

#805 Eating Invasive Species

Yonglin Wang, Xiaoyu Lu, Yun-jing Lee, Ruobin Hu

Outline

- Retrieval & Reranking Approaches
 - Query Expansion
 - Longformer Embedding
- Corpus-wide Evaluation for Longformer
- Web UI
 - Overall UI Design
 - Keyword Highlight Algorithm
- Demonstration
- Conclusions

Topic Overview

Title:

eating invasive species

Description:

Would eating invasive species be a viable method of controlling/eradicating them?

Narrative:

Relevant documents identify **invasive species in the U.S.** that can be eaten by human beings and discuss the likelihood of controlling the species through consumption. Comments on the color, taste, texture, or other culinary qualities of particular invasive species are also relevant.

Query Term Expansion

eating invasive species

PA5 results

Score Type	Best Score	Best Combo
AP	0.533	Default + FastText
Precision	0.35	Custom + BM25
NDCG@20	0.695	Default + FastText

Query Term Expansion

eating invasive species

PA5 results

Score Type	Best Score	Best Combo
AP	0.533	Default + FastText
Precision	0.35	Custom + BM25
NDCG@20	0.695	Default + FastText

False Positives

An enemy in our midst: Maryland tries to fight invasive plant species [805-0]

This little flea — and its huge appetite — could ruin the Great Lakes [805-0]

How growing global trade is carrying pests all over the world [805-0]

Query Term Expansion (1)

eating **invasive species**



Expand “invasive species”...

common_name([List of invasive species in North America](#) \cap [List of edible invasive species](#))

Query Term Expansion (1)

eating **invasive species**



Expand “invasive species”...

common_name(List of invasive species in North America \cap List of edible invasive species) = {Kudzu, Zebra mussel, Asian carp, Snakehead, Giant African land snail, Lionfish}



Query Term Expansion (1)

eating

invasive species Kudzu Zebra mussel Asian carp
Snakehead Giant African land snail Lionfish

- Combined with “eating invasive species”, reranking w/ sBERT

Score Type	Wiki Expand	PA5 Best
AP	0.659 ▲	0.533
Precision	0.45 ▲	0.35
NDCG@20	0.772 ▲	0.695

Query Term Expansion (2)

eating

invasive species Kudzu Zebra mussel Asian carp
Snakehead Giant African land snail Lionfish

- Persistent problem: who's doing the eating?

**Divers try spoon feeding lionfish to sharks, a method
that could come back to bite them [805-0]**

Query Term Expansion (2)

eating

invasive species Kudzu Zebra mussel Asian carp
Snakehead Giant African land snail Lionfish

- Persistent problem: who's doing the eating?

**Divers try spoon feeding lionfish to sharks, a method
that could come back to bite them [805-0]**

- Search for a better word, informed by GloVe embeddings



```
gensim.downloader.load("glove-wiki-gigaword-50").most_similar("eat")
```

Query Term Expansion (2)

eating

invasive species Kudzu Zebra mussel Asian carp
Snakehead Giant African land snail Lionfish

- Persistent problem: who's doing the eating?

**Divers try spoon feeding lionfish to sharks, a method
that could come back to bite them [805-0]**

- Search for a better word, informed by GloVe embeddings



```
gensim.downloader.load("glove-wiki-gigaword-50").most_similar("eat")
```

“eating ate eaten eats meal chicken meat cooked fish drink”

Query Term Expansion (2)

eating cooking

invasive species Kudzu Zebra mussel Asian carp
Snakehead Giant African land snail Lionfish

- Persistent problem: who's doing the eating?

**Divers try spoon feeding lionfish to sharks, a method
that could come back to bite them [805-0]**

- Search for a better word, informed by GloVe embeddings



```
gensim.downloader.load("glove-wiki-gigaword-50").most_similar("eat")
```

“eating ate eaten eats meal chicken meat cooked fish drink”

Query Term Expansion (2)

eating cooking

invasive species Kudzu Zebra mussel Asian carp
Snakehead Giant African land snail Lionfish

Score Type	Wiki + Cook	Wiki Expand	PA5 Best
AP	0.772 ▲	0.659	0.533
Precision	0.5 ▲	0.45	0.35
NDCG@20	0.836 ▲	0.772	0.695

Query Term Expansion (2)

eating cooking

invasive species Kudzu Zebra mussel Asian carp
Snakehead Giant African land snail Lionfish

