

DeepCT-enhanced Lexical Argument Retrieval

August 15, 2024



Alexander
Bondarenko



Maik
Fröbe



Danik
Hollatz



Jan Heinrich
Merker



Matthias
Hagen

Friedrich-Schiller-Universität Jena

<https://webis.de>

DeepCT-enhanced Lexical Argument Retrieval

Argument retrieval

Existing argument search engines use lexical retrieval models.

- ❑ Examples: args.me [Wachsmuth et al. EMNLP'17] OR ArgumenText [Stab et al. NAACL-HLT'18].
- ❑ Goal: Relevance to query, presence/quality of arguments.
- ❑ Focus on controversial topics, e.g., 'ban bottled water'.

The screenshot shows the args.me search engine interface. The search bar contains the query "Should bottled water be banned?". Below the search bar, there are navigation links for "All", "Discussions", "News", and "People". The search results are displayed in two columns, one for "PRO" (Pro) and one for "CON" (Con) arguments. Each result includes a title, a "Show full argument" link, a snippet of the argument text, and a URL with a score.

args Should bottled water be banned?

All Discussions News People Pro vs. con view - 88 arguments retrieved in 613.0 ms

PRO

Even though we use them for daily use, animals shouldn't...
▶ Show full argument
Even though we use them for daily use, animals shouldn't pay because you wish to consume some over-sugared drink. ... We need to put a stop to the destruction of animal life and think before we take action. ...
<https://www.debate.org/debates/plastic-bottles-should-be-banned/1/> score -

Wildlife damage: People most of the time do not take care...
▶ Show full argument
Wildlife damage: People most of the time do not take care about how they dispose their trash, Especially **plastic**. ... Sometimes they only leave their their **plastic** backs on the street, Without worrying about animals such as ...
<https://www.debate.org/debates/Plastic-bags-should-be-banned/2/> score -

it should be banned because oil is wasted making the...
▶ Show full argument
it **should be banned** because oil is wasted making the **plastic bottles** and it is also very unhealthy for you by drinking something that is contained in **plastic** because the **plastic** mixes into the water giving it that **plastic** ...
<https://www.debate.org/debates/Bottled-water-should-be-banned/1/> score -

CON

But, if we were to ban all plastic bottles, or plastics...
▶ Show full argument
But, if we were to ban all **plastic bottles**, or plastics overall, then where would we put the **plastic**? ... But face it: if we're going to eliminate all the **plastic**, we have to consider the cost, as well as where it would be ...
<https://www.debate.org/debates/plastic-bottles-should-be-banned/1/> score -





If you want to ban plastic water bottles almost...
▶ Show full argument
If you want to ban **plastic water bottles** almost exclusively because they contain oil, then by that logic all the other products that I listed **should** also **be banned** on the sole grounds that they contain oil. Drinking clean ...
<https://www.debate.org/debates/Bottled-water-should-be-banned/1/> score -

Whilst banning some bags and single-use items like...
▶ Show full argument
Whilst banning some bags and single-use items like straws, Etc might make sense. ... A flat out all-out ban cannot work because of the requirements of sterility. For which there are no other acceptable solutions. ...
<https://www.debate.org/debates/Plastic-bags-should-be-banned/2/> score -

DeepCT-enhanced Lexical Argument Retrieval

Curious case of Touché

- ❑ Touché is a series of shared tasks on computational argumentation.
- ❑ Includes argument retrieval tasks [Bondarenko et al. CLEF'20, CLEF'21].
- ❑ Most effective participant approaches used lexical retrieval models.
- ❑ Touché in the BEIR benchmark for zero-shot retrieval [Thakur et al. SIGIR'24].

Model	Type	Effectiveness
BM25	Lexical	
E5 _{large}	Dense	
CITADEL+	Multi-vector	
uniCOIL	Sparse	

DeepCT-enhanced Lexical Argument Retrieval

Touché test collection

Test collection:

- ❑ Args.me corpus (400 000 documents) [Ajjour et al. KI'19].
- ❑ Argument passages from debate portals: idebate.org, debate.org, ...
- ❑ 99 queries (topics) on controversial topics, 6 000 relevance judgments.

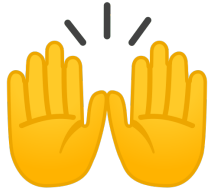
Example topic:

Title	<i>Should bottled water be banned?</i>
Description	<i>Tap water is a valid alternative to bottled water. A user wonders why tap water is not the norm, and whether, to save resources, bottled water should be banned.</i>
Narrative	<i>Highly relevant arguments argue for or against bottled water, giving clear reasons for their conclusions. Relevant arguments mention specific kinds of bottled water (e.g., plastic bottles), or drinking water quality and the industry in general.</i>

DeepCT-enhanced Lexical Argument Retrieval

Argument relevance

Relevance: How well do arguments fit the topic?



Highly relevant



Relevant



Not relevant



Spam

Topic: *Should bottled water be banned?*



Takes stance for/against bottled water.



About plastic bottles, but not bottled water.



Argumentative, but not about bottled water or plastic bottles.



Not an argument.

