

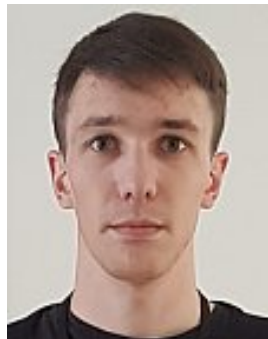
# Using Keyqueries to Reduce Misinformation in Health-Related Search Results

---

April 10, 2022



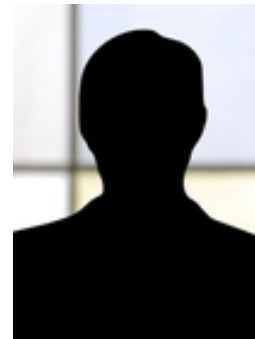
Maik Fröbe



Sebastian  
Günther



Alexander  
Bondarenko



Johannes Huck



Matthias Hagen

Martin-Luther-Universität Halle-Wittenberg

first.last@informatik.uni-halle.de

@webis\_de

www.webis.de

# Challenges in Health-Related Searches

## Example Information Need

[Clarke et al.; TREC'21]

Does selenium help prevent cancer?

# Challenges in Health-Related Searches

## Example Information Need

[Clarke et al.; TREC'21]

Does selenium help prevent cancer?



# SELENIUM AUTOMATION TESTING

# Challenges in Health-Related Searches

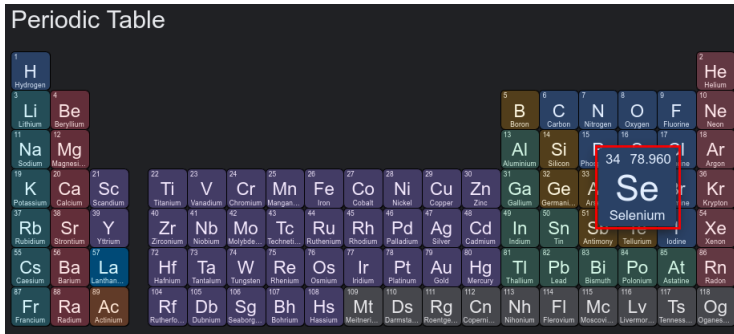
## Example Information Need

[Clarke et al.; TREC'21]

Does selenium help prevent cancer?



Periodic Table



1																	2
H																	He
3	4															10	
Li	Be															Ne	
11	12															18	
Na	Mg															Ar	
19	20	21												36			
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
27	28	29	30	31	32	33	34	35	36								54
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
55	56	57												84	85	86	
Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104
Fr	Ra	Ac	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Nh	Fl	Mc	Lv	Ts	Og

- ❑ Selenium is a chemical element
- ❑ Medical use-cases of selenium:
  - selenium deficiency
  - high cholesterol

# Challenges in Health-Related Searches

## Example Information Need

[Clarke et al.; TREC'21]

Does selenium help prevent cancer?



Periodic Table

- ❑ Selenium is a chemical element
- ❑ Medical use-cases of selenium:
  - selenium deficiency
  - high cholesterol

CHAT NOIR

selenium cancer

Settings

Search results 1-10 for *selenium cancer* Total results: 235,985 (retrieved in 1518.6ms)

**Selenium May Prevent Advanced Prostate Cancer**  
Conclusion: In our study, **selenium** was associated with a substantial decrease in risk of advanced prostate **cancer**.

**Selenium Reduces the Progression of Colorectal Cancer**  
**Selenium** reduces colorectal **cancer** progression by various mechanisms such as reduction in the number and size of ACF and alteration in the function of the proteins such as P53, and BCL2.

**Vitamins E and C and Selenium do Not Protect Against Cancer**  
Two new studies in the Journal of the American Medical Association have found that people who supplement their diet with vitamins E and C or the mineral **selenium** do not have a lower risk of **cancer**.

**National Cancer Institute ends Selenium and Vitamin E Cancer Prevention Trial**  
Vitamin E and **selenium** are ineffective in preventing **cancer** and there's a hint of risk.

