Knowledge-based Counter-Claim Generation

Milad Alshomary and Felix Lange

10.03.2023

Leibniz Universität Hannover

Introduction

- **Task:** Given an input claim, generate a counter claim to it.
  - There exists research on counter-claim generation *(Hidey and McKown 2019, Bilu et al 2015)*

Introduction

- **Task:** Given an input claim, generate a counter claim to it.
  - There exists research on counter-claim generation *(Hidey and McKown 2019, Bilu et al 2015)*
  - Linquistic approaches (depending on either rules or learned from data) to apply a set of edits in order to reach a counter.

Introduction

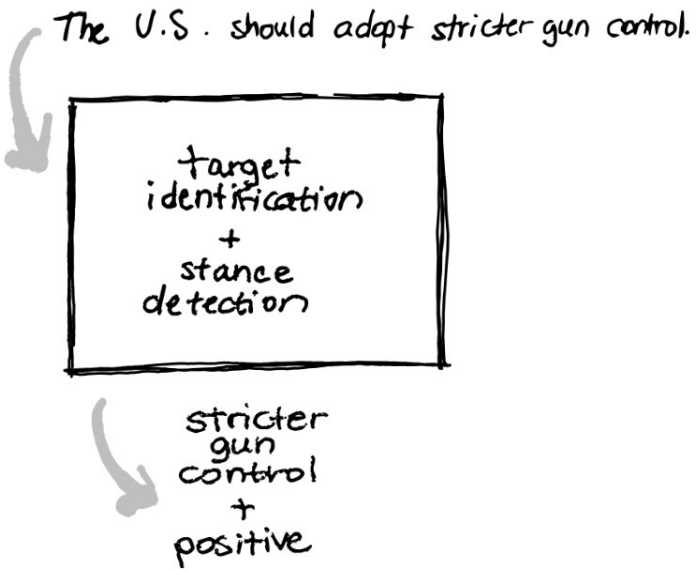
- **Task:** Given an input claim, generate a counter claim to it.
  - There exists research on counter-claim generation *(Hidey and McKown 2019, Bilu et al 2015)*
  - Linquistic approaches (depending on either rules or learned from data) to apply a set of edits in order to reach a counter.
- **However:**
  - Generating Counter claims relies heavily on commonsense knowledge.
  - Empowering language models with commonsense knowledge prove to work

Introduction

- **Task:** Given an input claim, generate a counter claim to it.
  - There exists research on counter-claim generation *(Hidey and McKown 2019, Bilu et al 2015)*
  - Linquistic approaches (depending on either rules or learned from data) to apply a set of edits in order to reach a counter.
- **However:**
  - Generating Counter claims relies heavily on commonsense knowledge.
  - Empowering language models with commonsense knowledge prove to work

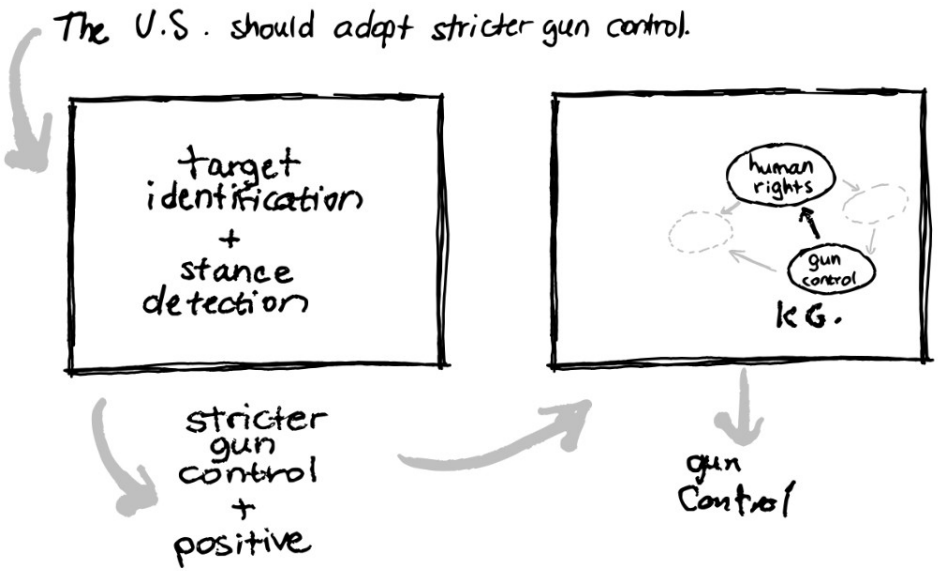
➔ In the following we will propose a simple a pporach to generate counter claims utilizing commonsense knowledge

Approach



I. **Claim analysis:** Extracting claim target and the stance towards it

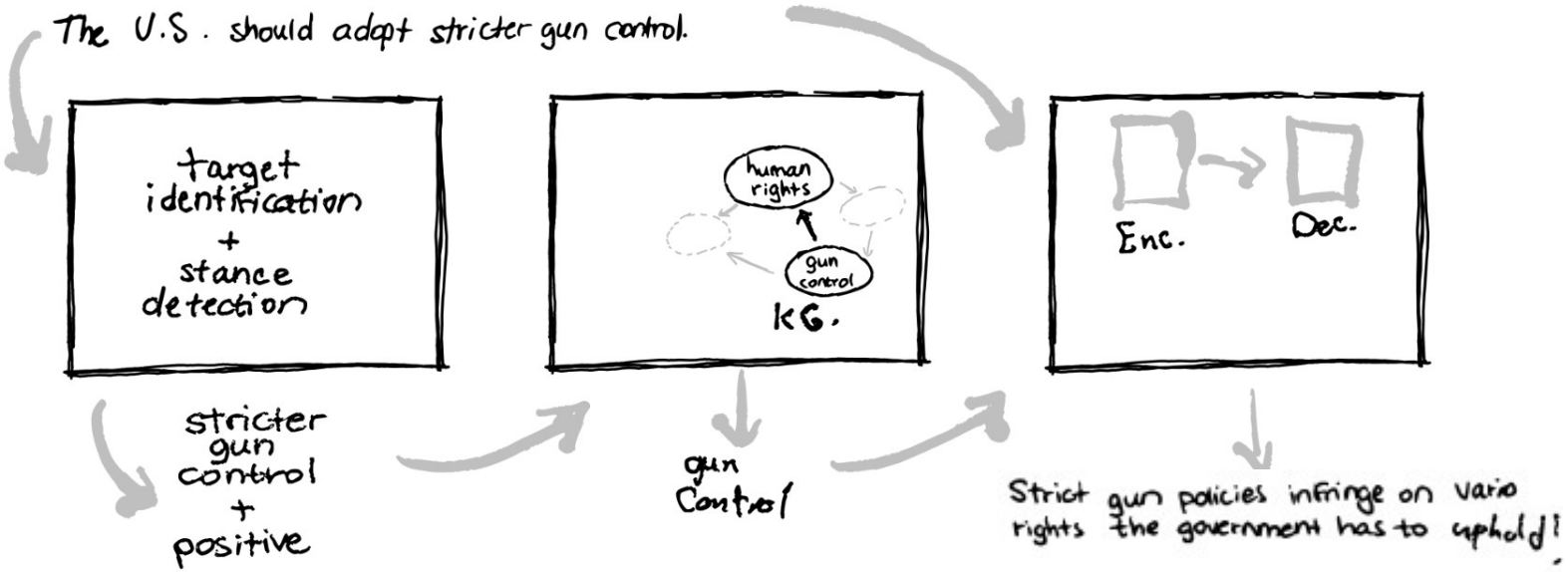
Approach



I. **Claim analysis:** Extracting claim target and the stance towards it

II. **Knowledge Extraction:** Finding appropriate knowledge path „Gun Control supresses human rights“

Approach



I. **Claim analysis:** Extracting claim target and the stance towards it

II. **Knowledge Extraction:** Finding appropriate knowledge path „Gun Control supresses human rights“

III. **Controlled Claim Generatoin:** Generate a counter claim based on the extracted knowledge path

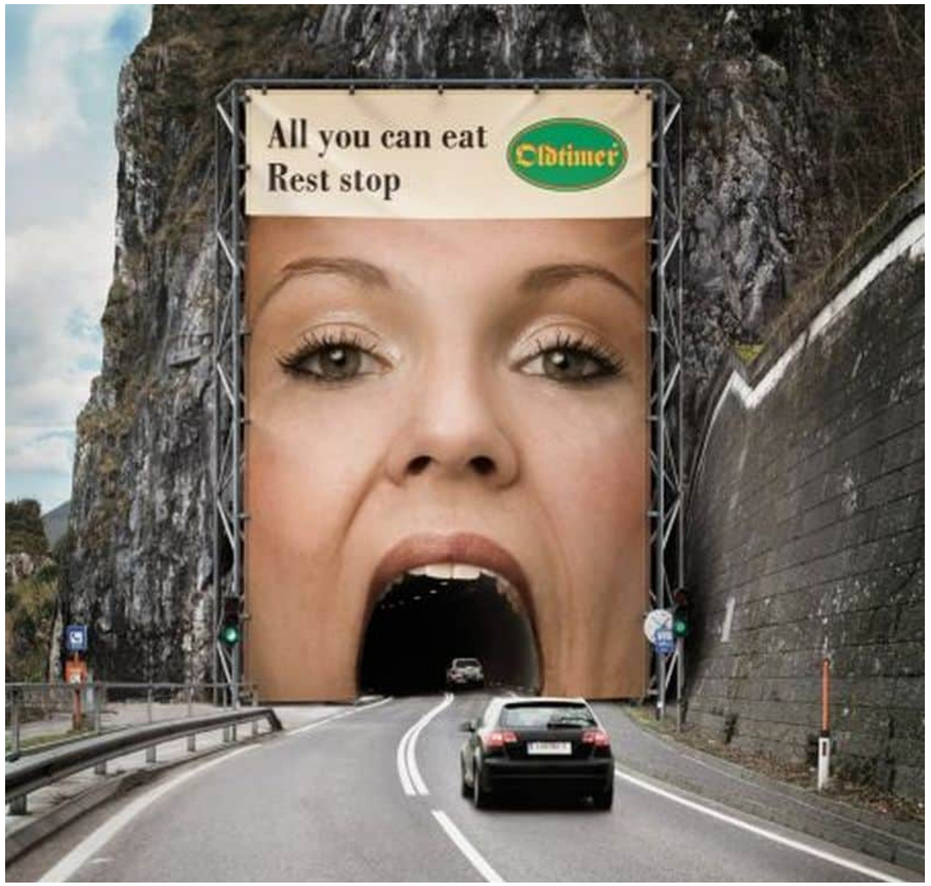
Thank You

Classifying Humans

Theresa Elstner, März 2023

Examples of Classifiers for Humans

Advertisment: Whom to show what?



https://drprem.com/digital-marketing/oldtimer-restaurants-all-you-can-eat-rest-stop

Examples of Classifiers for Humans

Banking: Whom to give a loan?

https://www.spiegel.de/kultur/kino/der-hobbit-smaugs-einoede-mittelerde-strahlt-im-mittelteil-a-938083.html

Examples of Classifiers for Humans

Hiring: Whom to invite to an interview?

https://www.insider.com/kim-kardashian-said-she-had-never-seen-snl-before-hosting-2022-6

Examples of Classifiers for Humans

...

Algorithmic Fairness

Why?

- Classification of humans through humans is biased

Algorithmic Fairness

Why?

- Classification of humans through humans is biased
- Classification of humans through algorithms is

Algorithmic Fairness

Why?

- Classification of humans through humans is biased
- Classification of humans through algorithms is ***ALSO BIASED***

Algorithmic Fairness

Why?

