From Fundamentals to Recent Advances A Tutorial on Keyphrasification

Part 1.2 Evaluation and Benchmarking

Rui Meng, Debanjan Mahata, Florian Boudin

ECIR 2022

Evaluation Strategies

- Exact Match Evaluation - *most popular*
- Partial Match Evaluation
- Downstream Evaluation
- Human Evaluation

Automated

Evaluation Metrics

- Precision@K, Recall@K, F1@K
  - Most widely used metric in traditional methods
- Precision, Recall and F1
  - Most popularly used in evaluating keyphrase extraction as sequence tagging
- Precision@O/M, Recall@O/M, F1@O/M
  - Mostly used for evaluating keyphrase generation

Sample Example

Input Text - The development of a mobile manipulator imaging system for bridge crack inspection. A mobile manipulator imaging system is developed for the automation of bridge crack inspection. During bridge safety inspections, an eyesight inspection is made for preliminary evaluation and screening before a more precise inspection. The inspection for cracks is an important part of the preliminary evaluation . Currently, the inspectors must stand on the platform of a bridge inspection vehicle or a temporarily erected scaffolding to examine the underside of a bridge. However, such a procedure is risky. To help automate the bridge crack inspection process, we installed two CCD cameras and a four-axis manipulator system on a mobile vehicle. The parallel cameras are used to detect cracks. The manipulator system is equipped with binocular charge coupled devices for examining structures that may not be accessible to the eye. The system also reduces the danger of accidents to the human inspectors. The manipulator system consists of four arms. Balance weights are placed at the ends of arms 2 and 4, respectively, to maintain the center of gravity during operation. Mechanically, arms 2 and 4 can revolve smoothly. Experiments indicated that the system could be useful for bridge crack inspections.

Gold Keyphrases (Y_{gold})

'mobile manipulator', 'imaging system', 'bridge crack inspection', 'automation', 'eyesight inspection', 'ccd cameras', 'parallel cameras', 'charge coupled devices', 'four-axis manipulator'

Predicted Keyphrases (Y_{pred})

'bridge crack inspection process', 'bridge inspection vehicle', 'bridge safety inspections', 'bridge crack inspection', 'mobile manipulator imaging system', 'precise inspection', 'eyesight inspection', 'four-axis manipulator system', 'inspection', 'manipulator system'

Gold keyphrases

Precision, Recall, F1@K

Precision, Recall, F1@K

$$precision@K = \frac{top\ K\ in\ Y_{pred} \cap Y_{gold}}{K}$$

$$recall@K = \frac{top\ K\ in\ Y_{pred} \cap Y_{gold}}{|Y_{gold}|}$$

$$F1@K = 2 \times \frac{precision@K \times recall@K}{precision@K + recall@K}$$

Gold Keyphrases (Y_{gold})

'mobile manipulator', 'imaging system', 'bridge crack inspection', 'automation', 'eyesight inspection', 'ccd cameras', 'parallel cameras', 'charge coupled devices', 'four-axis manipulator'

Predicted Keyphrases (Y_{pred})

['bridge crack inspection process', 'bridge inspection vehicle', 'bridge safety inspections', 'bridge crack inspection', 'mobile manipulator imaging system'], 'precise inspection', 'eyesight inspection', 'four-axis manipulator system', 'inspection', 'manipulator system'

K = 5, **Precision@5** = 1 / 5 = 0.2, **Recall@5** = 1 / 9 = 0.11,

F1@5 = 2 * (0.2 * 0.11) / (0.2 + 0.11) = 0.14

What should be the right @K?

- High variance in number of ground truth phrases
  - often $|Y_{pred}| \leq K < |Y_{gold}|$
  - **Recall@K** and **F1@K** can then never be 1

- Simple solution
  - Set K as a variable number
  - Depends on the specific data example

Precision, Recall, F1@M

precision, recall, F1@M

$$precision@M = \frac{top\ M\ in\ Y_{pred} \cap Y_{gold}}{M}$$

$$recall@M = \frac{top\ M\ in\ Y_{pred} \cap Y_{gold}}{|Y_{gold}|}$$

$$F1@M = 2 \times \frac{precision@M \times recall@M}{precision@M + recall@M}$$

M = No. of Predicted Keyphrases

Gold Keyphrases (Y_{gold})

'mobile manipulator', 'imaging system', 'bridge crack inspection', 'automation', 'eyesight inspection', 'ccd cameras', 'parallel cameras', 'charge coupled devices', 'four-axis manipulator'

Predicted Keyphrases (Y_{pred})

['bridge crack inspection process', 'bridge inspection vehicle', 'bridge safety inspections', 'bridge crack inspection', 'mobile manipulator imaging system', 'precise inspection', 'eyesight inspection', 'four-axis manipulator system', 'inspection', 'manipulator system', 'parallel cameras', 'ccd cameras', 'balance weights', 'human inspectors']

M=15, Precision@M = 4 / 15 = 0.26, **Recall@M** = 4 / 9 = 0.44,

F1@M = 2 * (0.26 * 0.44) / (0.26 + 0.44) = 0.3268

Yuan, X., Wang, T., Meng, R., Thaker, K., Brusilovsky, P., He, D., & Trischler, A. (2018). One size does not fit all: Generating and evaluating variable number of keyphrases. *arXiv preprint arXiv:1810.05241*.