After retrieve with keywords above, rerank with: **eating and cooking invasive species such as**
Kudzu Zebra mussel Asian carp Snakehead Giant African land snail and Lionfish

Query Term Expansion (2)

eating cooking

invasive species Kudzu Zebra mussel Asian carp
Snakehead Giant African land snail Lionfish

After retrieve with keywords above, rerank with: **eating** and **cooking** **invasive species** such as
Kudzu Zebra mussel Asian carp Snakehead Giant African land snail and **Lionfish**

Score Type	Wiki + Cook + NL rerank	Wiki + Cook	Wiki Expand	PA5 Best
AP	0.837 ▲	0.772	0.659	0.533
Precision	0.5	0.5	0.45	0.35
NDCG@20	0.917 ▲	0.836	0.772	0.695

Query Specific Term Expansion

- A persistent problem: in the false positives, the agents of “eating” or “consuming” are usually the invasive species and involve no humans/people
 - “[the land snails] are known to consume over 500 species of plants”
- Substitute “eating” and “consuming”
 -
 - A missing keyword: cooking
- eating cooking invasive species Kudzu Zebra mussel Asian carp Snakehead Giant African land snail Lionfish
 - AP-0.772 Precision-0.5 NDCG@20-0.836
- eating **and** cooking invasive species **such as** Kudzu Zebra mussel Asian carp Snakehead Giant African land snail **and** Lionfish
 - AP-0.795 Precision-0.55 NDCG@20-0.89
- Keyword for search: AP-0.837 Precision-0.5 NDCG@20-0.917

Separating Query Texts

- BERT is probably better at reranking when query is “natural”
- Keyword jumbo w/ sBERT + custom: AP-0.675 Precision-0.65 NDCG@20-0.828
 - “eating invasive species cooking baking frying restaurant culinary Kudzu Zebra mussel Asian carp Snakehead Giant African land snail Lionfish”
- Natural language rewrite:
 - how do people eat invasive species such as Kudzu, Zebra mussel, Asian carp, Snakehead, Giant African land snail, or Lionfish at restaurants, by cooking, baking or frying?
 - Kudzu Zebra mussel Asian carp Snakehead Giant African land snail Lionfish restaurant eat invasive species cooking baking frying
 - Kudzu Zebra mussel Asian carp Snakehead Giant African land snail Lionfish: how do restaurants eat invasive species by cooking baking or frying

Longformer

- Motivation
- High-level: How does it work?
- Configurations: All default tokenizer; BM25 alone, BM25+reranked w/Longformer, Longformer alone
- Look at scores @35 (total relevant docs); heatmap results if needed?
- Does it retrieve more relevant documents?
 - Use this embedding alone, w/o BM25
 - Look at precision @35, before and after
- Does it help with reranking?
 - Use this embedding after BM25 retrieval
 - Look at NDCG @35, before and after
- Does it help overall?
 - Look at AP

Longformer Embedding

- Pretrained, context-aware embedding model just like sBERT.
- But crucially: max input length is 4096, while sBERT is 510.

Longformer Embedding

- Pretrained, context-aware embedding model just like sBERT.
- But crucially: max input length is 4096, while sBERT is 510.
- However, didn't bring improvement to our topic...

Longformer Embedding

- Pretrained, context-aware embedding model just like sBERT.
 - But crucially: max input length is 4096, while sBERT is 510.
 - However, didn't bring improvement to our topic...
-
- What about other topics, compared to sBERT and baseline (BM25)?