- Classification of humans through humans is biased
- Classification of humans through algorithms is ***ALSO BIASED***
  - Algorithmic fairness should change that as much as possible

Algorithmic Fairness

Concepts

- Group Fairness
  - relates to statistics of classification outcomes among members of protected groups
  - e.g.: Protected groups should have the same number of bad classification outcomes than other groups.
  - Does not always guarantee fairness
- Individual Fairness
  - Treat similar individuals similarly with respect to a task

Unfairness

Unequal Outcomes/Ignorance of Decision Maker

https://www.youtube.com/watch?v=-KSryJXDpZo

Implementing Unfairness-Aware Algorithms

Let humans stand up for themselves

- Not implemented in algorithmic fairness approaches yet:
  - Feedback loops that allow humans to:
    - Notify the algorithm about unfairness
    - Correct their own algorithmic representation

Thanks.

Ernesto Priego

@ernestopriego

For instance, if you are referencing an arxiv pre-print for example, please link to the output via its arxiv.org/abs/ location (abs stands for abstract; it would be followed by DOI number), not to arxiv.org/pdf/ (also followed by DOI number)

10:12 AM · Mar 2, 2021

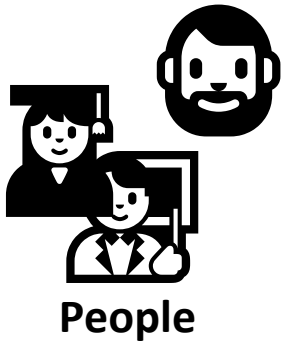
Looking for /abs/
Linking Authors to Twitter

Webis Flash Talks 2023

Sebastian Günther

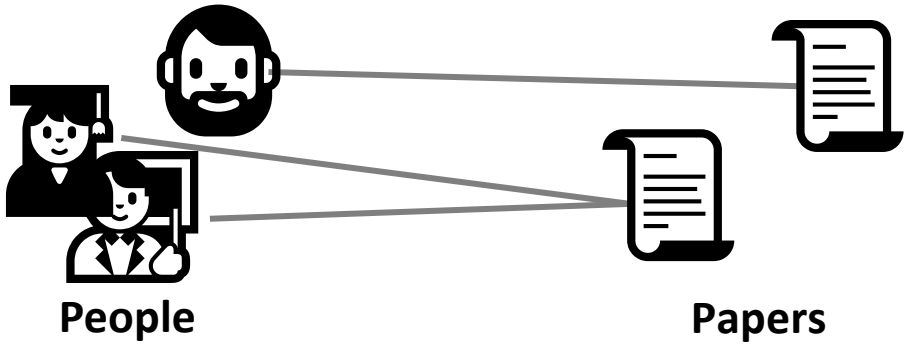
The initial idea…

Initial Idea



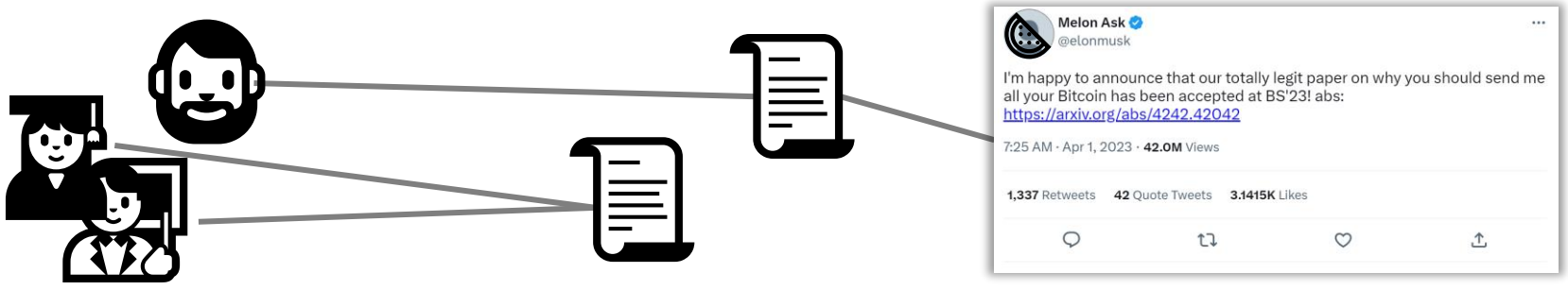
Initial Idea

- People publish papers



Initial Idea

- People publish papers
- People tweet about papers



People **Papers** **Tweet**
(with link to paper)

Initial Idea

- People publish papers
- People tweet about papers
- Tweets have authors

People **Papers** **Tweet** (with link to paper) **Twitter Profile**

Initial Idea

- People publish papers
- People tweet about papers
- Tweets have authors

But wait, there‘s more!

More Information

- Summarization by the author
- @mentioning co-authors & project partners

Sagnik Majumder @sagnikmjr · Mar 6

Replying to @sagnikmjr

Key insight: egocentric audio-visual observations of participants in a natural conversation can be used to map a 3D scene in a cost-efficient way, such that the visual capture cost can be kept to a minimum by relying on audio for mapping in moments of visual redundancy.

(2/n)

1 62

Sagnik Majumder @sagnikmjr · Mar 6

Our approach: use an audio-visual mapper for scene mapping and an audio-visual RL policy to decide when to capture or skip a visual frame.

(3/n)

1 36

Sagnik Majumder @sagnikmjr · Mar 6

Results: our model produces strong results in both simulation and real world. With full visual capture, it improves over previous SOTA scene mappers. When actively caputring useful frames, it sees a small drop in mapping quality, while sharply reducing the capture cost.

(4/n)

1 57

Sagnik Majumder @sagnikmjr · Mar 6

Work done in collaboration with @RealityLabs and @UTCompSci.

(n/n)

44

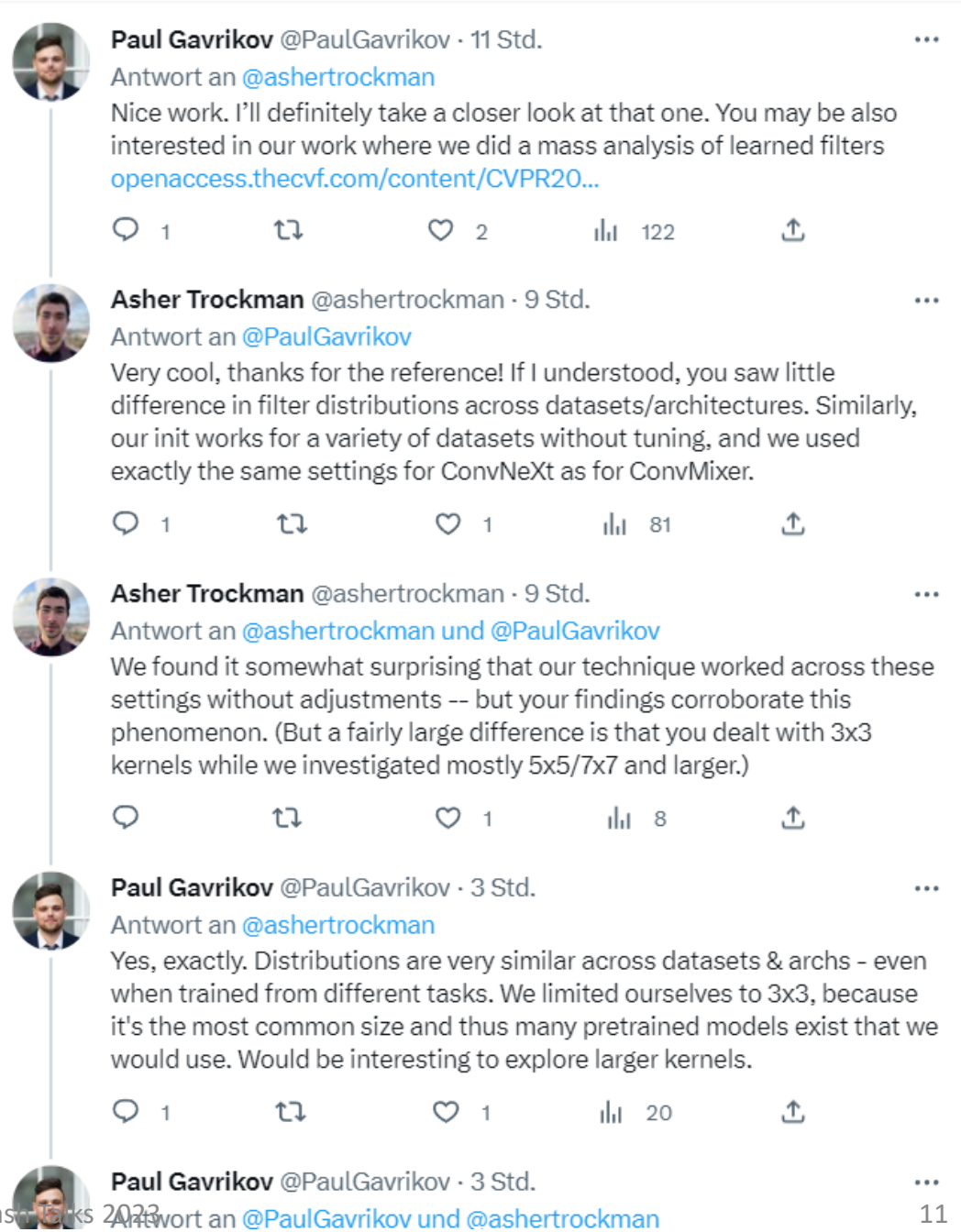
More Information

- Summarization by the author
- @mentioning co-authors & project partners

- Github (code), project website

More Information

- Summarization by the author
- @mentioning co-authors & project partners
- Github (code), project website
- Threads: discussions, explanations



Paul Gavrikov @PaulGavrikov · 11 Std.
Antwort an @ashertrockman
Nice work. I'll definitely take a closer look at that one. You may be also interested in our work where we did a mass analysis of learned filters openaccess.thecvf.com/content/CVPR20...
1 2 122

Asher Trockman @ashertrockman · 9 Std.
Antwort an @PaulGavrikov
Very cool, thanks for the reference! If I understood, you saw little difference in filter distributions across datasets/architectures. Similarly, our init works for a variety of datasets without tuning, and we used exactly the same settings for ConvNeXt as for ConvMixer.
1 1 81

Asher Trockman @ashertrockman · 9 Std.
Antwort an @ashertrockman und @PaulGavrikov
We found it somewhat surprising that our technique worked across these settings without adjustments -- but your findings corroborate this phenomenon. (But a fairly large difference is that you dealt with 3x3 kernels while we investigated mostly 5x5/7x7 and larger.)
1 8

Paul Gavrikov @PaulGavrikov · 3 Std.
Antwort an @ashertrockman
Yes, exactly. Distributions are very similar across datasets & archs - even when trained from different tasks. We limited ourselves to 3x3, because it's the most common size and thus many pretrained models exist that we would use. Would be interesting to explore larger kernels.
1 1 20

Paul Gavrikov @PaulGavrikov · 3 Std.
Antwort an @PaulGavrikov und @ashertrockman

But it's not that easy!

Issues

- Most papers: not mentioned
- People tweet about other‘s papers
- Twitter bots
- Broken links

Attack vectors:

- Impersonating people

How it‘s supposed to work

Approach

- Follow leads
  - Initial lead: arXiv id
  - Next: tweets with the arXiv „abs“/“pdf“ link
  - Next: tweet authors, answers, mentions
  - ...result: knowledge graph
- Credibility assessment per item + author
  - Score: 0.0 (no match) – 1.0 (match)
  - String similarity, # of tweets with arXiv links
  - Total credibility: multiply along the path

Does it work then!?