Precision, Recall, F1@O

precision, recall, F1@O

$$precision@O = \frac{top\ O\ in\ Y_{pred} \cap Y_{gold}}{O}$$

$$recall@O = \frac{top\ O\ in\ Y_{pred} \cap Y_{gold}}{O}$$

$$F1@O = 2 \times \frac{precision@O \times recall@O}{precision@O + recall@O}$$

O = No. of Gold Keyphrases

Gold Keyphrases (Y_{gold})

'mobile manipulator', 'imaging system', 'bridge crack inspection', 'automation', 'eyesight inspection', 'ccd cameras', 'parallel cameras', 'charge coupled devices', 'four-axis manipulator'

Predicted Keyphrases (Y_{pred})

['bridge crack inspection process', 'bridge inspection vehicle', 'bridge safety inspections', 'bridge crack inspection', 'mobile manipulator imaging system', 'precise inspection', 'eyesight inspection', 'four-axis manipulator system', 'inspection'], 'manipulator system', 'parallel cameras', 'ccd cameras', 'balance weights', 'human inspectors'

O = 9, **Precision@O** = 2 / 9 = 0.22, **Recall@O** = 2 / 9 = 0.22,

F1@O = 2 * (0.22 * 0.22) / (0.22 + 0.22) = 0.22

Yuan, X., Wang, T., Meng, R., Thaker, K., Brusilovsky, P., He, D., & Trischler, A. (2018). One size does not fit all: Generating and evaluating variable number of keyphrases. *arXiv preprint arXiv:1810.05241*.

Precision, Recall, F1@O

precision, recall, F1@O

$$precision@O = \frac{top\ O\ in\ Y_{pred}\ \cap\ Y_{gold}}{O}$$

$$recall@O = \frac{top\ O\ in\ Y_{pred}\ \cap\ Y_{gold}}{O}$$

$$F1@O = 2 \times \frac{precision@O\ \times\ recall@O}{precision@O\ +\ recall@O}$$

O = No. of Gold Keyphrases

Gold Keyphrases (κ_{gold})

'mobile manipulator', 'imaging system', 'bridge crack inspection', 'automation', 'eyesight inspection', 'ccd cameras', 'parallel cameras', 'charge coupled devices', 'four-axis manipulator'

Predicted Keyphrases (κ_{pred})

not considered

'bridge crack inspection process', 'bridge inspection vehicle', 'bridge safety inspections', 'bridge crack inspection', 'mobile manipulator imaging system', 'precise inspection', 'eyesight inspection', 'four-axis manipulator system', 'inspection', 'manipulator system', 'parallel cameras', 'ccd cameras', 'balance weights', 'human inspectors'

O = 9, Precision@O = 2 / 9 = 0.22, **Recall@O** = 2 / 9 = 0.22,

F1@O = 2 * (0.22 * 0.22) / (0.22 + 0.22) = 0.22

Yuan, X., Wang, T., Meng, R., Thaker, K., Brusilovsky, P., He, D., & Trischler, A. (2018). One size does not fit all: Generating and evaluating variable number of keyphrases. *arXiv preprint arXiv:1810.05241*.

F1@M and F1@O

- **M** - no. of predicted keyphrases
- **O** - no. of ground truth keyphrases
- **F1@M = F1@O**
  - no. of predicted keyphrases = no. of ground truth keyphrases
- Intuitively, @M/@O penalizes the cases where the number of predicted phrases and gold phrases mismatch
- a model predicts many bad phrases - poor F1@M
- a very conservative model predicts few good phrases - poor F1@O
- accounts for the variable number of phrases for each document and now models can achieve the maximum F1 score of 1

Precision, Recall, F1

precision, recall, F1

$$precision = \frac{Y_{pred} \cap Y_{gold}}{|Y_{pred}|}$$

$$recall = \frac{Y_{pred} \cap Y_{gold}}{|Y_{gold}|}$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$

Gold Keyphrases (κ_{gold})

'mobile manipulator', 'imaging system', 'bridge crack inspection', 'automation', 'eyesight inspection', 'ccd cameras', 'parallel cameras', 'charge coupled devices', 'four-axis manipulator'

Top 10 Extracted Keyphrases (κ_{pred})

'bridge crack inspection process', 'bridge inspection vehicle', 'bridge safety inspections', 'bridge crack inspection', 'mobile manipulator imaging system', 'precise inspection', 'eyesight inspection', 'four-axis manipulator system', 'inspection', 'manipulator system'