# Challenges in Health-Related Searches

## Example Information Need

[Clarke et al.; TREC'21]

Does selenium help prevent cancer?



Periodic Table

- ❑ Selenium is a chemical element
- ❑ Medical use-cases of selenium:
  - selenium deficiency
  - high cholesterol

CHAT NOIR

selenium cancer

Settings

Search results 1-10 for *selenium cancer* Total results: 235,985 (retrieved in 1518.6ms)

**Selenium May Prevent Advanced Prostate Cancer** Relevant  
Conclusion: In our study, **selenium** was associated with a substantial decrease in risk of advanced prostate **cancer**.

**Selenium Reduces the Progression of Colorectal Cancer** Relevant  
**Selenium** reduces colorectal **cancer** progression by various mechanisms such as reduction in the number and size of ACF and alteration in the function of the proteins such as P53, and BCL2.

**Vitamins E and C and Selenium do Not Protect Against Cancer** Relevant  
Two new studies in the Journal of the American Medical Association have found that people who supplement their diet with vitamins E and C or the mineral **selenium** do not have a lower risk of **cancer**.

**National Cancer Institute ends Selenium and Vitamin E Cancer Prevention Trial** Relevant  
Vitamin E and **selenium** are ineffective in preventing **cancer** and there's a hint of risk.

# Challenges in Health-Related Searches

## Example Information Need

[Clarke et al.; TREC'21]

Does selenium help prevent cancer?



Periodic Table

- ❑ Selenium is a chemical element
- ❑ Medical use-cases of selenium:
  - selenium deficiency
  - high cholesterol

CHAT NOIR

selenium cancer

Settings

Search results 1-10 for *selenium cancer* Total results: 235,985 (retrieved in 1518.6ms)

**Selenium May Prevent Advanced Prostate Cancer** Relevant + Harmful  
Conclusion: In our study, **selenium** was associated with a substantial decrease in risk of advanced prostate **cancer**.

**Selenium Reduces the Progression of Colorectal Cancer** Relevant + Harmful  
**Selenium** reduces colorectal **cancer** progression by various mechanisms such as reduction in the number and size of ACF and alteration in the function of the proteins such as P53, and BCL2.

**Vitamins E and C and Selenium do Not Protect Against Cancer** Relevant + Helpful  
Two new studies in the Journal of the American Medical Association have found that people who supplement their diet with vitamins E and C or the mineral **selenium** do not have a lower risk of **cancer**.

**National Cancer Institute ends Selenium and Vitamin E Cancer Prevention Trial** Relevant + Helpful  
Vitamin E and **selenium** are ineffective in preventing **cancer** and there's a hint of risk.

# Challenges in Health-Related Searches

## Example Information Need

[Clarke et al.; TREC'21]

Does selenium help prevent cancer?



Periodic Table

- ❑ Selenium is a chemical element
- ❑ Medical use-cases of selenium:
  - selenium deficiency
  - high cholesterol

The screenshot shows a search engine interface for 'selenium cancer'. The search bar contains the text 'selenium cancer' and a magnifying glass icon. Below the search bar, the search results are displayed. The first result is titled 'Selenium May...' and the second is 'Selenium Redu...'. The search results are filtered to show 'Search results 1-10 for selenium cancer' with a total of 235,985 results. The first two results are highlighted as 'Relevant + Helpful'.

CHAT NOIR

selenium cancer

Search results

Selenium May  
Conclusion: In our stu

Selenium Redu  
Selenium reduces o  
as P53, and BCL2.

Vitamins E and  
Two new studies in t  
have a lower risk of

National Cancer  
Vitamin E and selenium are ineffective in preventing cancer and there's a hint of risk.

CHAT NOIR

selenium cancer

Search results 1-10 for **selenium cancer**

Total results: 235,985 (retrieved in 1518.6ms)

**Vitamins E and C and Selenium do Not Protect Against Cancer** Relevant + Helpful  
Two new studies in the Journal of the American Medical Association have found that people who supplement their diet with vitamins E and C or the mineral **selenium** do not have a lower risk of **cancer**.