Does it work?


```
{ "errors": [ {
    "code": 215,
    "message": "Bad Authentication data."
} ] }
```

Thanks, Elon

Does it work?


```
{ "errors": [ {
    "code": 215,
    "message": "Bad Authentication data."
} ] }
```

Thanks, Elon

- It‘s a lot of manual work without the API
- Results match with what I remember from last year

Tweet Coverage for CS.IR

- For the last 7 papers:
  - 2023 (5), 2022 (1), 2021 (1)
  - 3 papers: no tweet
  - 3 papers: bot tweets (1x, 2x, 3x)
  - 1 paper: 2x non-author tweets, 1x bot tweet
  - More hits using „/pdf/“ or the title

- But: it gets better over time!
- General rule: older paper = more tweets

Twitter Profiles for Authors in arXiv CS.IR

- 16 authors in 5 papers
  - Match by name
  - Analyze Twitter bio and recent tweets
  - 12 author profiles found

- 70-80% hit rate
  - Depends on recent accepts (i.e., less for asian authors)

Conclusion

- Authors‘ Twitter profiles can be discovered by their preprint tweets
- Twitter threads may contain useful information

Conclusion

- Authors‘ Twitter profiles can be discovered by their preprint tweets
- Twitter threads may contain useful information

Thanks!

Automatic Image Captioning

Image Retrieval for Arguments using CLIP

Alice Rahn, Christiane Gütter, Maaz Hussain, Nazifa Kazimi,

Fatihah Hakiem

Project Leaders: Johannes Kiesel, Tim Gollub

Bauhaus Universität Weimar

Task Specification

- *Question:* Does pineapple belong on pizza?
- *Obvious solution:* Look up image of delicious, cheesy pizza covered in sweet pineapple chunks
- *Problem:* Is that pizza pro-pineapple or con-pineapple?

Picture generated using

`https://deepai.org/machine-learning-model/text2img`

Touché: Image Retrieval For Arguments

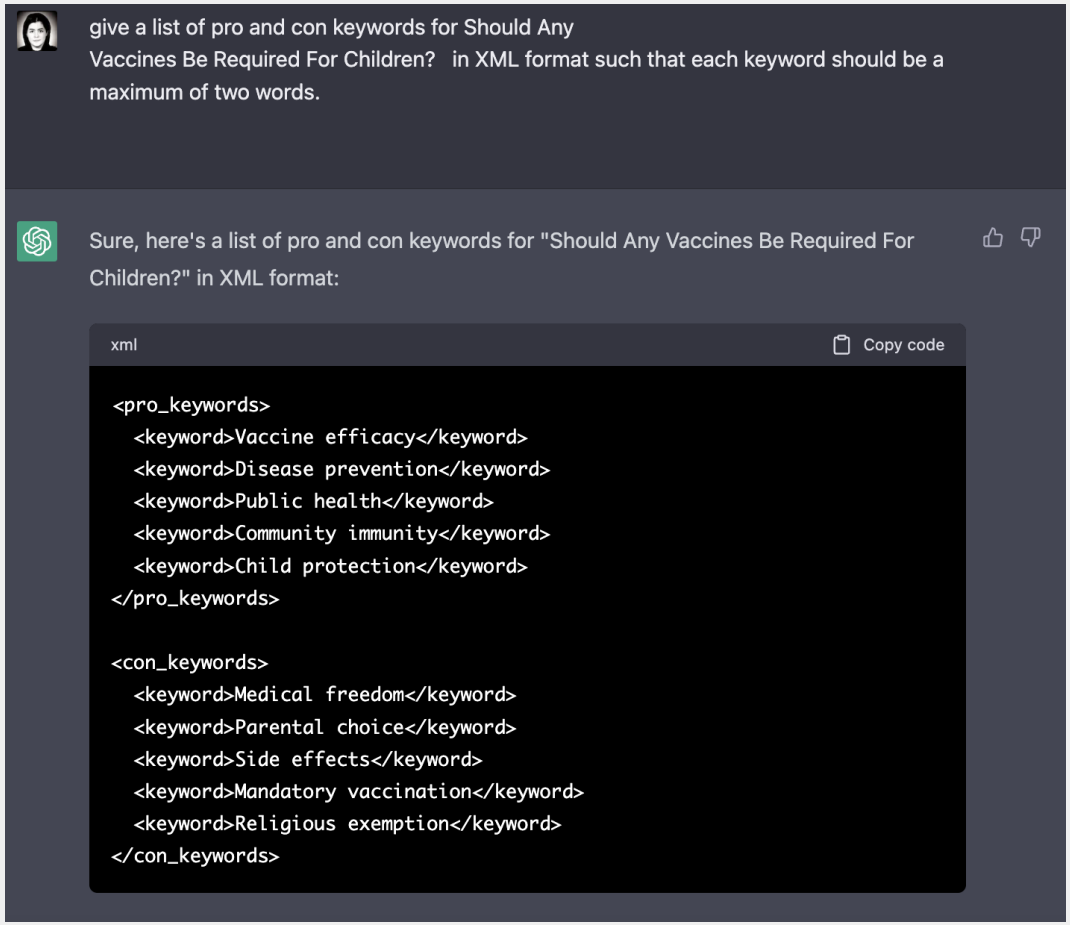
- 50 given topics in the form of a controversial question
- *Example:* Should children have mobile phones? (Or does pineapple belong on pizza…)
- Dataset of images that are found on Google images when looking up those topics
- *Task:* for each image, determine whether it is relevant for that topic, and whether it supports or opposes the question asked

Current Approach

- CLIP model to relate "labels" to the images
- Two runs:
  1. Determine whether image is on-topic
  2. Determine whether image is for or against the given statement
- Topics given as entire question (single label, considered on-topic at a certain threshold)
- Stances given as several (topic related) keywords per label, one representing "pro", one representing "con", higher score of the two is final stance
- result evaluated using a human annotated ground truth

Current State

- Topic relevance determined using similarity score of a question to the image → pretty accurate already (next slide)
- First attempts for stance classification: generation of keywords for all 50 topics using ChatGPT



give a list of pro and con keywords for Should Any Vaccines Be Required For Children? in XML format such that each keyword should be a maximum of two words.

Sure, here's a list of pro and con keywords for "Should Any Vaccines Be Required For Children?" in XML format:

xml Copy code

```
<pro_keywords>
  <keyword>Vaccine efficacy</keyword>
  <keyword>Disease prevention</keyword>
  <keyword>Public health</keyword>
  <keyword>Community immunity</keyword>
  <keyword>Child protection</keyword>
</pro_keywords>

<con_keywords>
  <keyword>Medical freedom</keyword>
  <keyword>Parental choice</keyword>
  <keyword>Side effects</keyword>
  <keyword>Mandatory vaccination</keyword>
  <keyword>Religious exemption</keyword>
</con_keywords>
```

Current Results

Date	Topic Relevance	Argumentativeness	Stance Relevance	Note	Dataset	Model-Type	Input
06.02.2023	0.806	0.702	0.382	Random stance	Ground-Truth	OpenAI Base	Topic Title
06.02.2023	0.847	0.749	0.406	Random stance	Ground-Truth	OpenAI Large	Topic Title
06.02.2023	0.824	0.734	0.398	Random stance	Ground-Truth	LAION Base	Topic Title
06.02.2023	0.834	0.738	0.383	Random stance	Ground-Truth	LAION H	Topic Title

Our Highest Results (Random Stance Estimation For Now)

TEAM	TAG	PRECISION@10		
		ON TOPIC	ARGUMENTATIVE	ON STANCE
Boromir	BERT, OCR, query-processing	0.878	0.768	0.425
Boromir	BERT, OCR, clustering, query-preprocessing	0.822	0.728	0.411
Boromir	AFINN, OCR	0.814	0.726	0.408
Minsc	Baseline	0.736	0.686	0.407

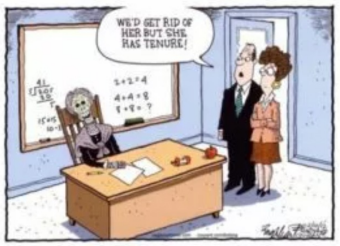
Highest Results From Last Year's Participants

https://touche.webis.de/clef22/touche22-web/image-retrieval-for-arguments.html#submission

Future Plans

- Fully implement stance classification, using mentioned approach
- Base line: Using two template labels for all images (not topic related)

Future Plans



Possible class names (comma-separated)

Long-term investment, Discourages new teachers

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: cached

Discourages new teachers 0.985

Long-term investment 0.015

Example Topic 1: Should Teachers Get Tenure?

Possible class names (comma-separated)

Disease prevention, Side effects

Compute

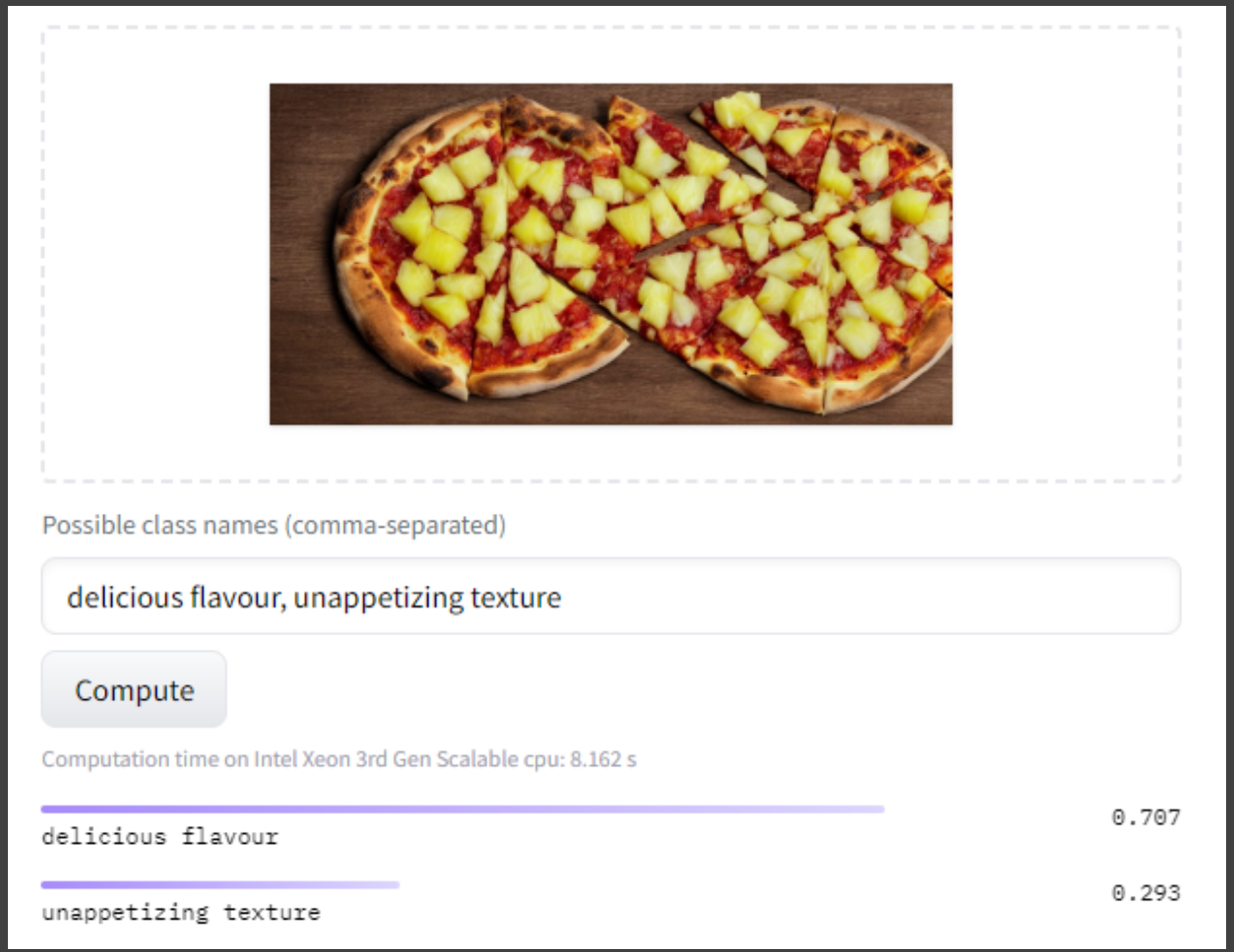
Computation time on Intel Xeon 3rd Gen Scalable cpu: 3.484 s

Disease prevention 0.972

Side effects 0.028

Example Topic 2: Should Any Vaccines Be Required For Children?