Same as F1@M

Precision = 2 / 10 = 0.2, **Recall** = 2 / 9 = 0.22,

F1 = 2 * (0.2 * 0.22) / (0.2 + 0.22) = 0.20

Gold Keyphrases

Used mainly in keyphrase extraction as sequence labeling task

Actual Matching Issues

Input Text - The development of a mobile manipulator imaging system for bridge crack inspection. A mobile manipulator imaging system is developed for the automation of bridge crack inspection. During bridge safety inspections, an eyesight inspection is made for preliminary evaluation and screening before a more precise inspection. The inspection for cracks is an important part of the preliminary evaluation . Currently, the inspectors must stand on the platform of a bridge inspection vehicle or a temporarily erected scaffolding to examine the underside of a bridge. However, such a procedure is risky. To help automate the bridge crack inspection process, we installed two CCD cameras and a four-axis manipulator system on a mobile vehicle. The parallel cameras are used to detect cracks. The manipulator system is equipped with binocular charge coupled devices for examining structures that may not be accessible to the eye. The system also reduces the danger of accidents to the human inspectors. The manipulator system consists of four arms. Balance weights are placed at the ends of arms 2 and 4, respectively, to maintain the center of gravity during operation. Mechanically, arms 2 and 4 can revolve smoothly. Experiments indicated that the system could be useful for bridge crack inspections.

Actual Matching

Gold Keyphrases (κ_{gold})

'mobile manipulator', 'imaging system', 'bridge crack inspection', 'automation', 'eyesight inspection', 'ccd cameras', 'parallel cameras', 'charge coupled devices', 'four-axis manipulator'

Predicted Keyphrases (κ_{pred})

'bridge crack inspection process', 'bridge inspection vehicle', 'bridge safety inspections', 'bridge crack inspection', 'mobile manipulator imaging system', 'precise inspection', 'eyesight inspection', 'four-axis manipulator system', 'inspection', 'manipulator system', 'parallel camera'

semantic variations overlapping phrases morphological variants → **Stemming**

Partial Matching

Gold Keyphrases (κ_{gold})

'mobile manipulator', 'imaging system', 'bridge crack inspection', 'automation', 'eyesight inspection', 'ccd cameras', 'parallel cameras', 'charge coupled devices', 'four-axis manipulator'

Predicted Keyphrases (κ_{pred})

'bridge crack inspection process', 'bridge inspection vehicle', 'bridge safety inspections', 'bridge crack inspection', 'mobile manipulator imaging system', 'precise inspection', 'eyesight inspection', 'four-axis manipulator system', 'inspection', 'manipulator system', 'parallel camera'

semantic variations overlapping phrases morphological variants

INCLUDE - extracted keyphrase includes the gold keyphrase

PARTOF - extracted keyphrase is a part of gold keyphrase

MORPH - extracted and gold keyphrases are morphological variants

Zesch, T., & Gurevych, I. (2009, September). Approximate matching for evaluating keyphrase extraction. In *Proceedings of the International Conference RANLP-2009* (pp. 484-489).

R-precision

- Precision when the number of predicted keyphrases equals the number of gold keyphrases $|\kappa_{pred}| = |\kappa_{gold}|$
- Puts focus on the top keyphrases predicted by the system

$$R - p = \frac{no.\ of\ predicted\ keyphrases\ actually\ or\ partially\ matching\ gold\ keyphrases}{|\kappa_{gold}|}$$

	#	Judges accepting matchings 4	≥ 3
INCLUDES	274	.58	.80
PARTOF	239	.31	.44
MORPH	53	.96	.96
MORPH+INCLUDES	327	.65	.83

Table 2: *Ratio of approximate keyphrase matchings acceptable to human judges (4 = all judges; ≥ 3 = at least 3 out of 4 judges).*

Zesch, T., & Gurevych, I. (2009, September). Approximate matching for evaluating keyphrase extraction. In *Proceedings of the International Conference RANLP-2009* (pp. 484-489).

Other Evaluation Metrics

- Information Retrieval Metrics
  - Mean Average Precision (Jiang et al.)
  - Mean Reciprocal Rank (Liu et al.)
  - Binary Preference Measure (Liu et al.)

- Summarization Metrics
  - BLEU, METEOR, NIST, ROUGE (Kim et al.)
    - Useful for handling near-misses

Jiang, X., Hu, Y., & Li, H. (2009, July). A ranking approach to keyphrase extraction. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 756-757).
Kim, S. N., Baldwin, T., & Kan, M. Y. (2010, August). Evaluating n-gram based evaluation metrics for automatic keyphrase extraction. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)* (pp. 572-580).
Liu, Z., Huang, W., Zheng, Y., & Sun, M. (2010, October). Automatic keyphrase extraction via topic decomposition. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 366-376).

