

Information Extraction Topics WiSe 2022/23

Katharina Hämmerl
haemmerl@cis.lmu.de

CIS, LMU

09 November 2022

Table of Contents

- ① Canonicalisation of Open Knowledge Bases
- ② Song Lyric Segmentation
- ③ Human-in-the-Loop Summarisation

- ① Canonicalisation of Open Knowledge Bases
- ② Song Lyric Segmentation
- ③ Human-in-the-Loop Summarisation

Canonicalisation of Open Knowledge Bases

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

Canonicalisation 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

Canonicalisation 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
- Canonicalisation: Recognizing such synonyms and clustering them together

Canonicalisation 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
- Canonicalisation: Recognizing such synonyms and clustering them together

Example

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

Canonicalisation of Open Knowledge Bases

- Luis Galárraga, Jeremy 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)*.
- Wei Shen, Yang Yang, and Yinan Liu. 2022. Multi-View Clustering for Open Knowledge Base Canonicalization. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

Table of Contents

- ① Canonicalisation of Open Knowledge Bases
- ② Song Lyric Segmentation
- ③ Human-in-the-Loop Summarisation

Song Lyric Segmentation

- A lot of work analysing music focuses on the audio, but song lyrics are often available as plain text

Song Lyric Segmentation

- A lot of work analysing music focuses on the audio, but song lyrics are often available as plain text
- Lyrics websites or streaming services want to analyse them along with the songs

Song Lyric Segmentation

- A lot of work analysing music focuses on the audio, but song lyrics are often available as plain text
- Lyrics websites or streaming services want to analyse them along with the songs
- Finding the structural parts as one of the first steps

Song Lyric Segmentation

- Heng Tze Cheng, Yi-Hsuan Yang, Yu-Ching Lin, and Homer H. Chen. 2009. Multimodal structure segmentation and analysis of music using audio and textual information. In *2009 IEEE International Symposium on Circuits and Systems*.
- Michael Fell, Yaroslav Nechaev, Elena Cabrio, and Fabien Gandon. 2018. Lyrics Segmentation: Textual Macrostructure Detection using Convolutions. In *Proceedings of the 27th International Conference on Computational Linguistics*.
- Kento Watanabe, Masataka Goto. 2020. A Chorus-Section Detection Method for Lyrics Text. In *Proceedings of the 21st International Society for Music Information Retrieval Conference*.

Table of Contents

- ① Canonicalisation of Open Knowledge Bases
- ② Song Lyric Segmentation
- ③ Human-in-the-Loop Summarisation

- Summarisation is a complex information extraction task

Human-in-the-Loop Summarisation

- Summarisation is a complex information extraction task
- Especially if you want to personalize, involve social aspects

Human-in-the-Loop Summarisation

- Summarisation is a complex information extraction task
- Especially if you want to personalize, involve social aspects
- "Human-in-the-loop": User collaborates with system; system should learn from feedback

Human-in-the-loop Summarisation

- 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)*.
- Avinesh P.V.S and Christian M. Meyer. 2017. Joint Optimization of User-desired Content in Multi-document Summaries by Learning from User Feedback. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.
- Duy-Hung Nguyen, Nguyen Viet Dung Nghiem, Bao-Sinh Nguyen, Dung Tien Tien Le, Shahab Sabahi, Minh-Tien Nguyen, and Hung Le. 2022. Make The Most of Prior Data: A Solution for Interactive Text Summarization with Preference Feedback. In *Findings of the Association for Computational Linguistics: NAACL 2022*.