But what does CLIP say about pineapple on pizza?



API used: https://huggingface.co/openai/clip-vit-large-patch14


Argumentation Now!
Maximilian Heinrich
2023-03-10

What is Argumentation?

Created by Dall-E, Prompt: *cats having a heavy argument*

Intro Image by Dall-E: Prompt: *fire red bright sky with a giant full moon rising*

What is Argumentation?

communicative activity of producing and exchanging reasons in order to support claims or defend/challenge positions, especially in situations of doubt or disagreement (Lewiński & Mohammed 2016).

Lewiński, Marcin and Dima Mohammed, 2016, "Argumentation Theory", in *The International Encyclopedia of Communication Theory and Philosophy*, Klaus Bruhn Jensen, Robert T. craig, Jefferson Pooley, and Eric W. Rothenbuhler (eds.), Hoboken, NJ: Wiley.

What is good Argumentation?

- Persuasive?
- Emotional?
- Logical? Valid?
- Topic Related?

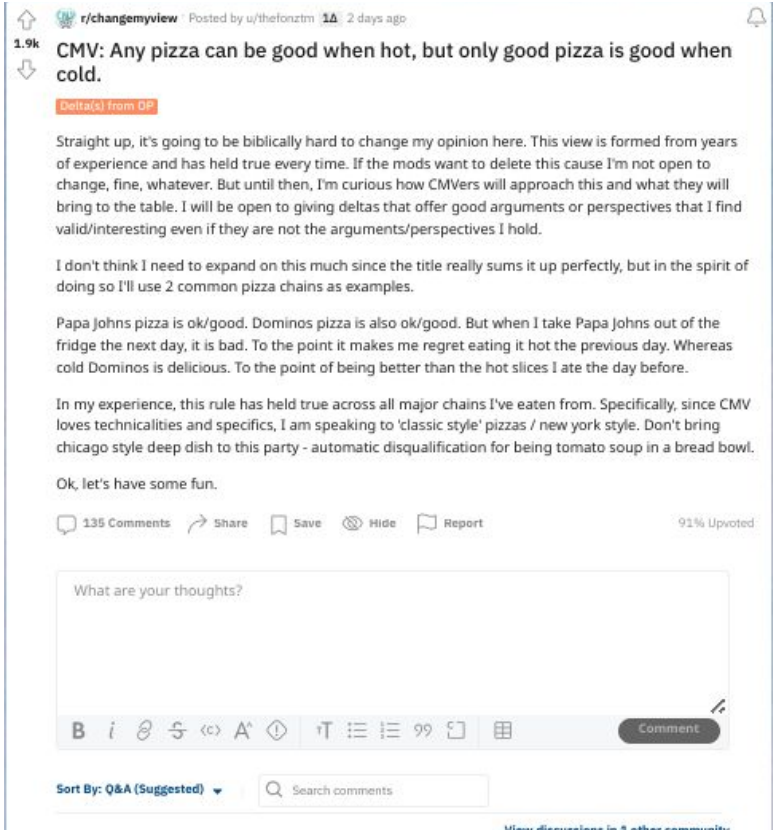
different realms have different methods

Goal - Search for Axioms

Axioms:

- local constraints that capture individual aspects of rhetorically-savvy behaviour
- use as shortcuts or *cooking recipes* for finding good, convincing etc. arguments

Argumentation on Change My View (CMV)



- [subreddit](#) on Reddit
- OriginalPoster (OP) states a claim with premises
- users try to change the position of the OP
- if attempt is successful, a delta is awarded

- discussion is moderated
- data is publicly available
- CMV Argumentation Goal: Persuasion

Effects of Persuasion - Starting Point for Axioms

Sympathy

Reciprocity

Social Proof

Authority

Cialdini, Robert B. 2021. *Influence, New and Expanded: The Psychology of Persuasion.*

Effect of Sympathy

sympathy in form of similarity

- similar experience or background story
- same writing style
- similarity of the username can be influential

Effect of Reciprocity

do sth. and get sth.

- influence of concessions as partial admissions
- a statement cannot be refused too many times
- one does not get persuaded by the first post

Effect of Social Proof

6 out of 7 students want to give a Flash Talks!

or

1 out 7 students does not like Flash Talks!

Effect of Social Proof

look what others are doing

- statistics with higher numbers are more convincing
- posts with a lot of upvotes more are persuasive

Effect of Authority

reference to respected persons

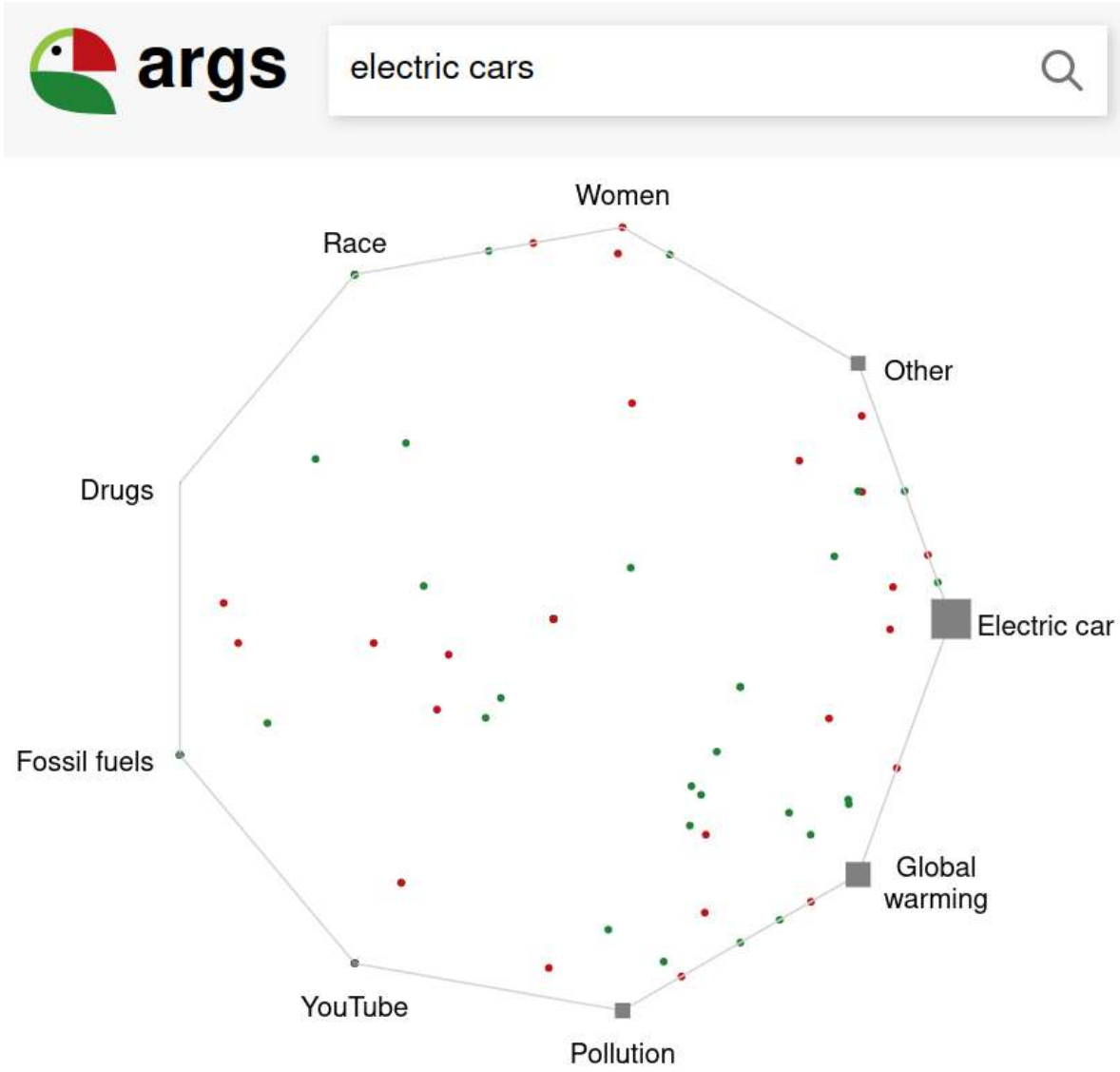
- users, who state that they have professional experience with the topic
- use of links and sources
- delta score of the user

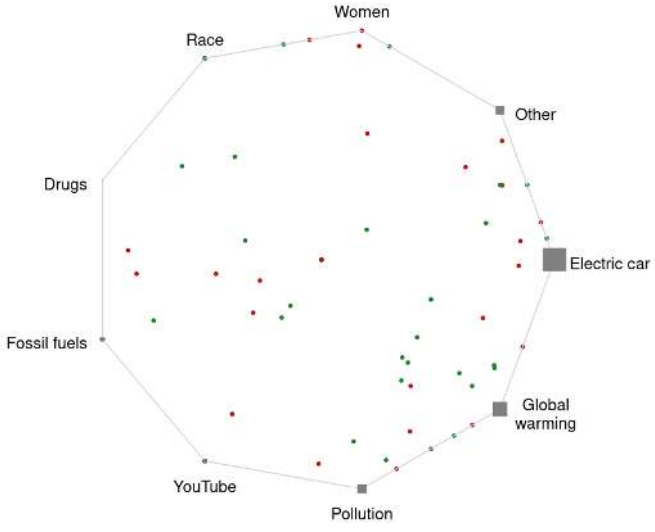
Conclusion

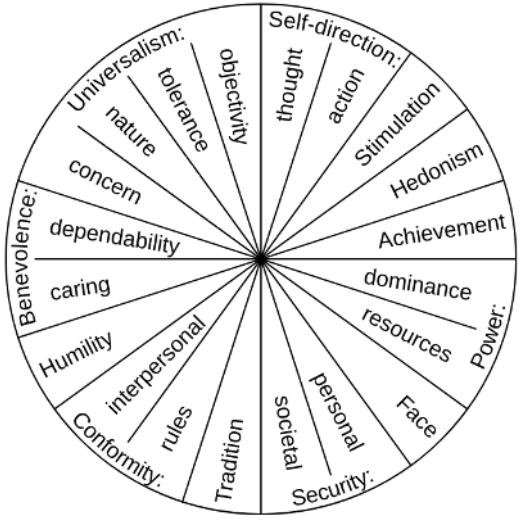
- use a computational approach in order to understand patterns of human behavior
- combine axioms with of other existing metrics, e.g. structure of Elementary Discourse Units (EDUs)
- motivation to use insights for args.me

Thank You!