**National Cancer Institute ends Selenium and Vitamin E Cancer Prevention Trial** Relevant + Helpful  
Vitamin E and **selenium** are ineffective in preventing **cancer** and there's a hint of risk.



# Automatic Approaches for Health-Related Searches?

## Fully Automatic Approaches Seem “Utopical”

- ❑ Scientific knowledge changes rapidly
- ❑ Distinguishing helpful and harmful documents is difficult
  - maybe not possible using only the documents text?

## We Take a Step Back

Can explicit relevance feedback by experts guide query expansion methods to formulate queries that return fewer misleading or wrong results?

- ❑ Approach:
  - Simulate expert providing explicit relevance feedback
  - Explicit relevance feedback is input for query expansion approach
  - Return ranking for expanded query

# Baselines

## Query Expansion with RM3

[Jaleel et al.; TREC'04]

- ❑ Adds new terms with weights to a query
- ❑ Exploits information from feedback documents
- ❑ Step 1 (i.e., RM1): Estimate relevance model for feedback documents
  - Each term gets an score.
  - E.g., for feedback documents  $R$  for the original query  $q$ , a term  $t$  obtains:

$$\sum_{d \in R} P(t|d) \frac{score(q, d)}{\sum_{d' \in R} score(q, d')},$$

- ❑ Step 2 (i.e., RM3): Linearly combine original query weight with RM1

Observation: RM3 might be ineffective for explicit relevance feedback

- ❑ RM3 does not consider ranks of feedback documents
- ❑ RM3 does not check whether all expansion terms are actually needed
- ❑ Result: Costly relevance feedback by experts is maybe not fully leveraged

# Can We Make Better Use of Explicit Feedback From Experts?

The key are ...

# Can We Make Better Use of Explicit Feedback From Experts?

The key are ... keyqueries

# Query Expansion with Keyqueries

What is a keyquery? [Hagen et al.; ECIR'16]

Query  $q$  is a keyquery for a set  $D$  of target documents against a search engine iff

1. Every  $d \in D$  is in the top- $k$  results. (specificity)
2. Query  $q$  has at least  $l$  results. (generality)
3. No subquery  $q' \subset q$  satisfies the above. (minimality)

Remark: For small  $|D| \leq 5$ , typically  $l \geq 10$  and  $k = 10$ .

# Query Expansion with Keyqueries

Example: Keyquery for a paper ( $l \geq 1000$ ,  $k = 3$ )

## Elastic ChatNoir: Search Engine for the ClueWeb and the Common Crawl

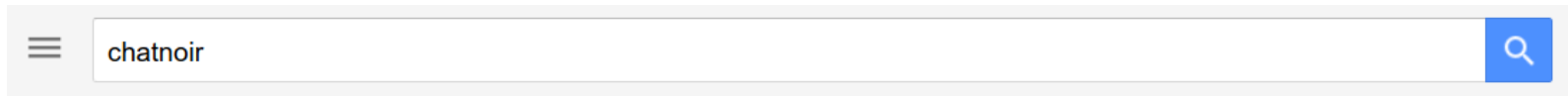
Janek Bevendorff, Benno Stein, Matthias Hagen, and Martin Potthast

Bauhaus-Universität Weimar and Leipzig University  
<first name>.<last name>@uni-weimar.de and martin.pothast@uni-leipzig.de

**Abstract** Elastic ChatNoir<sup>1</sup> is an Elasticsearch-based search engine offering a freely accessible search interface for the two ClueWeb corpora and the Common Crawl, together about 3 billion web pages. Running across 130 nodes, Elastic ChatNoir features subsecond response times comparable to commercial search engines. Unlike most commercial search engines, it also offers a powerful API that is available free of charge to IR researchers. Elastic ChatNoir's main purpose is to serve as a baseline for reproducible IR experiments and user studies for the coming years, empowering research at a scale not attainable to many labs beforehand, and to provide a platform for experimenting with new approaches to web search.