Evaluation Errors

- Overgeneration errors
- Infrequency errors
- Redundancy errors
- Evaluation errors

Canadian **Ben Johnson** left the **Olympics** today "in a complete state of shock," accused of cheating with drugs in the world's fastest **100-meter dash** and stripped of his **gold medal**. The prize went to American **Carl Lewis**. Many athletes accepted the accusation that Johnson used a muscle-building but dangerous and illegal anabolic steroid called **stanozolol** as confirmation of what they said they know has been going on in track and field. Two tests of Johnson's urine sample proved positive and his denials of **drug use** were rejected today. "This is a blow for the Olympic Games and the Olympic movement," said International Olympic Committee President Juan Antonio Samaranch.

Hasan, K. S., & Ng, V. (2014, June). Automatic keyphrase extraction: A survey of the state of the art. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1262-1273).

Datasets

Scientific Documents

News Documents

Web

Annotation Process

Author Assigned

Reader Assigned

Professional Indexer or Editor Assigned

Annotation Guidelines

Keyphrases?
The keywords or 'keyphrases' are defined as "a selection of short, significant expressions consisting of one or more words that can summarize the article very compactly."

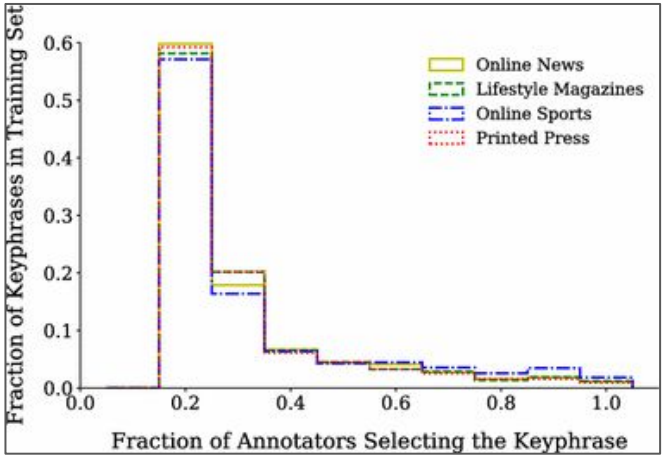
Method
There are some tips to get to the best keywords:
Ask yourself, "What words summarize the content of the article?" Or "What words are most representative of the contents of the article?" This can be an event or an object, the crucial entities, or organizations that are mentioned in the article. Try to keep the keyphrase as short as possible. Words that do not contribute may be omitted to the meaning of the keyphrases. The number of keywords per article depends largely on the length of the article and the various topics discussed in it. It is rare to select more than 10 keywords per article.

We demonstrate this with an example:
"Higher education is bracing itself. Once it had ample offer, now it calls for a economization of supply. The Flemish coalition agrees that the universities themselves must make proposals to achieve a constrained and transparent offer. In interviews, the Minister of Education, Hilde Crevits (ISA), indicates that the offer can be safely pruned to one hundred of majors. "
in this example "higher education" , "economization of supply" and " Hilde Crevits" would be appropriate keywords.

Sterckx, L., Demeester, T., Deleu, J., & Develder, C. (2018). Creation and evaluation of large keyphrase extraction collections with multiple opinions. *Language Resources and Evaluation*, *52*(2), 503-532.

Inter-annotator Agreements

In 2013 **Anastacia**[1,2,3,4,5,6,7,8,9,10](45) was struck for the second time with **breast cancer**[1,2,3,4,5,6,7,8,9,10]. A difficult recovery, including a double **mastectomy**[4] and reconstruction followed. She gave her **first concert**[2] in Belgium after this dark period, with her best friend **Natalia**[3,4,8,9,10] in the front row. In January 2013 Anastacia was known for world hits like '**I'm outta love**'[3] and '**Left Outside Alone**'[3]. Busy working on a new album, her doctor called with bad news. The lump she felt in her breast was cancerous. Ten years earlier she had already been treated for breast cancer, but the tumor appeared stronger. Anastacia underwent a **double mastectomy**[10] and reconstruction, but now feels "great again". "I've learned love to see myself," she says. "I am happy, you see. And I let myself be carried away by my work easily. Now I try to pay more attention to myself. Absolutely. I want to make people happy with my songs. 'I'm Outta Love', for example, I still sing that with pleasure. While other artists are often tired of their first songs. Take Madonna. Who said herself that she can't hear 'Like A Virgin', let alone sing it herself. That's right, and even the title ('**Resurrection**[7]', ed) seems to refer to it. But this was already determined before I was told I had cancer again. I am resurrected in different areas, so to speak. I'm more balanced, a lot calmer … Yes, but the result of the **breast reconstruction**[5] is fantastic. I'm not ashamed of my scars. I still feel sexy. Such a **mastectomy**[8] is not a fairy tale, huh. You must accept the reality. There are a lot of emotions involved, and you have to beat you through. And my **first concert**[3,10] is in **Belgium**[2] On October 19, in the AB in Brussels I can finally do what I love to do: act and sing. Of course, Natalia is one of my best friends. She has supported me the whole recovery period. We call regularly. No, but I'm not searching. Love has always happened to me. Which does not mean that true love may turn up little by little. But I'm not desperate. I take life as it comes and I am grateful for what I've got."