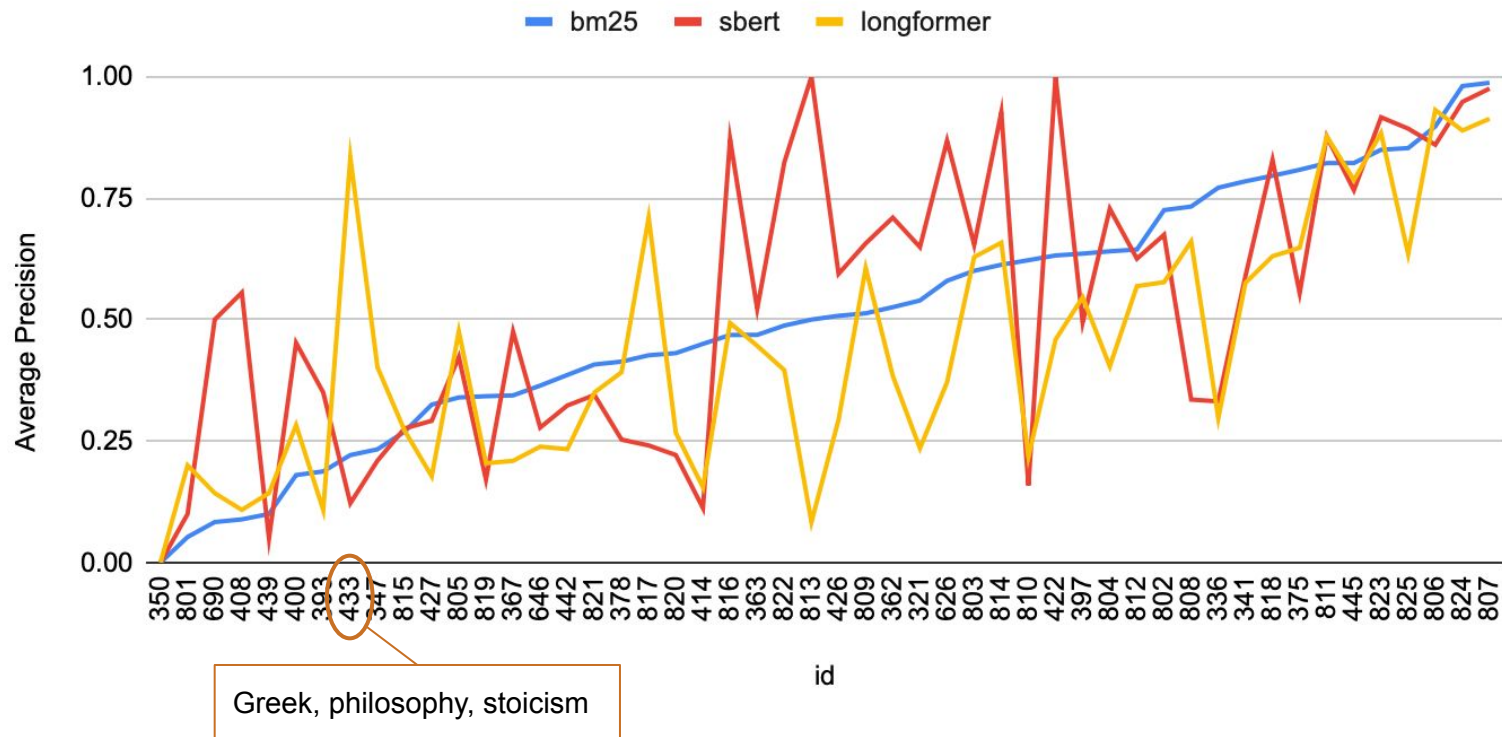
Generalizing to other queries

MAP of our automatable approach (Longformer) vs previous approaches (BM25 and sBERT):

- Title
- Narrative
- Description

Average Precision by Topic

Average Precision on Title



Web UI

Washington Post Search

Query: ☒ BM25+default ☐ BM25+custom ☐ ftvector+default ☐ sbertvector+default



Washington Post Search

Query:

Analyzer: ☒ Basic analyzer ☐ Custom analyzer

Embedding: ☒ BM25 ☐ FastText ☐ Sentence Bert ☐ Longformer

Web UI

Before

Washington Post Search

Query: ☒ BM25+default ☐ BM25+custom ☐ ftvector+default ☐ sbertvector+default

[Home](#)

9. [Goats battle the invasion of alien plant life](#)

When officials from the Belmont Country Club homeowners association realized that weeds were threatening trees in the neighborhood' s protected areas, ...

10. [Where have all the animals gone?](#)

Here' s a phrase to remember: the Anthropocene defaunation. That' s a fancy way of saying that the animals are disappearing. Humans already did their wo ...

11. [A scientist discovered a new insect species living in New Jersey. He' s already worried about its survival.](#)

A grass-eating species of insect that makes its home in the Pine Barrens of New Jersey - just south of New York City - has until recently escaped the ...

12. [An enemy in our midst: Maryland tries to fight invasive plant species](#)

The fig buttercup looks harmless, its tiny yellow flowers and wide green leaves seemingly perfect for a garden bed or patio pot. Even its name sounds ...

13. [Spread of snakehead fish in Potomac River concerns area officials](#)

A decade or so ago, no snakehead fish were in the Washington region. Today, more than 20,000 are in the Potomac River, experts say, a population appro ...

14. [Invasive Asian carp could overtake Lake Erie, study finds](#)

A new study suggests that Asian carp, which have been spotted in watersheds close to the Great Lakes, would make themselves right at home if they made ...