# Query Expansion with Keyqueries

Example: `chatnoir` is a keyquery for Google Scholar



 Scholar About 9.930 results (0,04 sec)

## ChatNoir: a search engine for the ClueWeb09 corpus

[M Potthast](#), [M Hagen](#), [B Stein](#), [J Graßegger](#)... - [Proceedings of the 35th ...](#), 2012 - [dl.acm.org](#)

We present the **ChatNoir** search engine which indexes the entire English part of the ClueWeb09 corpus. Besides Carnegie Mellon's Indri system, **ChatNoir** is the second publicly available search engine for this corpus. It implements the classic BM25F information ...

☆  Cited by 54 [Related articles](#) [All 7 versions](#) 

## Elastic ChatNoir: search engine for the ClueWeb and the common crawl

[J Bevendorff](#), [B Stein](#), [M Hagen](#), [M Potthast](#) - [European Conference on ...](#), 2018 - Springer



Abstract Elastic **ChatNoir** (Search: [www.chatnoir.eu](#) Code: [www.github.com/chatnoir-eu](#)) is an Elasticsearch-based search engine offering a freely accessible search interface for the two ClueWeb corpora and the Common Crawl, together about 3 billion web pages. Running ...

☆  Cited by 12 [Related articles](#) [All 3 versions](#) 

## [book] Le Chat Noir: A Montmartre cabaret and its artists in turn-of-the century Paris

[A Fields](#) - 1993 - [openbibart.fr](#)

Help. Search. XML. Title : Le **Chat noir** : a Montmartre cabaret and its artists in turn-of-the-century Paris ... Includes discussion of some of the shadow theater productions. ; Pagination/Size : 54 ; 54 p.; ill. (some col.); Topic : 530 - BIBLIOGRAPHIE D'HISTOIRE DE ...

☆  Cited by 28 [Related articles](#) 

## Generalized Chat Noir is PSPACE-complete

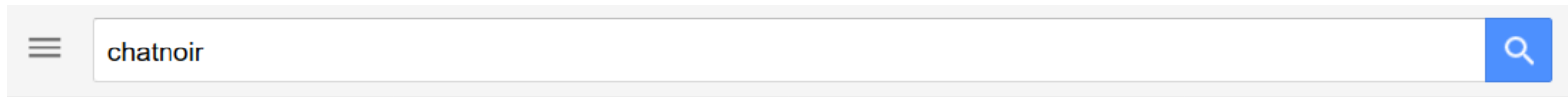
[C Iwamoto](#), [Y Mukai](#), [Y Sumida](#)... - [IEICE transactions on ...](#), 2013 - [search.ieice.org](#)

We study the computational complexity of the following two-player game. The instance is a graph  $G=(V, E)$ , an initial vertex  $s \in V$ , and a target set  $T \subseteq V$ . A "cat" is initially placed on  $s$ . Player 1 chooses a vertex in the graph and removes it and its incident edges from the graph ...

☆  Cited by 3 [Related articles](#) [All 8 versions](#) 

# Query Expansion with Keyqueries

Example: `chatnoir` is a keyquery for Google Scholar



Scholar About 9.930 results (0,04 sec)

## ChatNoir: a search engine for the ClueWeb09 corpus

[M Potthast](#), [M Hagen](#), [B Stein](#), [J Graßegger](#)... - Proceedings of the 35th ..., 2012 - dl.acm.org

We present the **ChatNoir** search engine which indexes the entire English part of the ClueWeb09 corpus. Besides Carnegie Mellon's Indri system, **ChatNoir** is the second publicly available search engine for this corpus. It implements the classic BM25F information ...

