Information Extraction Topics WiSe 2021/22

Katharina Hämmerl
haemmerl@cis.lmu.de

CIS, LMU

03 November 2021

Table of Contents

Table of Contents

Clinical Trial Eligibility Criteria Extraction

- Clinical trial descriptions usually include patient eligibility criteria—important for evaluating studies

Clinical Trial Eligibility Criteria Extraction

- Clinical trial descriptions usually include patient eligibility criteria—important for evaluating studies
- For large survey studies, patient matching, etc.: want computer-readable criteria

Clinical Trial Eligibility Criteria Extraction

- Tu, Samson W. et al. "A practical method for transforming free-text eligibility criteria into computable criteria." *Journal of Biomedical Informatics* 44 2 (2011): 239-50.
- Kang, Tian et al. "ElilE: An open-source information extraction system for clinical trial eligibility criteria." *Journal of the American Medical Informatics Association* 24 (2017): 1062–1071.
- Liu, Xiong et al. "Attention-Based LSTM Network for COVID-19 Clinical Trial Parsing." *2020 IEEE International Conference on Big Data (Big Data)* (2020): 3761-3766.

Table of Contents

Canonicalization of Open Knowledge Bases

- Open knowledge bases can be extracted from web text e.g. as triples of noun phrase, verb phrase, noun phrase

Canonicalization of Open Knowledge Bases

- Open knowledge bases can be extracted from web text e.g. as triples of noun phrase, verb phrase, noun phrase
- Multiple NPs can mean same entity; multiple VPs can mean same relation

Canonicalization of Open Knowledge Bases

- Open knowledge bases can be extracted from web text e.g. as triples of noun phrase, verb phrase, noun phrase
- Multiple NPs can mean same entity; multiple VPs can mean same relation
- Canonicalization: Recognizing such synonyms and clustering them together

Canonicalization of Open Knowledge Bases

- Open knowledge bases can be extracted from web text e.g. as triples of noun phrase, verb phrase, noun phrase
- Multiple NPs can mean same entity; multiple VPs can mean same relation
- Canonicalization: Recognizing such synonyms and clustering them together

Example

"Obama was born in Honolulu" == "Barack Obamas birthplace is Honolulu"

Canonicalization of Open Knowledge Bases

- Luis Galárraga, Geremy Heitz, Kevin Murphy, and Fabian M. Suchanek. 2014. Canonicalizing Open Knowledge Bases. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM '14)*.
- Shikhar Vashishth, Prince Jain, and Partha Talukdar. 2018. CESI: Canonicalizing Open Knowledge Bases using Embeddings and Side Information. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*.
- Yinan Liu, Wei Shen, Yuanfei Wang, Jianyong Wang, Zhenglu Yang, and Xiaojie Yuan. 2021. Joint Open Knowledge Base Canonicalization and Linking. In *Proceedings of the 2021 International Conference on Management of Data (SIGMOD/PODS '21)*.

Table of Contents

Song Lyric Segmentation

- Song lyrics are typically available as plain text

Song Lyric Segmentation

- Song lyrics are typically available as plain text
- Lyrics websites, streaming services want to analyze them along with songs

Song Lyric Segmentation

- Song lyrics are typically available as plain text
- Lyrics websites, streaming services want to analyze them along with songs
- Finding the structural parts as one of the first steps

Song Lyric Segmentation

- Baratè, Adriano et al. "A Semantics-Driven Approach to Lyrics Segmentation." *2013 8th International Workshop on Semantic and Social Media Adaptation and Personalization* (2013): 73-79.
- Fell, Michael et al. "Lyrics Segmentation: Textual Macrostructure Detection using Convolutions." *COLING* (2018).
- Watanabe, Kento et al. "Modeling Discourse Segments in Lyrics Using Repeated Patterns." *COLING* (2016).

Table of Contents

Human-in-the-Loop IE

- IE tasks can be highly complex

Human-in-the-Loop IE

- IE tasks can be highly complex
- Especially if you want to personalize, involve social aspects

Human-in-the-Loop IE

- IE tasks can be highly complex
- Especially if you want to personalize, involve social aspects
- "Human-in-the-loop": User collaborates with system; system (ideally) learns from feedback

Human-in-the-loop IE

- Constantin Orăsan and Laura Hasler. 2006. Computer-aided summarisation – what the user really wants. In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*
- Beheshti, Amin et al. 2018. CrowdCorrect: A Curation Pipeline for Social Data Cleansing and Curation. In: Mendling J., Mouratidis H. (eds) *Information Systems in the Big Data Era. CAiSE 2018. Lecture Notes in Business Information Processing*, vol 317.
- Ghodratnama, Samira et al. 2020. Adaptive Summaries: A Personalized Concept-based Summarization Approach by Learning from Users' Feedback. *arXiv cs.AI* 2012.13387