Created by Dall-E, Prompt: *cats having a peaceful argument*


args
electric cars
Women
Race
Other
Drugs
Electric car
Fossil fuels
Global warming
YouTube
Pollution

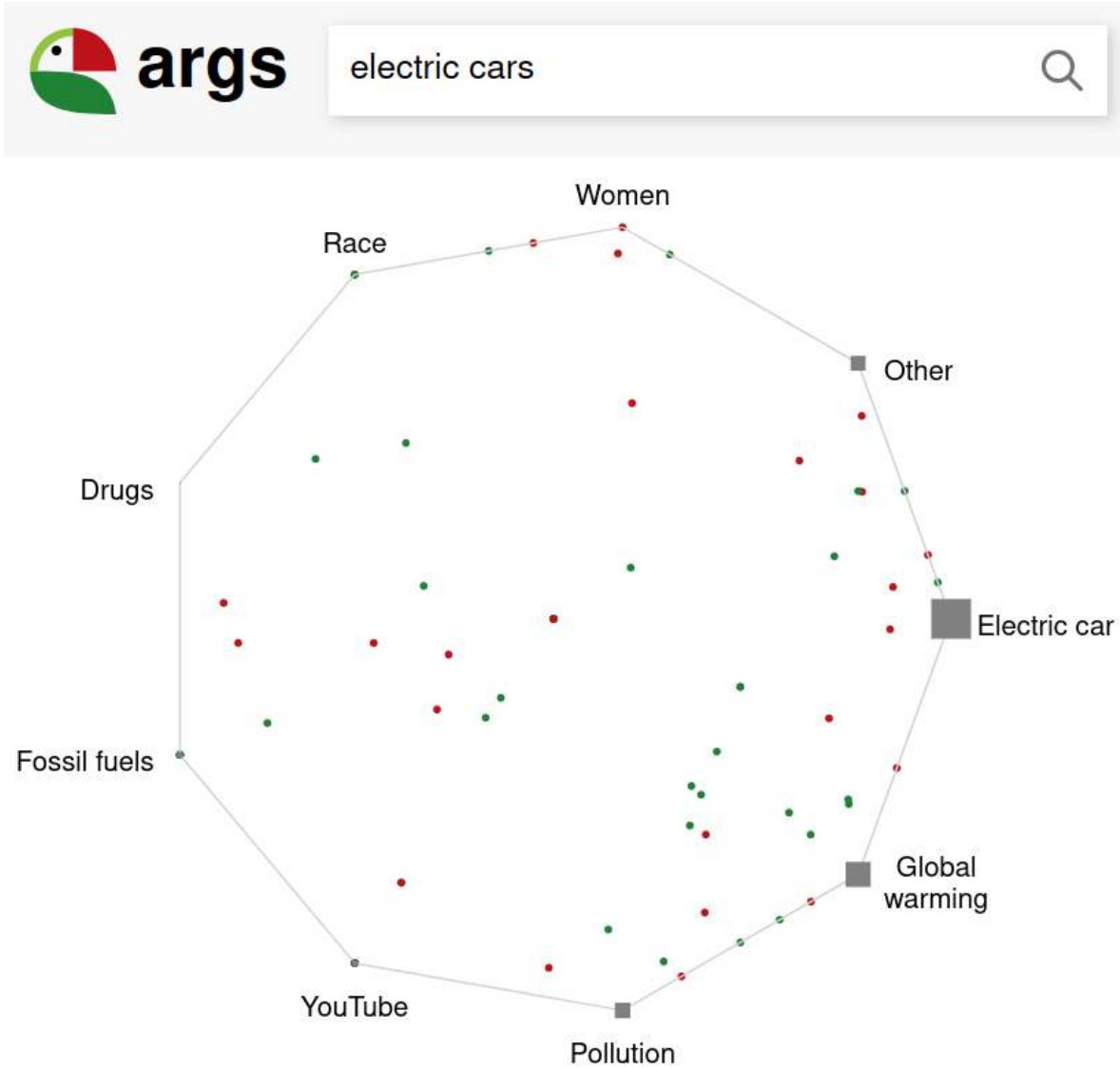


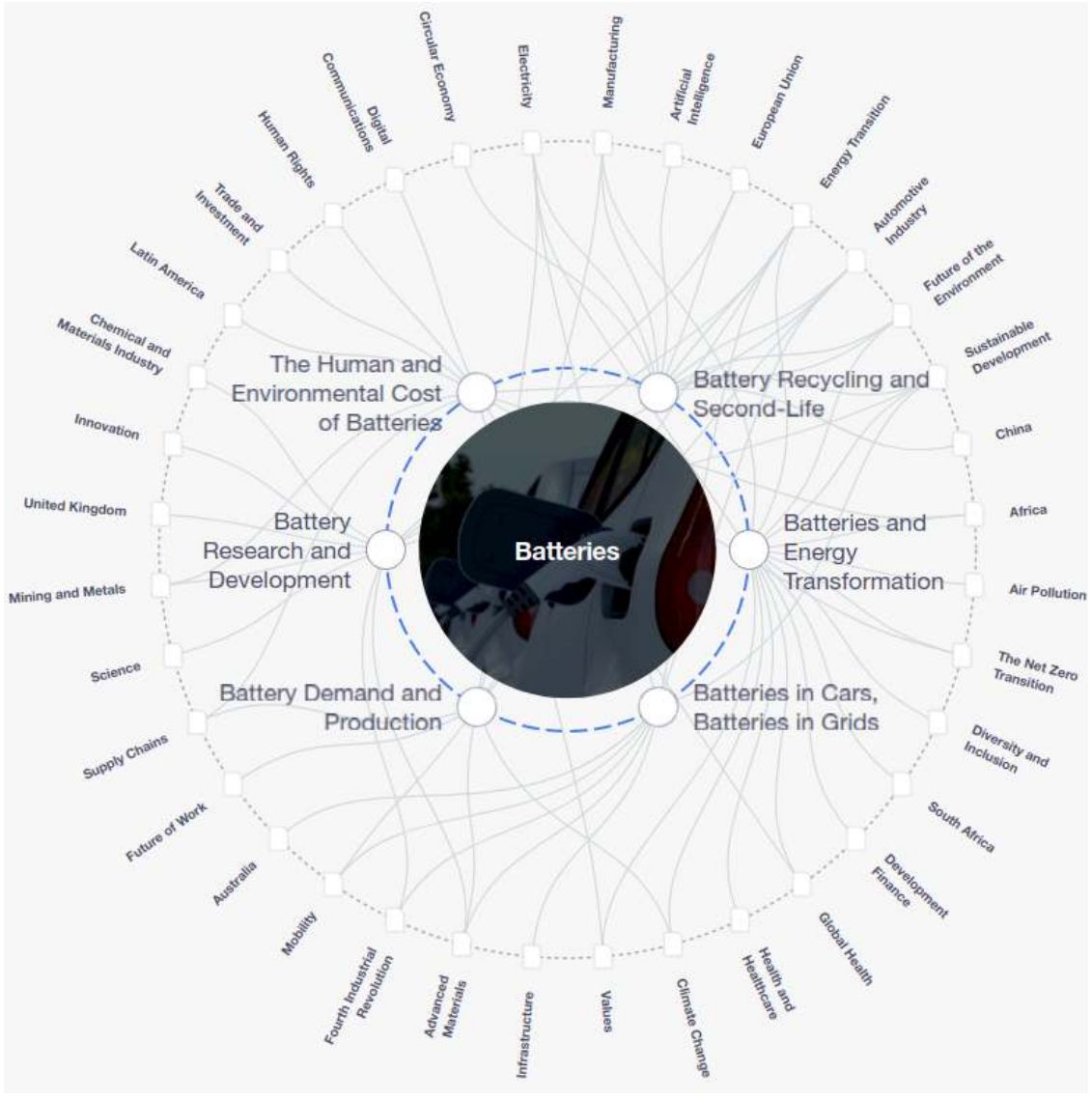


Human Value Detection: A Research Story

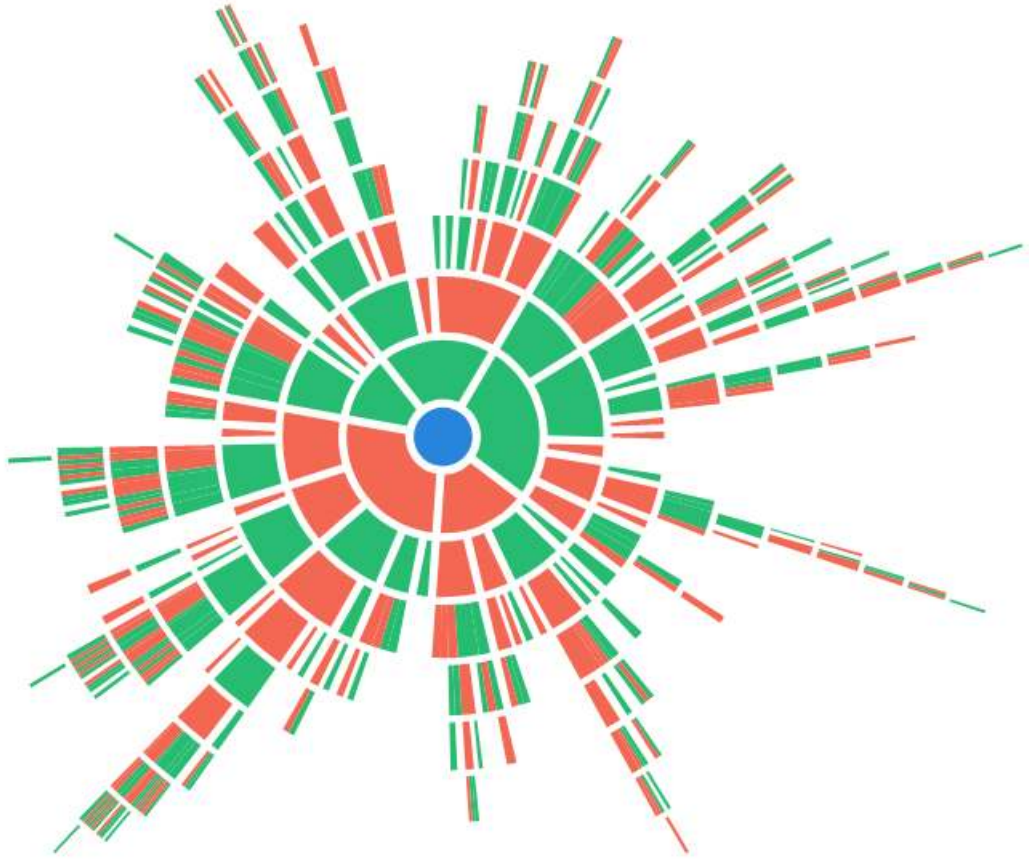
Johannes Kiesel

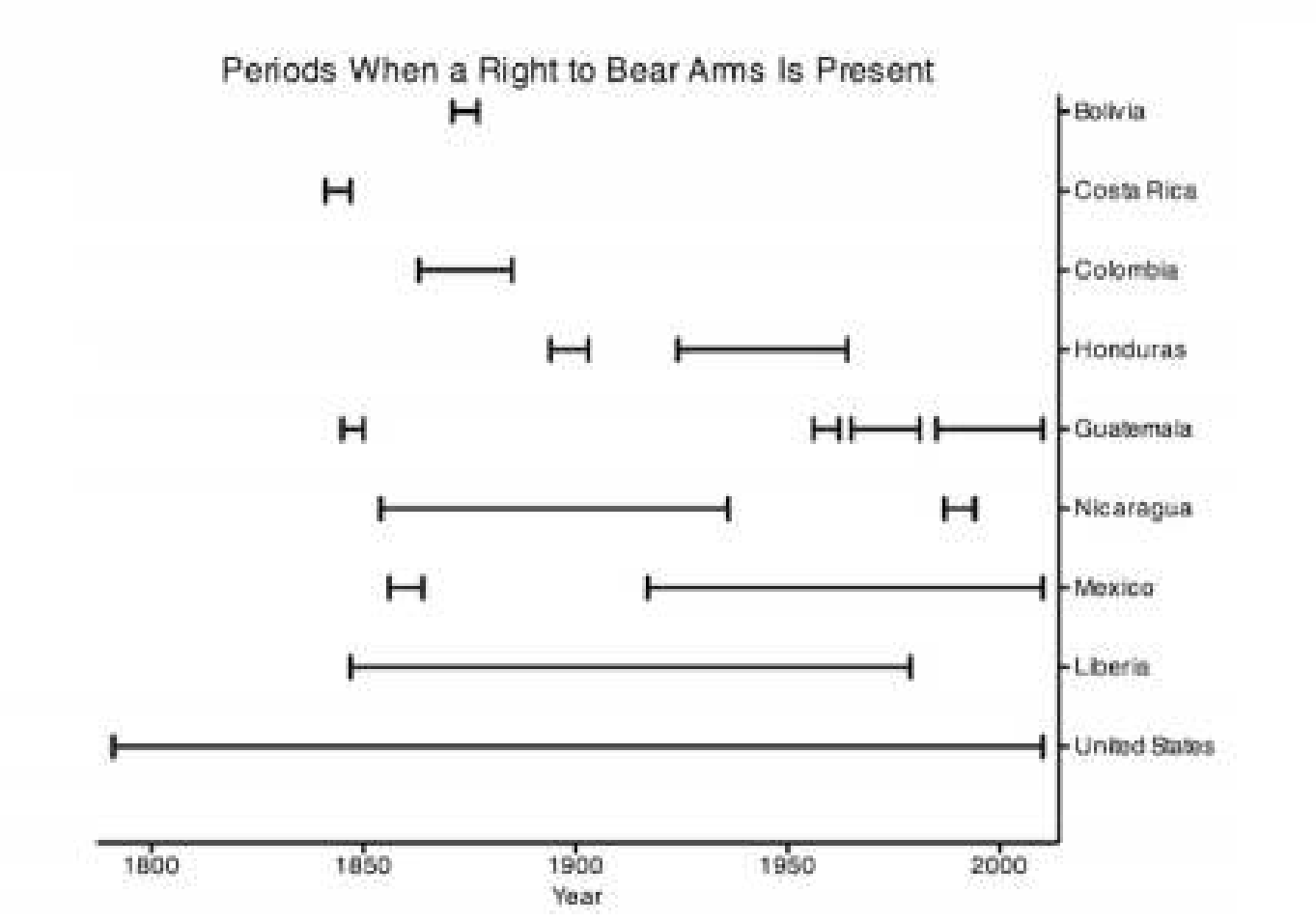
Webis Flash Talks 2023


args
electric cars
Women
Race
Other
Drugs
Electric car
Fossil fuels
Global warming
YouTube
Pollution


Batteries
The Human and Environmental Cost of Batteries
Battery Recycling and Second-Life
Battery Research and Development
Batteries and Energy Transformation
Battery Demand and Production
Batteries in Cars, Batteries in Grids
Circular Economy
Electricity
Manufacturing
Artificial Intelligence