Sterckx, L., Demeester, T., Deleu, J., & Develder, C. (2018). Creation and evaluation of large keyphrase extraction collections with multiple opinions. *Language Resources and Evaluation*, *52*(2), 503-532.

Inter-annotator Agreements

Student Annotator	IAA
1	0.85
2	0.66
3	0.63
4	0.60
5	0.50
6	0.48
7	0.47
8	0.45
9*	0.25
10*	0.22
11*	0.20
12*	0.15
13*	0.06

Inter-annotator agreement between the student annotator and the expert annotator, measured with Cohen's Kappa [1].

Judge Depth	Exact Match	Unigram Match
Keyphrase@1	64.74%	64.74%
Keyphrase@2	48.30%	63.12%
Keyphrase@3	43.51%	57.66%

The agreements between pairs of expert judges at different annotation depth. Exact and Unigram show the percentage of judge agreement on exact keyphrases and overlapped unigrams [2].

[1] Augenstein, I., Das, M., Riedel, S., Vikraman, L., & McCallum, A. (2017). Semeval 2017 task 10: Scienceie-extracting keyphrases and relations from scientific publications. *arXiv preprint arXiv:1704.02853*.

[2] Xiong, L., Hu, C., Xiong, C., Campos, D., & Overwijk, A. (2019). Open domain web keyphrase extraction beyond language modeling. *arXiv preprint arXiv:1911.02671*.

Data Annotation Challenges

- Annotation guidelines are problem and domain dependent
- No standard guidelines

Author Assigned	Reader Assigned	Indexer/Editor Assigned
• Subjective • Intent bias • Abstractive • Exhaustive	• Subjective • Lack of expertise • Extractive • Specific • Low agreement	• Subjective • Controlled • Depends on thesaurus quality and the controlled vocabulary

Popular Benchmarks in Scientific Domain

Abstracts

SCIENCE	Annotator	No. of Docs	Avg. Length (in words)	Avg. # of Keyphrases
Inspec	I	Train - 1,000 Dev - 500 Test - 500	134.6	9.8
KDD	A	Test - 755	190.7	4.1
WWW	A	Test - 1,330	163.5	4.8
KP20K	A	Train - 527,090 Dev - 50,000 Test - 50,000	176	5.3
OAGKx	A	23 million	175.1	5.9
SemEval 2017	I + R	Train - 350, Dev - 50, Test - 100	NA	NA

Full Texts

SCIENCE	Annotator	No. of Docs	Avg. Length (in words)	Avg. # of Keyphrases
CSTR	A	Train - 130 Test - 500	11,501.4	5.4
NUS	A + R	Test - 211	8398.3	11
PubMed	A	Test - 1320	5322.9	5.4
ACM	A	Test - 2304	9197.6	5.3
CiteULike	R	Test - 182	8589.7	5.4
SemEval 2010	A + R	Train - 144 Test - 100	7961.2	14.7

A - Author assigned, **R** - Reader assigned, **I** - Professional indexer assigned

Sources: 1. https://github.com/boudinfl/ake-datasets 2. https://huggingface.co/midas

Popular Benchmarks in News and Web Domains

	Annotator	No. of Docs	Avg. Length (in words)	Avg. # of Keyphrases
KPTimes	E	Train - 259,923 Dev - 10,000 Test - 20,000	921	5
DUC-2001	R	Test - 308	847.2	8.1
500N-KPCrowd	R	Train - 450 Test - 50	465.3	46.2
110-PT-BN-KP (Portuguese)	R	Train - 100 Test - 10	439.4	27.6
Wikinews-Keyphrase (French)	R	Test - 100	313.6	9.7

- **OpenKP** - consists of documents from web indexed by the BING search engine.
  - Train - 134,894
  - Dev - 6,616
  - Test - 6,614
  - Reader assigned keyphrases
  - Avg. Length - 900.4
  - Avg. # of Keyphrases - 1.8

A - Author assigned, **R** - Reader assigned, **E** - Professional editor assigned

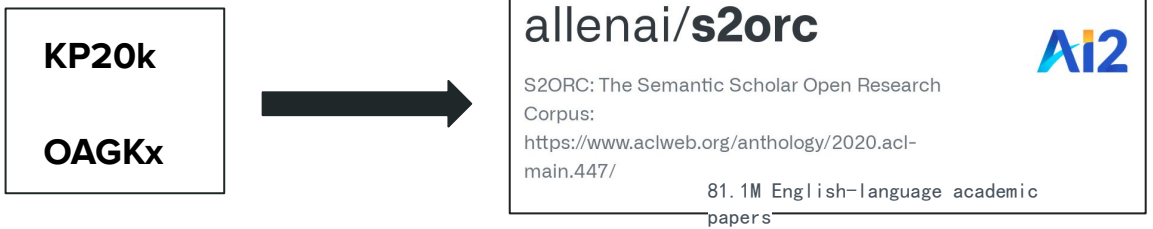
Sources: 1. https://github.com/boudinfl/ake-datasets 2. https://huggingface.co/midas