☆ [🔗](#) Cited by 54 [Related articles](#) [All 7 versions](#) [🔗](#)

## Elastic ChatNoir: search engine for the ClueWeb and the common crawl

[J Bevendorff](#), [B Stein](#), [M Hagen](#), [M Potthast](#) - European Conference on ..., 2018 - Springer

Abstract Elastic **ChatNoir** (Search: [www.chatnoir.eu](#) Code: [www.github.com/chatnoir-eu](#)) is an Elasticsearch-based search engine offering a freely accessible search interface for the two ClueWeb corpora and the Common Crawl, together about 3 billion web pages. Running ...

☆ [🔗](#) Cited by 12 [Related articles](#) [All 3 versions](#) [🔗](#)

## [book] Le Chat Noir: A Montmartre cabaret and its artists in turn-of-the century Paris

[A Fields](#) - 1993 - openbibart.fr

Help. Search. XML. Title : Le **Chat noir** : a Montmartre cabaret and its artists in turn-of-the-century Paris ... Includes discussion of some of the shadow theater productions. ; Pagination/Size : 54 ; 54 p.; ill. (some col.); Topic : 530 - BIBLIOGRAPHIE D'HISTOIRE DE ...

☆ [🔗](#) Cited by 28 [Related articles](#) [🔗](#)

## Generalized Chat Noir is PSPACE-complete

[C Iwamoto](#), [Y Mukai](#), [Y Sumida](#)... - IEICE transactions on ..., 2013 - search.ieice.org

We study the computational complexity of the following two-player game. The instance is a graph  $G=(V, E)$ , an initial vertex  $s \in V$ , and a target set  $T \subseteq V$ . A "cat" is initially placed on  $s$ . Player 1 chooses a vertex in the graph and removes it and its incident edges from the graph ...

☆ [🔗](#) Cited by 3 [Related articles](#) [All 8 versions](#) [🔗](#)



# Query Expansion with Keyqueries

Example: ... but not for Google

The screenshot shows a Google search interface with the query 'chatnoir' in the search bar. Below the search bar, there are navigation options: All, Images, Maps, Shopping, Videos, More, Settings, and Tools. The search results indicate approximately 216,000,000 results found in 0.71 seconds. The first result is from en.wikipedia.org, titled 'Le Chat Noir - Wikipedia'. The snippet describes it as a nineteenth-century entertainment establishment in the bohemian Montmartre district of Paris, opened on 18 November 1881. Below this result is a 'People also ask' section with four questions: 'What does Le Chat Noir mean?', 'Where is the Chat Noir painting?', 'Who made Le Chat Noir?', and 'What happened to Henri Le Chat Noir?'. The second result is from de.wikipedia.org, titled 'Le Chat Noir – Wikipedia'. The snippet describes it as a popular Parisian cabaret on Montmartre, founded by Rodolphe Salis. The third result is from www.amazon.de, titled 'Close Up Le Chat Noir Poster (61cm x 91,5cm): Amazon.de ...'. The snippet describes it as a poster available for purchase on Amazon.de.

chatnoir

All Images Maps Shopping Videos More Settings Tools

About 216.000.000 results (0,71 seconds)

en.wikipedia.org › wiki › Le\_Chat\_Noir ▾  
**Le Chat Noir - Wikipedia**  
Le Chat Noir was a nineteenth-century entertainment establishment, in the bohemian Montmartre district of Paris. It was opened on 18 November 1881 at 84 ...  
Second site · Last location · Shadow play · References

People also ask

What does Le Chat Noir mean? ▾

Where is the Chat Noir painting? ▾

Who made Le Chat Noir? ▾

What happened to Henri Le Chat Noir? ▾

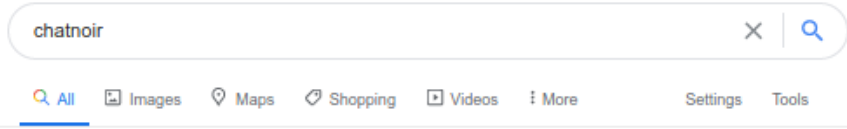
Feedback

de.wikipedia.org › wiki › Le\_Chat\_... ▾ Translate this page  
**Le Chat Noir – Wikipedia**  
Le Chat Noir (deutsch Der schwarze Kater) war von 1881 bis 1897 ein beliebtes Pariser Kabarett auf dem Montmartre, das von Rodolphe Salis gegründet wurde ...  
Anfänge am Boulevard ... · Rue Victor-Masse · Die Zeitschrift Le Chat Noir