15. [Aggressive lionfish invade non-native waters. Eat them.](#)

The gorgeous, deadly lionfish, native to the Indian and Pacific oceans, is presenting a human-propelled threat to fisheries in the rest of the world, ...

16. [In the Everglades, hunters help to check a slithery invader](#)

MIAMI — As they bounced into the Everglades in a beat-up truck, Ruben Ramirez and George Brana overflowed with confidence. They were in the final days ...

After

Washington Post Search

Query:

Analyzer: ☒ Basic analyzer ☐ Custom analyzer

Embedding: ☒ BM25 ☐ FastText ☐ Sentence Bert ☐ Longformer

Total hits: 20

Eat more fish; risks overstated

2012/04/02 - Even so, if you **eat** more than two servings of fish per week, both his group and the IOM advise that you don ' t always **eat** the same type, so you ' re not **eating** one contaminant over and over.

Ornamental fish face threat of infection as their resistance to antibiotics grows

2013/02/04 - Scientists studied the bacteria found in 32 **species** of freshwater fish that were imported from Singapore, Colombia and Florida, the major hubs for ornamental fish export . They found nine bacterial **species** that were not susceptible to any of ...

No, journalists, global warming is not going to lead to the end of fish and chips

2015/04/14 - However, she acknowledged that “ in terms of where fish and chips come from you ' re right, since we still want to **eat** traditional colder water fish much of it has to be imported since the North Sea has experienced massive stock reductions due to fishing pressure and ...

Web UI

Washington Post Search

Ornamental fish face threat of infection as their resistance to antibiotics grows

Author: ScienceNOW

Date: 2013/02/04

Trouble in the fish tank Ornamental fish face a rough ride when shipped from their home waters to aquariums around the world. To stave off infection, antibiotic use is widespread in the ornamental fish industry. But the practice is proving increasingly ineffective and contributes to antibiotic resistance in these ornamental fish, according to a study published last month in the Journal of Fish Diseases. Scientists studied the bacteria found in 32 species of freshwater fish that were imported from Singapore, Colombia and Florida, the major hubs for ornamental fish export. They found nine bacterial species that were not susceptible to any of the antibiotics tested. The most effective antibiotic had 16 percent resistance. The least effective drug faced 77 percent resistance. Though the health risks to humans remains low, the bigger problem is that the \$15 billion ornamental fish industry will face a growing challenge treating diseased fish as antibiotics remain unregulated. Researchers plan to use this information to educate fish farmers on the need to curb the blanket use of antibiotics in fish feed. — ScienceNOW

Bold the text

Query	Doc content
"Fishes eating noth invasive species!!!"	<div data-bbox="966 347 1139 380">Sentence 1</div> <div data-bbox="1472 347 1646 380">Sentence 2</div> <p data-bbox="774 385 1843 578">"... . Fishing season. Taking Frankfish, creeping into your northern — invasive species are everywhere. ..."</p>

Bold the text

Query	Doc content
" Fishes eating noth invasive species!!! "	"... . Fishing season. Taking Frankfish , creeping into your northern — invasive species are everywhere. ..."

misspelling

Sentence 1

Sentence 2

Bold the text

Query	Doc content
" Fishes eating noth invasive species!!! "	<div>Sentence 1</div> <div>Sentence 2</div> <p>"... . Fishing season. Taking Frankfish, creeping into your northern — invasive species are everywhere. ..."</p>

Display on UI: "Taking **Frankfish**, creeping into your northern — **invasive species** are everywhere."

- fishes -> fishing
- Fishes -> Frankfish

Bold the text

- **Fishes** -> **Fishing**

```
>>> lemmatizer.lemmatize("fishing", "v")  
'fish'  
>>> lemmatizer.lemmatize("fishes", "n")  
'fish'
```

Bold the text

- **Fishes -> Fishing**

```
>>> lemmatizer.lemmatize("fishing", "v")  
'fish'  
>>> lemmatizer.lemmatize("fishes", "n")  
'fish'
```

- **Fishes -> Frankfish**

```
>>> hyphenate_word("Frankfish")  
['Frank', 'fish']
```

Bold the text

Query	Doc content
" Fishes eating noth invasive species!!!! "	<div>Sentence 1</div> <div>Sentence 2</div> <p>"... . Fishing season. Taking Frankfish, creeping into your northern — invasive species are everywhere. ..."</p>

↓

ALGORITHM

↓

Display on UI: "Taking **Frankfish**, creeping into your northern — **invasive species** are everywhere."