Dataset Challenges

- Language coverage
- Domain coverage
- Lack of large dataset on long documents
- Dataset for performing keyphrase extraction in the wild

LDKP - A Dataset of Long Documents

Dataset	Size	Long Doc	Avg # Sentences	Avg # Words	Present KPs	Absent KPs
SemEval2017 (Augenstein et al., 2017)	0.5k	×	7.36	176.13	42.01%	57.69%
KDD (Caragea et al., 2014)	0.75k	×	8.05	188.43	45.99%	54.01%
Inspec (Hulth, 2003)	2k	×	5.45	130.57	55.69%	44.31%
KP20k (Meng et al., 2017)	568k	×	7.42	188.47	57.4%	42.6%
OAGKx (Çano, 2019)	22M	×	8.87	228.50	52.7%	47.3%
NUS (Nguyen and Kan, 2007)	0.21k	✓	375.93	7644.43	67.75%	32.25%
SemEval2010 (Kim et al., 2010)	0.24k	✓	319.32	7434.52	42.01%	57.99%
Krapivin (Krapivin et al., 2010)	2.3k	✓	370.48	8420.76	44.74%	52.26%
LDKP3K (S2ORC ← KP20K)	**100k**	✓	280.67	6027.10	76.11%	23.89%
LDKP10K (S2ORC ← OAGKx)	**1.3M**	✓	194.76	4384.58	63.65%	36.35%



Debanjan Mahata, Naveen Agarwal, Dibya Gautam, Amardeep Kumar, Swapnil Parekh, Yaman Kumar Singla, Anish Acharya, Rajiv Ratn Shah. LDKP: A Dataset for Identifying Keyphrases from Long Scientific Documents. https://arxiv.org/abs/2203.15349

LDKP - A Dataset of Long Documents

Data Stats

Dataset		LDKP3K	LDKP10K
Train	Small	20,000	20,000
	Medium	50,000	50,000
	Large	90,019	1,296,613
Test		3,413	10,000
Validation		3,339	10,000

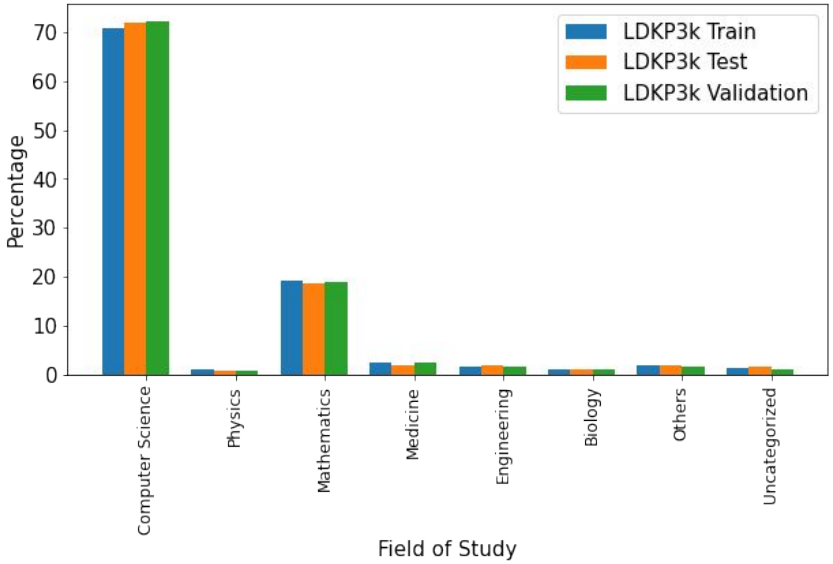
Metadata

Paper details	Paper Identifier	Citations and References
Paper ID	ArXiv ID	Outbound Citations
Title	ACL ID	Inbound Citations
Authors	PMC ID	Bibliography
Year	PubMed ID	References
Venue	MAG ID	
Journal	DOI	
Field of Study	S2 URL	