www.amazon.de › Close-Up-Chat-N... ▾ Translate this page  
**Close Up Le Chat Noir Poster (61cm x 91,5cm): Amazon.de ...**  
Close Up Le Chat Noir Poster (61cm x 91, 5cm) günstig auf Amazon.de: Kostenlose Lieferung an den Aufstellort sowie kostenlose Rückgabe für qualifizierte ...

# Query Expansion with Keyqueries

Example: ... but not for Google



About 216.000.000 results (0,71 seconds)

en.wikipedia.org › wiki › Le\_Chat\_Noir

## Le Chat Noir - Wikipedia

Le Chat Noir was a nineteenth-century entertainment establishment, in the bohemian Montmartre district of Paris. It was opened on 18 November 1881 at 84 ...

Second site · Last location · Shadow play · References

**No!**

People also ask

- What does Le Chat Noir mean? ▾
- Where is the Chat Noir painting? ▾
- Who made Le Chat Noir? ▾
- What happened to Henri Le Chat Noir? ▾

Feedback

de.wikipedia.org › wiki › Le\_Chat\_... Translate this page

## Le Chat Noir – Wikipedia

Le Chat Noir (deutsch Der schwarze Kater) war von 1881 bis 1897 ein beliebtes Pariser Kabarett auf dem Montmartre, das von Rodolphe Salis gegründet wurde ...

Anfänge am Boulevard ... · Rue Victor-Masse · Die Zeitschrift Le Chat Noir

**No!**

www.amazon.de › Close-Up-Chat-N... Translate this page

## Close Up Le Chat Noir Poster (61cm x 91,5cm): Amazon.de ...

Close Up Le Chat Noir Poster (61cm x 91, 5cm) günstig auf Amazon.de: Kostenlose an den Aufstellort sowie kostenlose Rückgabe für qualifizierte ...

**No!**

# Query Expansion with Keyqueries

## Simulation of explicit relevance feedback

- First  $k$  relevant and helpful documents that appear in the BM25 ranking

## Implementation Details

- Generate vocabulary  $V$  of <term, weight> expansion terms with RM3
- $Q = 2^V \setminus \{\emptyset\}$  are the meaningful queries over  $V$
- Relaxation of the first keyquery criteria if necessary:
  - Retrieve  $|R| - 1, |R| - 2, \dots$  target documents
- Selection of expanded query if there are multiple keyqueries
  - Candidate with the highest nDCG@k for the target documents
- We use a simple brute-force implementation

# Evaluation

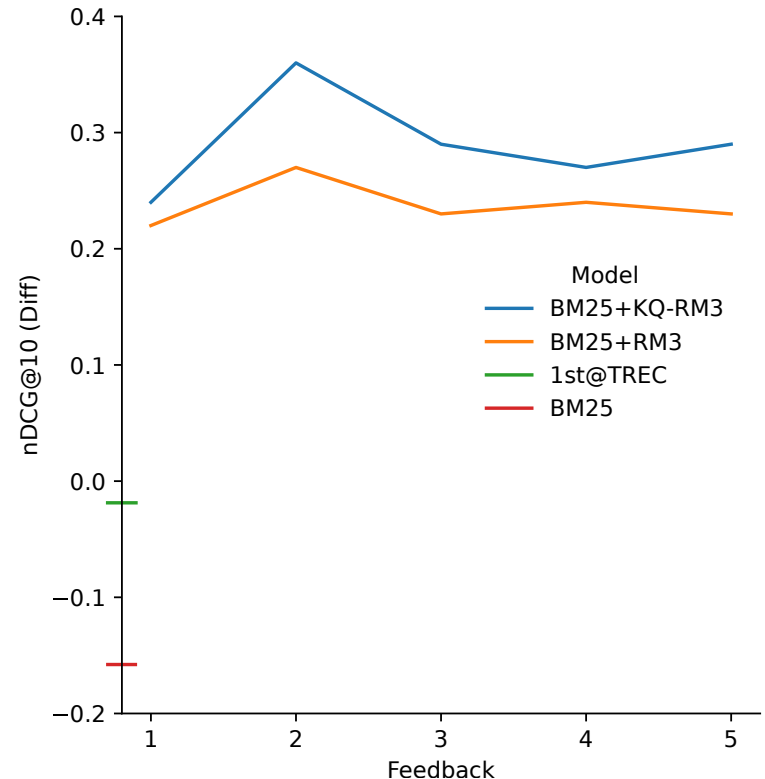
## Setup: TREC 2019–2021 Decision and Health Misinformation Tracks

