

Information Extraction – Seminar Topics

WS 2021–2022, English topics

Jindřich Libovický
libovicky@cis.lmu.de

LMU München, CIS

November 3, 2021

Multilingual Named Entity Recognition

The **Mona Lisa** WORK OF ART is a **16th century** DATE **oil painting** CONCEPT
created by **Leonardo** PERSON. It's held at the **Louvre** INSTITUTION in **Paris** LOCATION.

- Training data mostly exist in English and few other high-resource languages
- Pre-trained multilingual embeddings could be used to transfer the knowledge from English to less resourced languages

Multilingual NER: Suggested papers

- J. Ni and R. Florian. **Improving multilingual named entity recognition with Wikipedia entity type mapping.** In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1275–1284, Austin, Texas, Nov. 2016. Association for Computational Linguistics. doi: [10.18653/v1/D16-1135](https://doi.org/10.18653/v1/D16-1135). URL <https://aclanthology.org/D16-1135>
- M. Arkhipov, M. Trofimova, Y. Kuratov, and A. Sorokin. **Tuning multilingual transformers for language-specific named entity recognition.** In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 89–93, Florence, Italy, Aug. 2019. Association for Computational Linguistics. doi: [10.18653/v1/W19-3712](https://doi.org/10.18653/v1/W19-3712). URL <https://aclanthology.org/W19-3712>
- D. Mueller, N. Andrews, and M. Dredze. **Sources of transfer in multilingual named entity recognition.** In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8093–8104, Online, July 2020. Association for Computational Linguistics. doi: [10.18653/v1/2020.acl-main.720](https://doi.org/10.18653/v1/2020.acl-main.720). URL <https://aclanthology.org/2020.acl-main.720>

Stance detection = extraction of a subject's reaction to a claim made by a primary actor. It is a core part of a set of approaches to fake news assessment.

Topic: *Climate Change is a Real Concern*

Tweet: *Gone are the days where we would get temperatures of Min -2 and Max 5 in Cape Town*

Stance: *FAVOR*

Example from Schiller et al., 2021.

Stance Detection: Suggested papers

- M. Hardalov, A. Arora, P. Nakov, and I. Augenstein. A survey on stance detection for mis- and disinformation identification. *CoRR*, abs/2103.00242, 2021. URL <https://arxiv.org/abs/2103.00242>
- B. Schiller, J. Daxenberger, and I. Gurevych. Stance detection benchmark: How robust is your stance detection? *KI - Künstliche Intelligenz*, 2021. ISSN 1610-1987. doi: 10.1007/s13218-021-00714-w. URL <https://doi.org/10.1007/s13218-021-00714-w>
- J. Vamvas and R. Sennrich. X-stance: A multilingual multi-target dataset for stance detection. In S. Ebling, D. Tuggener, M. Hürlimann, M. Cieliebak, and M. Volk, editors, *Proceedings of the 5th Swiss Text Analytics Conference and the 16th Conference on Natural Language Processing, SwissText/KONVENS 2020, Zurich, Switzerland, June 23-25, 2020 [online only]*, volume 2624 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2020. URL <http://ceur-ws.org/Vol-2624/paper9.pdf>

Fake Reviews Detection

- Internet is full of fake reviews of products, restaurants, ...
- Humans are not particularly good in detecting what reviews are fake
- Can we do it automatically?
- Can fake reviews be automatically generated?



The screenshot shows a product listing for "100% Genuine Snake Oil" by "Scammer's Warehouse". The product is priced at \$89.70 and has a Prime badge. It features a 5-star rating from 42 customer reviews. A review by "Fake Jim" from June 19, 2017, is displayed, stating: "AMAZING healing qualities. Very good product. I can't prove this for certain, but I think it cured my cancer. I feel like I'm 17 again." The review is accompanied by a 5-star rating. The product image shows a bottle of snake oil. The Hustle logo is visible in the bottom right corner of the screenshot.

100% Genuine Snake Oil
By: [Scammer's Warehouse](#)
★★★★★ 42 customer reviews
Price: **\$89.70** ✓ Prime

★★★★★ **AMAZING** healing qualities
By: [Fake Jim](#) on June 19, 2017
Item: Snake oil, 4 oz.
Very good product. I can't prove this for certain, but I think it cured my cancer. I feel like I'm 17 again.

HUSTLE

Fake Reviews Detection: Papers

- A. Mukherjee, V. Venkataraman, B. Liu, N. Glance, et al. **Fake review detection: Classification and analysis of real and pseudo reviews**. *Technical Report UIC-CS-2013-03, University of Illinois at Chicago, Tech. Rep.*, 2013
- S. Kennedy, N. Walsh, K. Sloka, A. McCarren, and J. Foster. **Fact or factitious? contextualized opinion spam detection**. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 344–350, Florence, Italy, July 2019
- D. I. Adelani, H. Mai, F. Fang, H. H. Nguyen, J. Yamagishi, and I. Echizen. **Generating sentiment-preserving fake online reviews using neural language models and their human- and machine-based detection**. In L. Barolli, F. Amato, F. Moscato, T. Enokido, and M. Takizawa, editors, *Advanced Information Networking and Applications - Proceedings of the 34th International Conference on Advanced Information Networking and Applications, AINA-2020, Caserta, Italy, 15-17 April*, volume 1151 of *Advances in Intelligent Systems and Computing*, pages 1341–1354. Springer, 2020

Depressions in Online Forums

- People write various things on Reddit and other fora—e.g., that they depressed or have suicidal thoughts
- Can we use this data somehow to help people?
- Can we predict user will suffer with depression before he/she knows?

Depressions in Online Forums: Papers

- A. Yates, A. Cohan, and N. Goharian. **Depression and self-harm risk assessment in online forums.** In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2968–2978, Copenhagen, Denmark, Sept. 2017. Association for Computational Linguistics. doi: [10.18653/v1/D17-1322](https://doi.org/10.18653/v1/D17-1322). URL <https://www.aclweb.org/anthology/D17-1322>
- H.-C. Shing, S. Nair, A. Zirikly, M. Friedenberg, H. Daumé III, and P. Resnik. **Expert, crowdsourced, and machine assessment of suicide risk via online postings.** In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 25–36, New Orleans, LA, June 2018. Association for Computational Linguistics. doi: [10.18653/v1/W18-0603](https://doi.org/10.18653/v1/W18-0603). URL <https://www.aclweb.org/anthology/W18-0603>
- J. Ive, G. Gkotsis, R. Dutta, R. Stewart, and S. Velupillai. **Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health.** In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 69–77, New Orleans, LA, June 2018. Association for Computational Linguistics. doi: [10.18653/v1/W18-0607](https://doi.org/10.18653/v1/W18-0607). URL <https://www.aclweb.org/anthology/W18-0607>