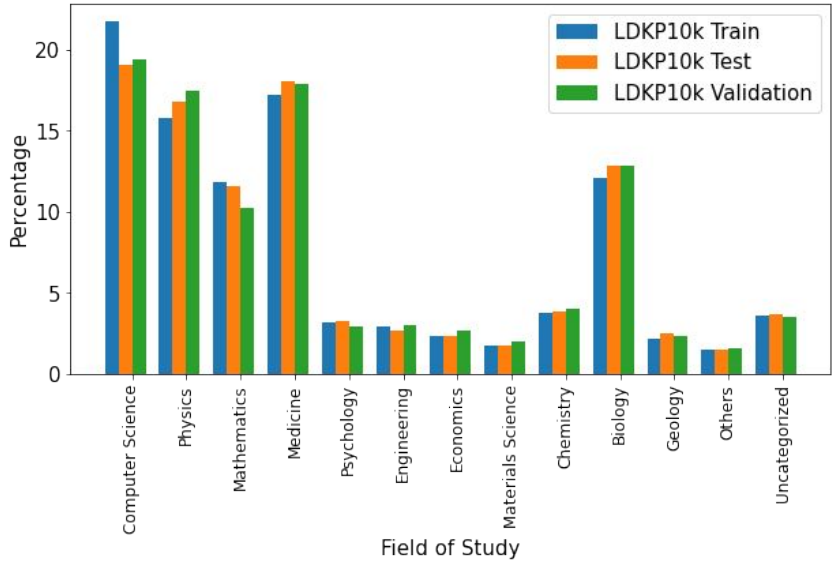
Debanjan Mahata, Naveen Agarwal, Dibya Gautam, Amardeep Kumar, Swapnil Parekh, Yaman Kumar Singla, Anish Acharya, Rajiv Ratn Shah. LDKP: A Dataset for Identifying Keyphrases from Long Scientific Documents. https://arxiv.org/abs/2203.15349

LDKP - A Dataset of Long Documents

LDKP3k



LDKP10k



Debanjan Mahata, Naveen Agarwal, Dibya Gautam, Amardeep Kumar, Swapnil Parekh, Yaman Kumar Singla, Anish Acharya, Rajiv Ratn Shah. LDKP: A Dataset for Identifying Keyphrases from Long Scientific Documents. https://arxiv.org/abs/2203.15349

Data Resources

https://huggingface.co/midas

midas/oagkx
Preview • Updated 1 day ago

midas/kpcrowd
Preview • Updated 7 days ago

midas/www
Preview • Updated 8 days ago

midas/semeval2017
Preview • Updated 8 days ago

midas/kp20k
Preview • Updated 12 days ago

midas/kptimes
Preview • Updated 13 days ago

midas/inspec
Preview • Updated 23 days ago • ↓ 1.01k

midas/semeval2010
Preview • Updated 23 days ago

midas/ldkp3k
Updated 25 days ago

midas/ldkp10k
Preview • Updated 25 days ago

midas/citeulike180
Preview • Updated 28 days ago

midas/cstr
Preview • Updated 28 days ago

Data Resources

```python
from datasets import load_dataset

# get entire dataset
dataset = load_dataset("midas/<dataset_name>", "raw")

# get the dataset only for keyphrase extraction
dataset = load_dataset("midas/<dataset_name>", "extraction")

# get the dataset only for keyphrase generation
dataset = load_dataset("midas/<dataset_name>", "generation")
```


```python
from datasets import load_dataset

# get ldkp3k small
dataset = load_dataset("midas/ldkp3k", "small")

# get ldkp3k medium
dataset = load_dataset("midas/ldkp3k", "medium")

# get ldkp3k large
dataset = load_dataset("midas/ldkp3k", "large")
```

```python
from datasets import load_dataset

# get ldkp10k small
dataset = load_dataset("midas/ldkp10k", "small")

# get ldkp10k medium
dataset = load_dataset("midas/ldkp10k", "medium")

# get ldkp10k large
dataset = load_dataset("midas/ldkp10k", "large")
```

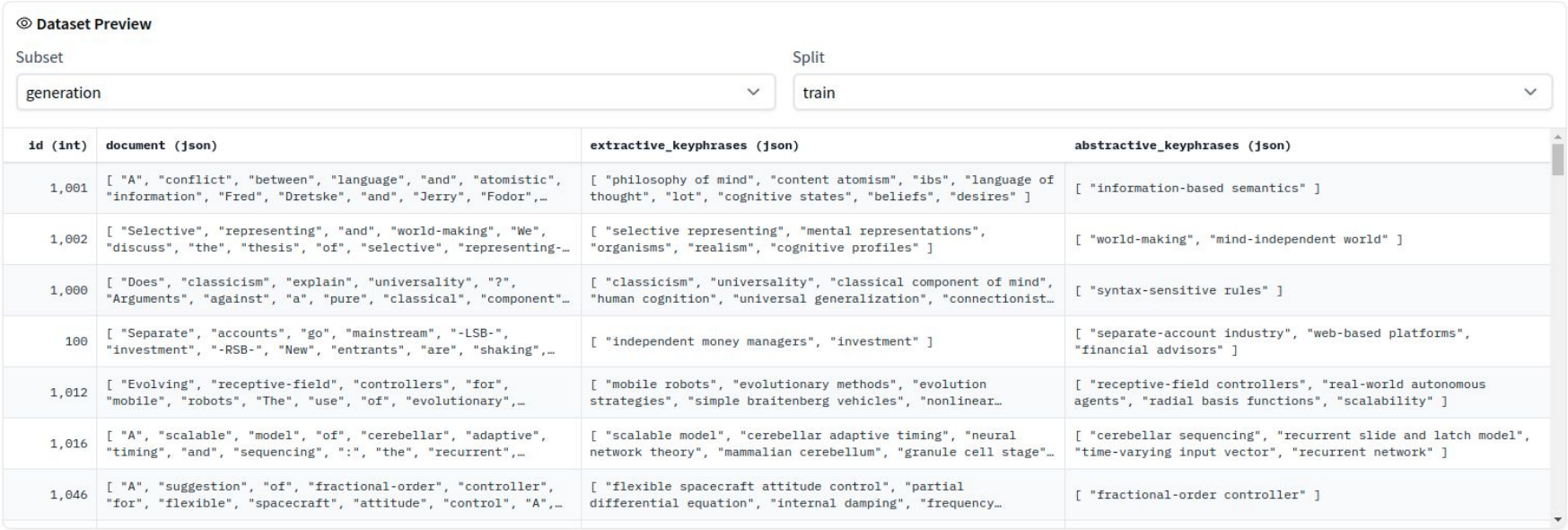
BIO Tagged Data for Keyphrase Extraction