"Fishes eating north invasive species!!!"

tokenize func (lemma + remove pun)

['fish', 'eat', 'noth', 'invasive', 'specie']

query tokens indices

[0, 4, 11, 12]

algo(binary search)

(1,4)

[4,11,12] - 3 => [1,8,9]

doc_id: 42

db_whole

"Fishing season. Taking Frankfish, creeping into your northern –
invasive species are everywhere."

0 1 2 3 4
11 12 13 14 15

EOS punc indices (. ! ?)

[2, 15]

(0,1)

[2+1: 15+1]

doc_id	title	author	annotation	published_date	content_str	words_dict
1	24 Spaceship Earth...	Joel Achenbach	805-0	2012/01/02	Spaceship Earth enters 2012 belch...	defaultdict(<class 'set', {'space': {0, 122}, 'ship': {0, 122...
2	154 Untitled	Rita Zeidner	805-0	2012/01/12	Ask Kevin Green about his researc...	defaultdict(<class 'set', {'ask': {0}, 'kevin': {1}, 'green': {...
3	221 Nearly 7 million ...	Darryl Fears	805-0	2012/01/17	More than five years since the dea...	defaultdict(<class 'set', {'more': {0, 392, 917, 495}, 'tha...

basic tokenize func

0 1 2 3
['Fishing', 'season', '.', 'Taking', 'Frankfish', ',', 'creeping', 'into',
'northern', 'everywhere', '-', 'invasive', 'species', 'everywhere', '.']

0 1 2 3 4 5
['Taking', 'Frankfish', ',', 'creeping', 'into', 'northern',
'everywhere', '-', 'invasive', 'species', 'everywhere', '.']

6 7 8 9 10 11
['Taking', 'Frankfish', ',', 'creeping', 'into', 'northern', 'everywhere',
'-', 'invasive', 'species', 'everywhere', '.']

detokenize

Taking Frankfish, creeping into northern everywhere
– invasive species everywhere...

input/output

algorithm/function

Bold the text

Query	Doc content
" Fishes eating noth invasive species!!! "	<div>Sentence 1</div> <div>Sentence 2</div> <p>"... . Fishing season. Taking Frankfish, creeping into your northern — invasive species are everywhere. ..."</p>



ALGORITHM

Display on UI: "Taking **Frankfish**, creeping into your northern — **invasive species** are everywhere."

Display on UI

Taking **Frankfish**, creeping into your northern — **invasive species** are everywhere.

Too short ???

Display on UI

Taking **Frankfish**, creeping into your northern — **invasive species** are everywhere. **John Waugh** is an independent adviser on environmental strategies and ...

Display on UI

Taking **Frankfish**, creeping into your northern — **invasive species** are everywhere. **John Waugh is an independent adviser on environmental strategies and ...**

In any case, state wildlife officials said, the goal of the hunt wasn't just bagging snakes, but also drawing attention to the havoc that pythons and other **invasive species** — such as two large and greedy lizards, the black and white Argentine tegu and Nile monitor, as well as the Cuban tree frog — wreak upon native wildlife in the Sunshine State.

Display on UI

Taking **Frankfish**, creeping into your northern — **invasive species** are everywhere. **John Waugh is an independent adviser on environmental strategies and ...**

Too long ???

In any case, state wildlife officials said, the goal of the hunt wasn't just bagging snakes, but also drawing attention to the havoc that pythons and other **invasive species** — such as two large and greedy lizards, the black and white Argentine tegu and Nile monitor, as well as the Cuban tree frog — wreak upon native wildlife in the Sunshine State.

Display on UI

Taking **Frankfish**, creeping into your northern — **invasive species** are everywhere. **John Waugh is an independent adviser on environmental strategies and ...**

In any case, state wildlife officials said, the goal of the hunt wasn't just bagging snakes, but also drawing attention to the havoc that pythons and other **invasive species** — such as two large and greedy lizards, the black and white Argentine tegu and Nile....

DEMO TIME

Conclusions

- Retrieval Approaches
 - Taking intersections on Wikipedia data helps narrow down keywords
 - Reranking with a different text might help, especially when expanding query terms
 - Implementing Longformer to include more document texts might be helpful, depending on the topic
- Corpus-wide Comparison and Analysis
 - Longformer *sometimes* helps with reranking BM25 results
- Web UI
 - Separate input texts for retrieving and reranking
 - Keyword Highlight
 - User-centric UI design