- ❑ Helpful qrels: Documents are relevant and helpful
- ❑ Harmful qrels: Documents are relevant but harmful
- ❑ Effectiveness measured as difference: Helpful - Harmful

## Results on the TREC 2019 Decision Track

- ❑ Corpus: ClueWeb12 category B
- ❑ 50 judged Topics. E.g.:

Does hypnotherapy help people quit smoking?



# Evaluation

## Setup: TREC 2019–2021 Decision and Health Misinformation Tracks

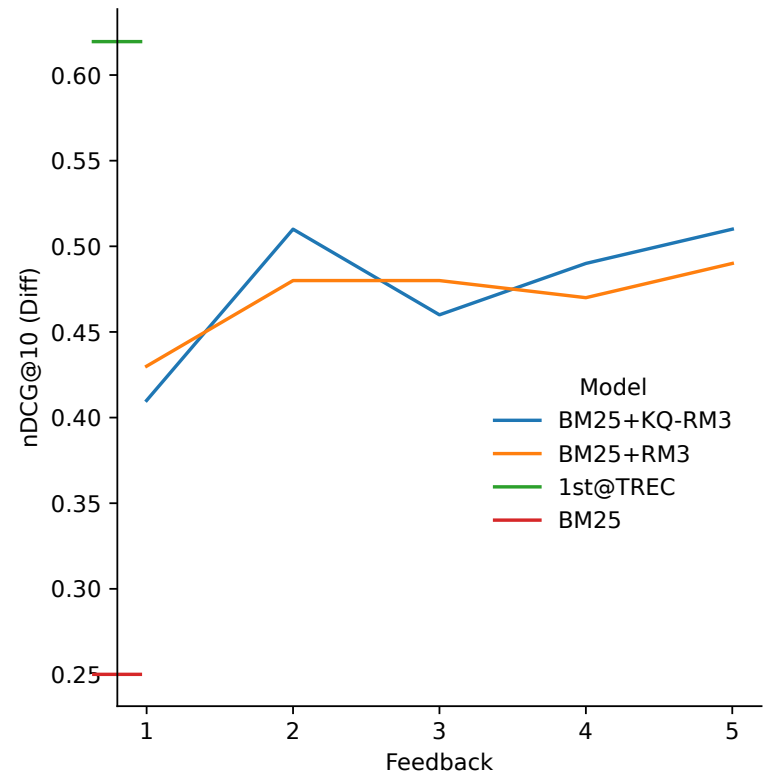
- Helpful qrels: Documents are relevant and helpful
- Harmful qrels: Documents are relevant but harmful
- Effectiveness measured as difference: Helpful - Harmful

## Results on the TREC 2020 Decision Track

- Corpus: Common Crawl News Crawl
- 46 judged Topics. E.g.:

Can ibuprofen worsen COVID-19?