European Union
Energy Transition
Automotive Industry
Future of the Environment
Sustainable Development
China
Africa
Air Pollution
The Net Zero Transition
Diversity and Inclusion
South Africa
Development Finance
Global Health
Health and Healthcare
Climate Change
Values
Infrastructure
Advanced Materials
Fourth Industrial Revolution
Mobility
Australia
Future of Work
Supply Chains
Science
Mining and Metals
United Kingdom
Innovation
Chemical and Materials Industry
Latin America
Trade and Investment
Human Rights
Digital Communications

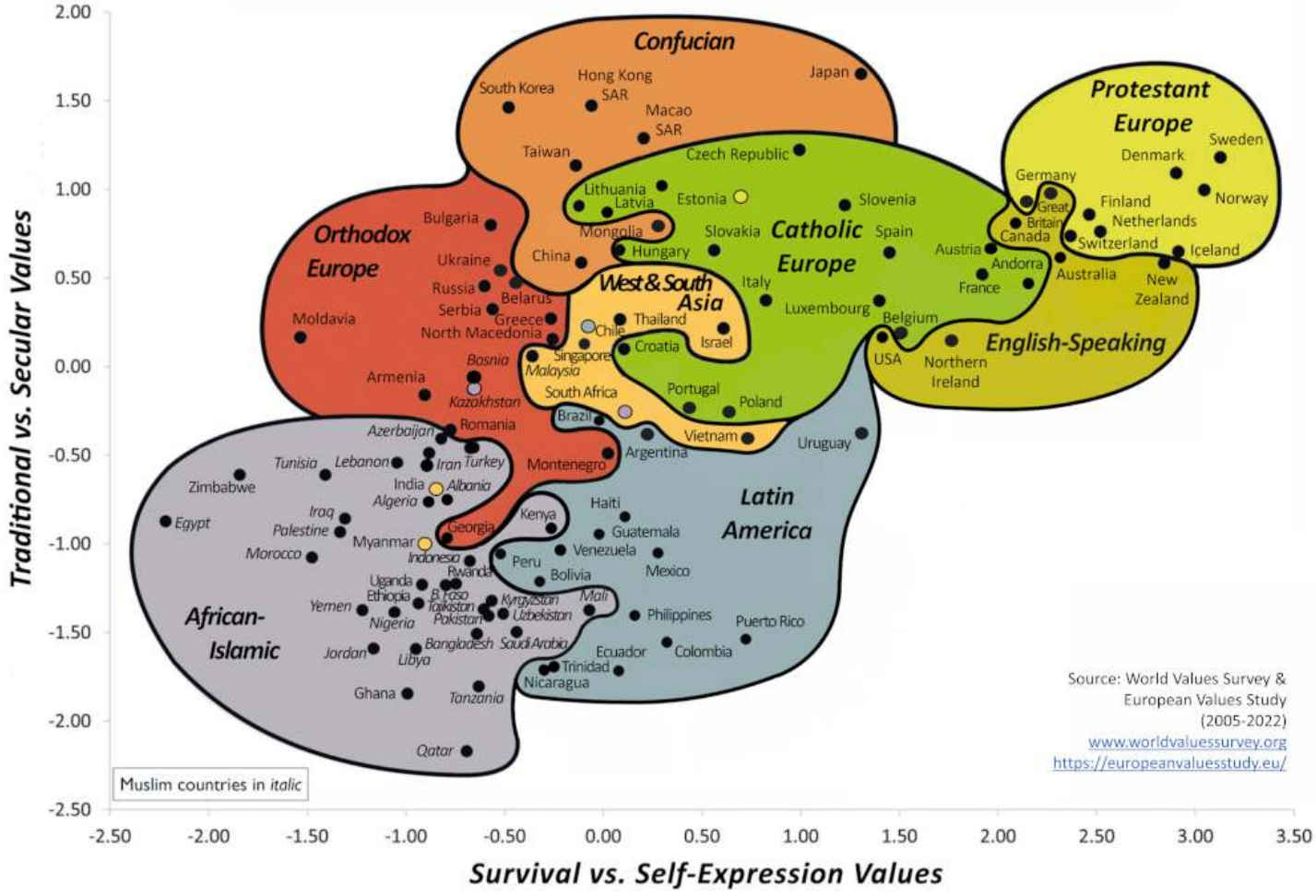
Are electric cars bad for the environment?

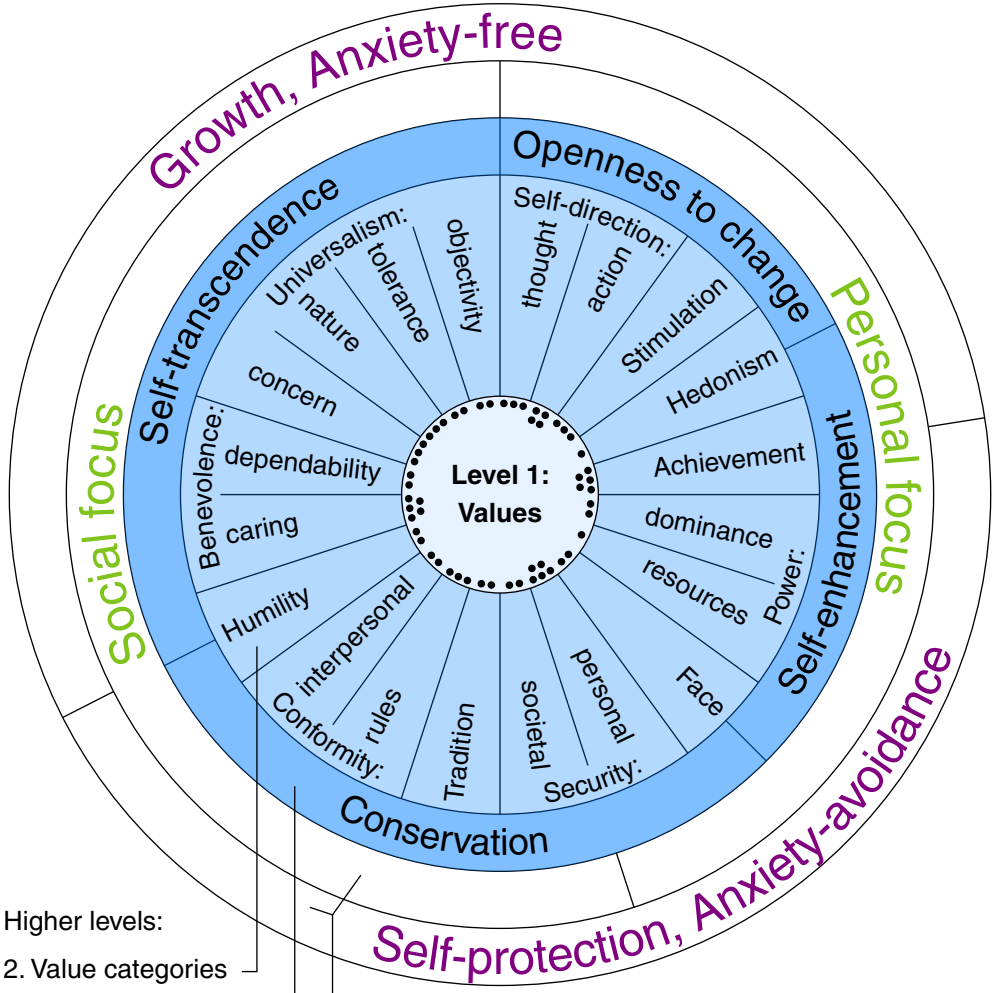


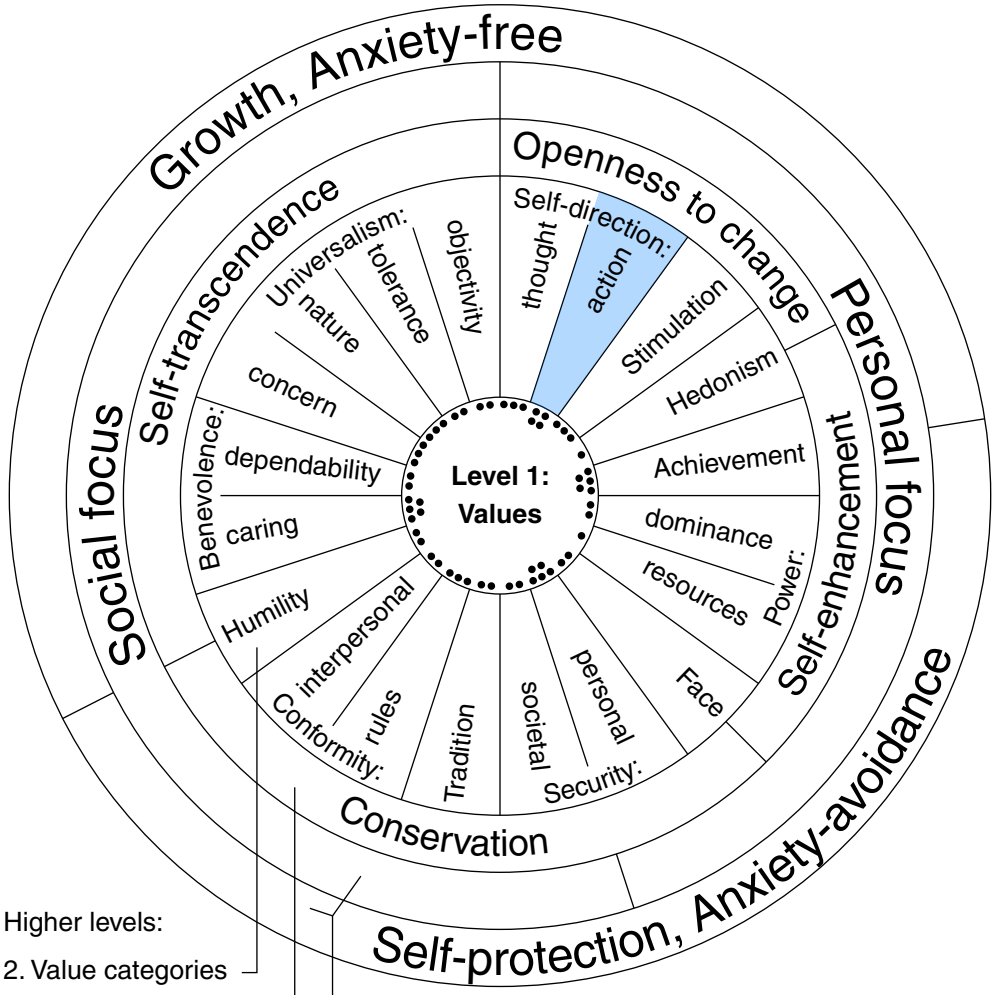


Note: Graphic and data from the Comparative Constitutions Project- Zachary Elkins, Tom Ginsburg, James Melton.