DeepCT-enhanced Lexical Argument Retrieval

Our approach

- ❑ Starting point: Lexical models \gg neural models but lack “semantics”.
- ❑ Intuition: Add “semantics” to lexical retrieval.
- ❑ But neural models need a lot of training data and compute.
- ❑ Idea: DeepCT-based document expansion [Dai and Callan WWW'20, SIGIR'20].
- ❑ Advantages: no relevance judgments needed and can be done offline.
- ❑ Evaluation: Measure retrieval effectiveness (nDCG@5, bpref).

DeepCT-enhanced Lexical Argument Retrieval

DeepCT term weighting

Document: *Plastic **water** **bottles** were the third most commonly collected **waste** during the Ocean Conservancy's International Coastal Cleanup behind cigarette butts and plastic food wrappers. [...] A nationwide **ban** on **bottled** **water** would lead to an estimated 68 billion fewer plastic **water** **bottles** being manufactured, purchased, used, and discarded.*

Reference: *water: 1.0, bottles: 1.0, waste: 1.0, ban: 1.0, bottled: 1.0.*

- ❑ BERT to predict the importance of words in documents w.r.t. reference terms.
- ❑ Terms are repeated: $w * 100$, where term weights $w \in [0, 1]$.
- ❑ Originally finetuned on MS MARCO [Nguyen et al. CoCo@NIPS'16].
- ❑ Finetune DeepCT for argument retrieval.
- ❑ Exploit the structure of args.me documents.

DeepCT-enhanced Lexical Argument Retrieval

Why repeating terms may work

Document: *Plastic water water water water water water water water water bottles bottles bottles bottles were the third most commonly collected waste waste waste waste waste waste waste waste during the Ocean Conservancy's International Coastal Cleanup behind cigarette butts and plastic food wrappers. [...]*

- TF weighting:

$$BM25(Q, D) = \sum_{i=1}^n IDF(q_i) \cdot \frac{TF(q_i, D) \cdot (k_1 + 1)}{TF(q_i, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{avgdl}\right)}$$

$$DirichletLM(Q, D) = \sum_{i=1}^n \log \left(1 + \frac{TF(q_i, D)}{\mu \cdot \frac{DF(q_i)+1}{N+1}} \right) + \log \left(\frac{\mu}{|D| + \mu} \right)$$

DeepCT-enhanced Lexical Argument Retrieval

Fine-tuning DeepCT for argument retrieval

Debate topic: *Bottle waste.*

Conclusion: *Ban bottled water.*

Premises: *Plastic water bottles were the third most commonly collected waste during the Ocean Conservancy's International Coastal Cleanup behind cigarette butts and plastic food wrappers. [...] A nationwide ban on bottled water would lead to an estimated 68 billion fewer plastic water bottles being manufactured, purchased, used, and discarded.*

- ❑ Split the premises into passages of 500 tokens.
- ❑ Remove stop words and do stemming.
- ❑ Look for overlap.
- ❑ Original form of the word is added as a reference term.
- ❑ No need for relevance judgments, query independent.

DeepCT-enhanced Lexical Argument Retrieval

Evaluation

Data transf.: (1) all documents, (2) judged, and (3) top-50 documents are removed.

Reference field: debate topic and conclusion (TC) or conclusion only (C).

Underlines: best system per metric and **bold:** significant equivalence to the best system ± 0.1 .

	Retrieval model	Data transf.	nDCG@5	bpref	judged@5
Touché 2020	DeepCT + DirichletLM + RM3	(1), C	0.88	0.71	0.45
	BM25 + monoT5	n/a	0.87	0.81	0.41
	DeepCT + BM25 + RM3	(2), TC	0.87	0.77	0.46
	BM25 + LiT5	n/a	0.86	0.51	0.39
	DeepCT + BM25	(2), TC	0.84	0.71	0.47
	Best Touché	n/a	0.83	0.70	1.00
...					
Touché 2021	BM25 + monoT5	n/a	0.77	0.80	0.70
	DeepCT + BM25	(3), TC	0.74	0.74	0.78
	DeepCT + BM25 + RM3	(2), TC	0.74	0.74	0.70
	Best Touché	n/a	0.74	0.73	1.00
	...				

- ❑ Our approach improves over the Touché best systems.
- ❑ On par with the monoT5 re-ranker but DeepCT is applied at index time and does not require model inference at query time.

DeepCT-enhanced Lexical Argument Retrieval

Summary

- ❑ Combine lexical retrieval models with semantic document expansion.
- ❑ Can be done in an offline fashion.
- ❑ No or little training data is available.
- ❑ Our approach is on par with modern neural re-rankers.
- ❑ Neural models can be more computationally expensive.
- ❑ Future work: argument mining step in the document expansion process.

Code and Data

 github.com/webis-de/argmining24-deepct-lexical-argument-retrieval/

DeepCT-enhanced Lexical Argument Retrieval

Summary

- ❑ Combine lexical retrieval models with semantic document expansion.
- ❑ Can be done in an offline fashion.
- ❑ No or little training data is available.
- ❑ Our approach is on par with modern neural re-rankers.
- ❑ Neural models can be more computationally expensive.
- ❑ Future work: argument mining step in the document expansion process.

Code and Data

🔗 github.com/webis-de/argmining24-deepct-lexical-argument-retrieval/

Thank you!