- VERA (1st@TREC)  
[Pradeep et al.; SIGIR'21]
  - Answer as explicit feedback
  - Predict with T5 whether the document supports the answer
  - Interpolated with Mono/DuoT5



# Evaluation

## Setup: TREC 2019–2021 Decision and Health Misinformation Tracks

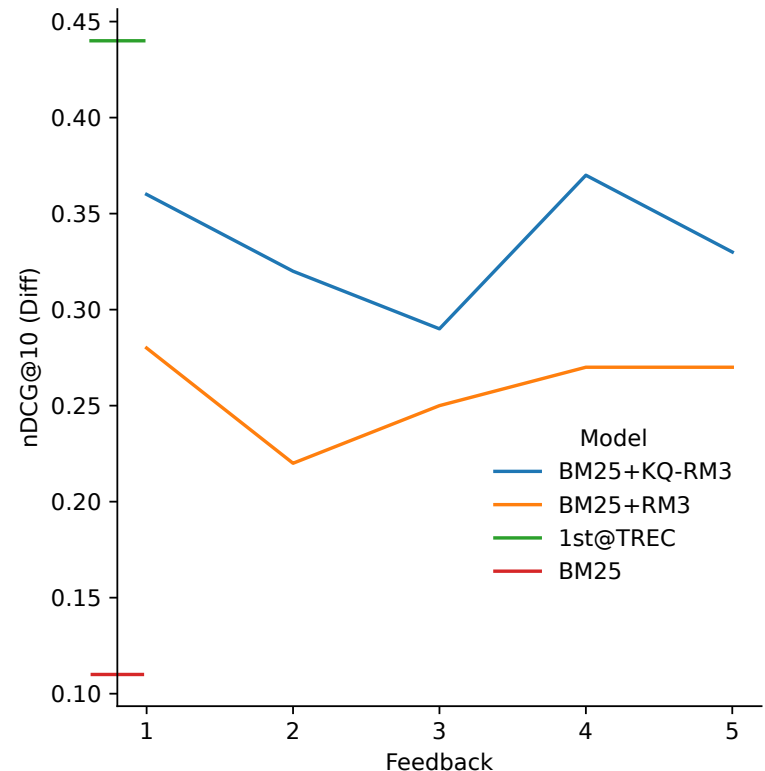
- Helpful qrels: Documents are relevant and helpful
- Harmful qrels: Documents are relevant but harmful
- Effectiveness measured as difference: Helpful - Harmful

## Results on the TREC 2021 Decision Track

- Corpus: noclean subset of C4
- 35 judged Topics. E.g.:

Does selenium help prevent cancer?

- VERA (1st@TREC)  
[Pradeep et al.; SIGIR'21]
  - Difference to
    - BM25+KQ-RM3 not significant
    - BM25+RM3 is significant



# Conclusions

## Summary

- ❑ RM3 query expansions can be enhanced with the concept of keyqueries
  - Leads to more effective BM25 queries: Increased help, decreased harm
- ❑ Explicit feedback is important for health-related searches
  - VERA is the state of the art

# Conclusions

## Summary

- ❑ RM3 query expansions can be enhanced with the concept of keyqueries
  - Leads to more effective BM25 queries: Increased help, decreased harm
- ❑ Explicit feedback is important for health-related searches
  - VERA is the state of the art

## Future work

- ❑ Expand study:
  - More relevance feedback approaches, query expansion approaches, retrieval models
- ❑ Replace brute-force implementation
  - Efficient enumeration or reverted indexes
- ❑ Replace costly relevance feedback with information available on the Web?
  - E.g., [schema.org/ClaimReview](http://schema.org/ClaimReview)



# Conclusions

## Summary

- ❑ RM3 query expansions can be enhanced with the concept of keyqueries
  - Leads to more effective BM25 queries: Increased help, decreased harm
- ❑ Explicit feedback is important for health-related searches
  - VERA is the state of the art

## Future work

- ❑ Expand study:
  - More relevance feedback approaches, query expansion approaches, retrieval models
- ❑ Replace brute-force implementation
  - Efficient enumeration or reverted indexes
- ❑ Replace costly relevance feedback with information available on the Web?
  - E.g., [schema.org/ClaimReview](http://schema.org/ClaimReview)

*thank you!*