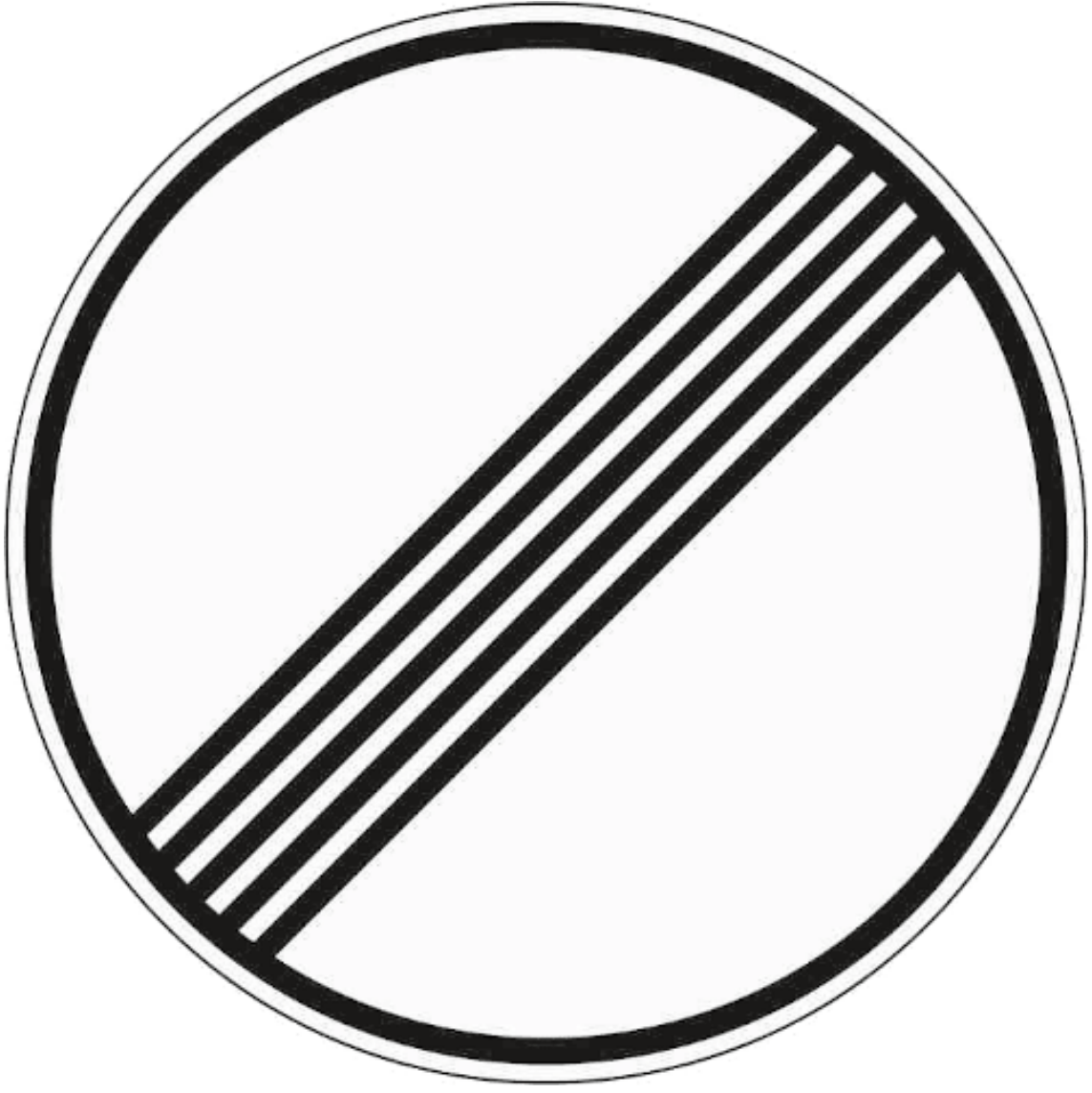
The Inglehart-Welzel World Cultural Map 2023



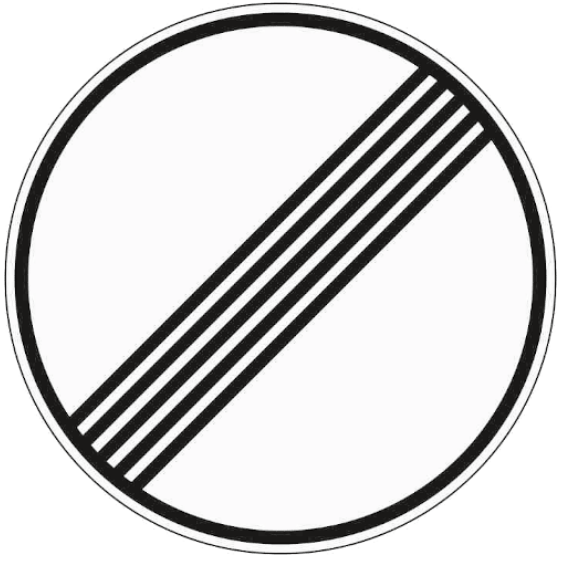

Growth, Anxiety-free
Openness to change
Self-direction:
thought
action
Stimulation
Hedonism
Achievement
Personal focus
Self-enhancement
dominance
Power:
resources
Face
personal
societal
Security:
Conservation
Tradition
Conformity:
rules
interpersonal
Humility
Benevolence:
caring
dependability
concern
Social focus
Self-transcendence
nature
Universalism:
tolerance
objectivity
Level 1:
Values
Self-protection, Anxiety-avoidance
Higher levels:
2. Value categories
3. Higher-order values
4. Base dichotomies

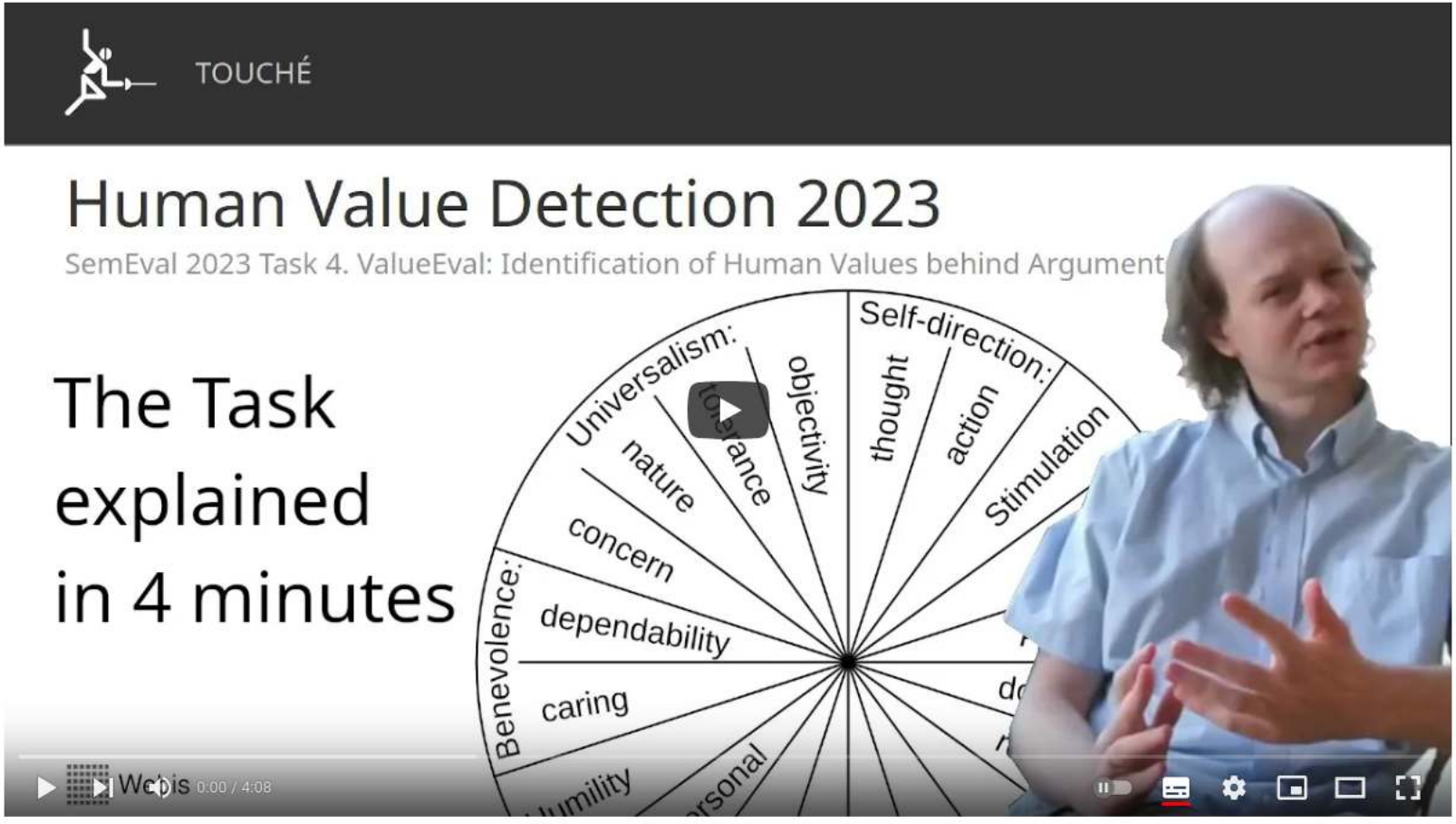


FREEDOM!



