.

A. M. Turing. 1950. Computing machinery and intelligence. *Mind* 59(236):433–460. http://www.jstor.org/stable/2251299.

Carel van Niekerk, Christian Geishauser, Michael Heck, Shutong Feng, Hsien chin Lin, Nurul Lubis, Benjamin Ruppik, Renato Vukovic, and Milica Gašić. 2023. CAMELL: Confidence-based Acquisition Model for Efficient Self-supervised Active Learning with Label Validation. *arXiv preprint arXiv:2310.08944* https://arxiv.org/abs/2310.08944.

Carel van Niekerk, Andrey Malinin, Christian Geishauser, Michael Heck, Hsien-chin Lin, Nurul Lubis, Shutong Feng, and Milica Gašić. 2021. Uncertainty measures in neural belief tracking and the effects on dialogue policy performance. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, pages 7901–7914. https://doi.org/10.18653/v1/2021.emnlp-main.623.

Renato Vukovic, David Arps, Carel van Niekerk, Benjamin Matthias Ruppik, Hsien-Chin Lin, Michael Heck, and Milica Gašić. 2024. Dialogue Ontology Relation Extraction via Constrained Chain-of-Thought Decoding. *arXiv preprint arXiv:2408.02361* https://arxiv.org/abs/2408.02361.

Renato Vukovic, Michael Heck, Benjamin Ruppik, Carel van Niekerk, Marcus Zibrowius, and Milica Gašić. 2022. Dialogue term extraction using transfer learning and topological data analysis. In Oliver Lemon, Dilek Hakkani-Tur, Junyi Jessy Li, Arash Ashrafzadeh, Daniel Hernández Garcia, Malihe Alikhani, David Vandyke, and Ondřej Dušek, editors, *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, Edinburgh, UK, pages 564–581. https://doi.org/10.18653/v1/2022.sigdial-1.53.

Xuezhi Wang and Denny Zhou. 2024. Chain-of-thought reasoning without prompting. *arXiv preprint arXiv:2402.10200* https://arxiv.org/abs/2402.10200.

Joachim Wermter and Udo Hahn. 2006. You can’t beat frequency (unless you use linguistic knowledge) – a qualitative evaluation of association measures for collocation and term extraction. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Sydney, Australia, pages 785–792. https://doi.org/10.3115/1220175.1220274.

Dian Yu, Mingqiu Wang, Yuan Cao, Izhak Shafran, Laurent El Shafey, and Hagen Soltau. 2022. Unsupervised slot schema induction for task-oriented dialog. https://doi.org/10.48550/ARXIV.2205.04515.