Dataset Preview — Go to dataset viewer

Subset: extraction | Split: train

id (int)	document (json)	doc_bio_tags (json)
1,001	["A", "conflict", "between", "language", "and", "atomistic", "information", "Fred",…	["O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O",…
1,002	["Selective", "representing", "and", "world-making", "We", "discuss", "the", "thesis",…	["B", "I", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O",…
1,000	["Does", "classicism", "explain", "universality", "?", "Arguments", "against",…	["O", "B", "O", "B", "O", "O", "O", "O", "O", "B", "I", "I", "I", "O", "O", "O", "O",…
100	["Separate", "accounts", "go", "mainstream", "-LSB-", "investment", "-RSB-", "New",…	["O", "O", "O", "O", "O", "B", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O",…
1,012	["Evolving", "receptive-field", "controllers", "for", "mobile", "robots",…	["O", "O", "O", "O", "B", "I", "O", "O", "O", "B", "I", "O", "O", "O", "O", "O", "O",…
1,016	["A", "scalable", "model", "of", "cerebellar", "adaptive", "timing", "and",…	["O", "B", "I", "O", "B", "I", "I", "O", "O", "O", "O", "O", "O", "O", "O", "O", "O",…
1,046	["A", "suggestion", "of", "fractional-order", "controller", "for", "flexible", "spacecraft"…	["O", "O", "O", "O", "O", "O", "O", "B", "I", "I", "I", "O", "O", "O", "O", "O", "B", "I",…

Present and Absent Keyphrases for Keyphrase Generation



Dataset Preview

Subset: generation

Split: train

id (int)	document (json)	extractive_keyphrases (json)	abstractive_keyphrases (json)
1,001	["A", "conflict", "between", "language", "and", "atomistic", "information", "Fred", "Dretske", "and", "Jerry", "Fodor",…	["philosophy of mind", "content atomism", "ibs", "language of thought", "lot", "cognitive states", "beliefs", "desires"]	["information-based semantics"]
1,002	["Selective", "representing", "and", "world-making", "We", "discuss", "the", "thesis", "of", "selective", "representing-…	["selective representing", "mental representations", "organisms", "realism", "cognitive profiles"]	["world-making", "mind-independent world"]
1,000	["Does", "classicism", "explain", "universality", "?", "Arguments", "against", "a", "pure", "classical", "component"…	["classicism", "universality", "classical component of mind", "human cognition", "universal generalization", "connectionist…	["syntax-sensitive rules"]
100	["Separate", "accounts", "go", "mainstream", "-LSB-", "investment", "-RSB-", "New", "entrants", "are", "shaking",…	["independent money managers", "investment"]	["separate-account industry", "web-based platforms", "financial advisors"]
1,012	["Evolving", "receptive-field", "controllers", "for", "mobile", "robots", "The", "use", "of", "evolutionary",…	["mobile robots", "evolutionary methods", "evolution strategies", "simple braitenberg vehicles", "nonlinear…	["receptive-field controllers", "real-world autonomous agents", "radial basis functions", "scalability"]
1,016	["A", "scalable", "model", "of", "cerebellar", "adaptive", "timing", "and", "sequencing", ":", "the", "recurrent",…	["scalable model", "cerebellar adaptive timing", "neural network theory", "mammalian cerebellum", "granule cell stage"…	["cerebellar sequencing", "recurrent slide and latch model", "time-varying input vector", "recurrent network"]
1,046	["A", "suggestion", "of", "fractional-order", "controller", "for", "flexible", "spacecraft", "attitude", "control", "A",…	["flexible spacecraft attitude control", "partial differential equation", "internal damping", "frequency…	["fractional-order controller"]

Part I - Outline

- Introduction to keyphrasification
  - Definitions and applications
- Datasets and evaluations
- Traditional methods for keyphrase extraction (**Florian**)
- Hands-on practice with PKE