From Fundamentals to Recent Advances
A Tutorial on Keyphrasification

Part 4 Closing Remarks

Rui Meng, Debanjan Mahata, Florian Boudin

ECIR 2022

Challenges in Keyphrasification

- Challenges in Understanding
  - Understand the human preference in keyphrases in different context
    - Why noun phrases are dominant in keyphrases?
    - When verb/adjective phrases can be preferred?
    - Why some phrases are more representative than others?
  - Lack of systematic categorization of keyphrases
    - present v.s absent (by appearance)
    - specific v.s topic/categorical (by specificity)
    - in-domain v.s out-of-domain (by domain relatedness)
    - etc.
  - Proper guidelines for annotating keyphrases
  - Structured relationship between phrases and documents
    - Scientific paper: method, metric, dataset etc.
    - News article: 5W1H
    - etc.

Challenges in Keyphrasification

- Challenges in Modeling
  - Model generalization (domains, annotation types)
  - Reduce the reliance on data annotation
    - Zero-/ few-shot keyphrase generation
  - Models for processing very short/long documents
  - Improving decoding strategies for keyphrase generation
    - More diverse, efficient, order-invariant
  - Controllable keyphrase generation
  - Multi-modal keyphrase identification
  - Representation learning of phrases

Challenges in Keyphrasification

- Challenges in Evaluation
  - Clearer definition and metrics for measuring the phraseness and keyness properties
  - Evaluating keyphrases through applications
- Challenges in Datasets
  - Few datasets in non-English Languages
  - Few datasets in domains other than scientific texts
  - Few high-quality (not authors or readers) annotated datasets
- Needs in exploiting its value in more applications

Thank you!

From Fundamentals to Recent Advances
A Tutorial on Keyphrasification

All materials available at
https://keyphrasification.github.io/

Q/A session