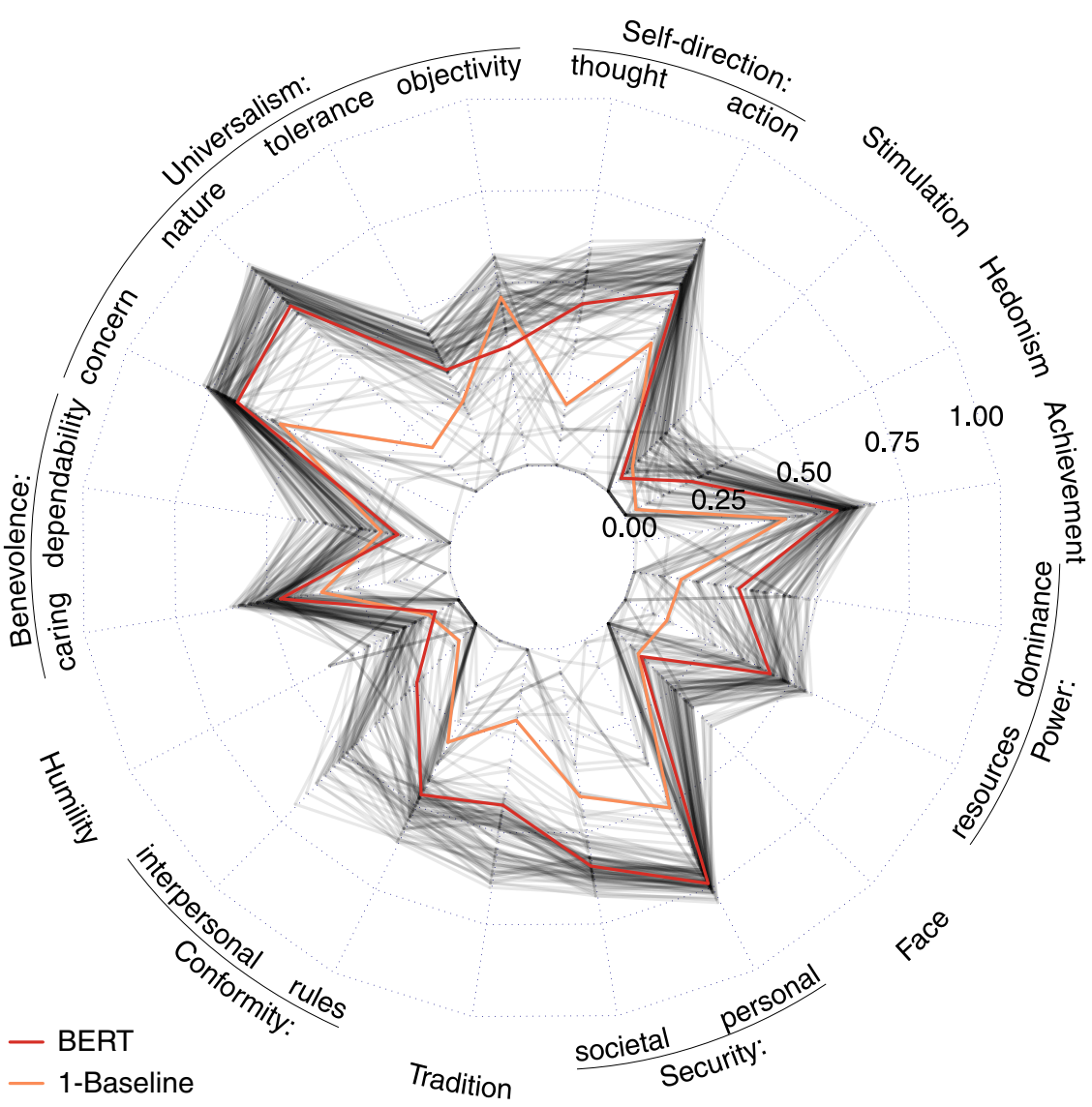

FREEDOM!

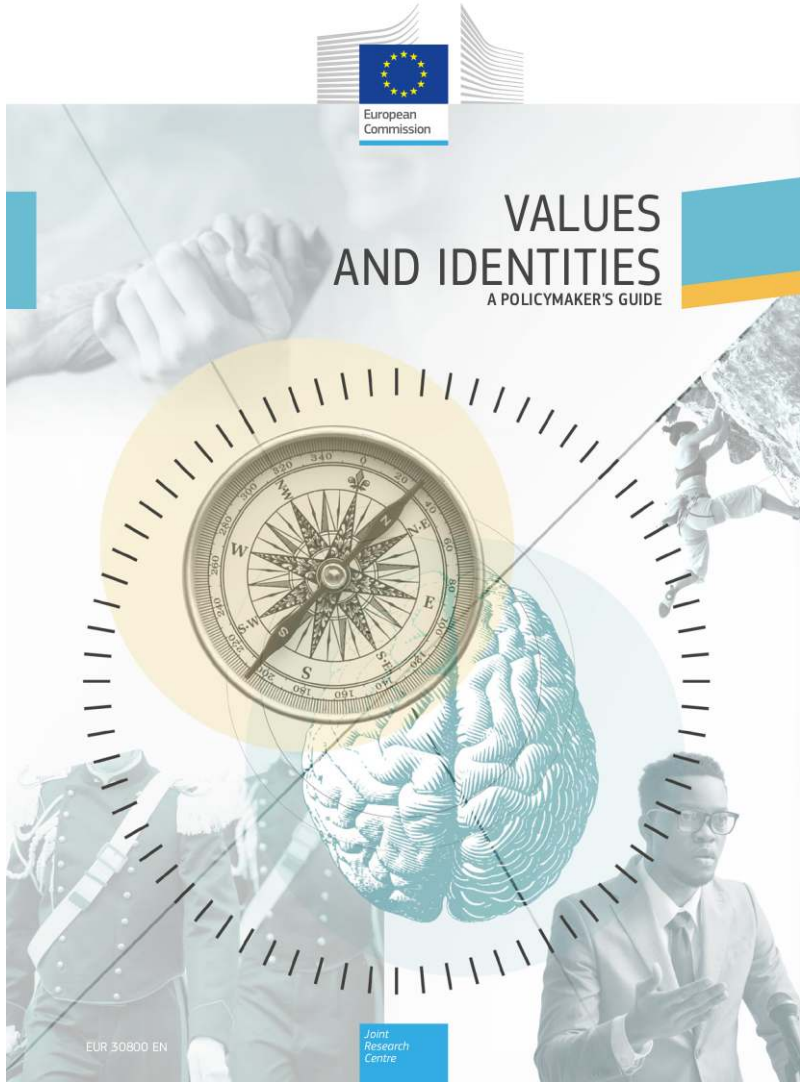




Introduction to the Human Value Detection Task (ValueEval; SemEval 2023 Task 4)

webis
293 Abonnenten
Analysen
Video bearbeiten
2
Teilen
Clip
Speichern




European Commission
VALUES AND IDENTITIES
A POLICYMAKER'S GUIDE
EUR 30800 EN
Joint Research Centre

Johannes Kiesel

Bauhaus-Universität Weimar

Milad Alshomary

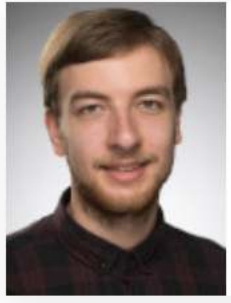
Leibniz University Hannover

Nailia Mirzakhmedova

Bauhaus-Universität Weimar

Maximilian Heinrich

Bauhaus-Universität Weimar



Nicolas Handke

Leipzig University

Henning Wachsmuth

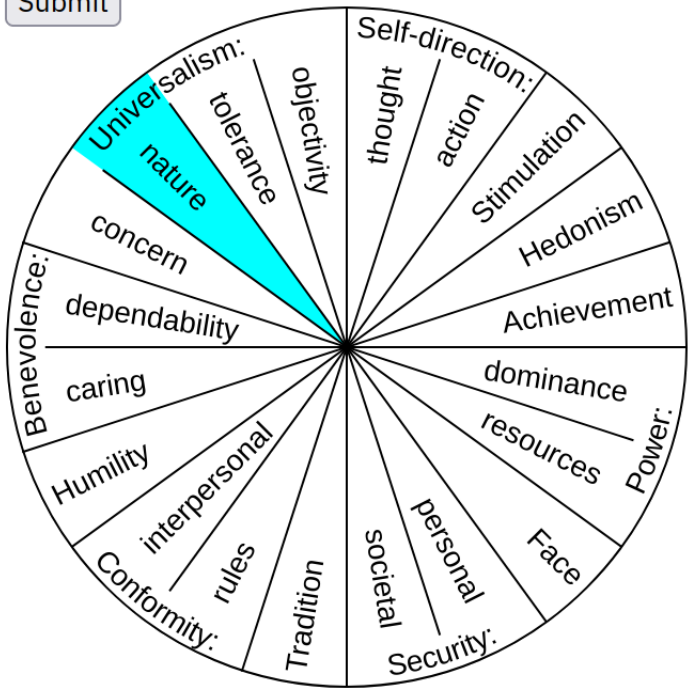
Leibniz University Hannover

Benno Stein

Bauhaus-Universität Weimar

We need to invest in electric cars to fight global warming

Submit



https://values.args.me

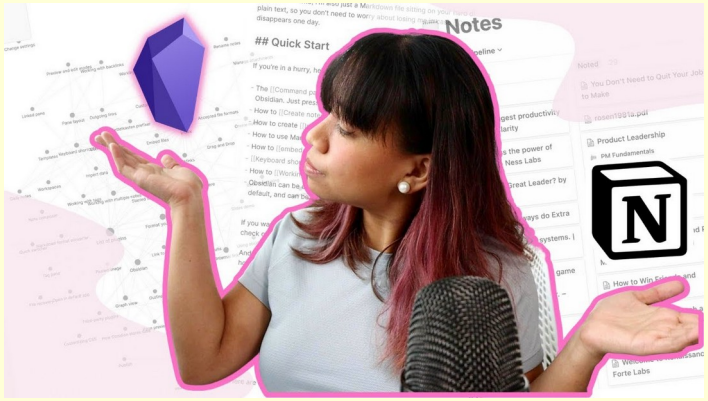
Towards the Personal Experience Base

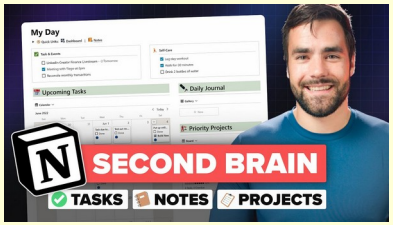
Ludwig Lorenz (ludwig.david.lorenz@uni-weimar.de)

10.03.2023

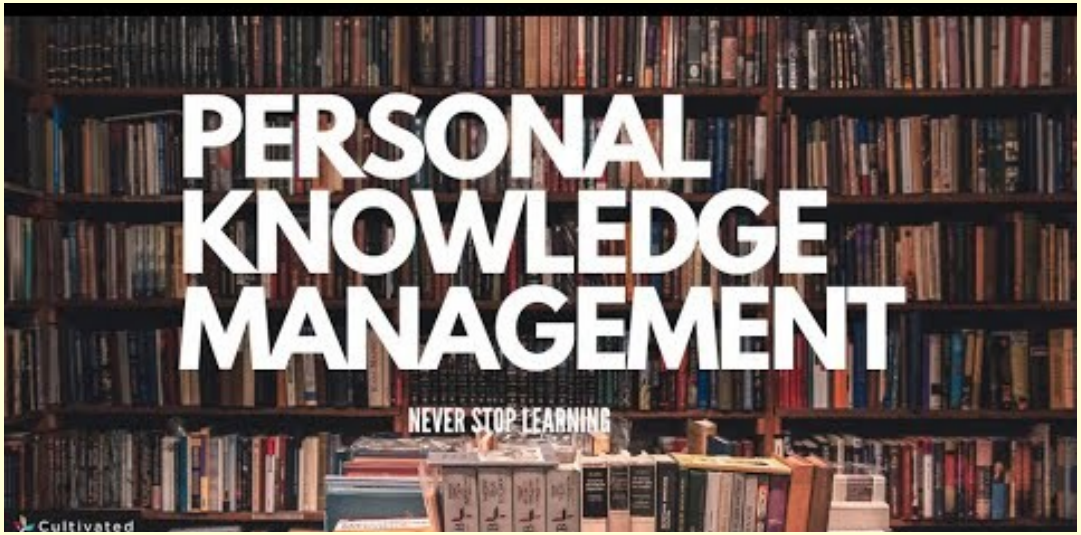
What is *Personal Knowledge Management*?

What is *Personal Knowledge Management*?

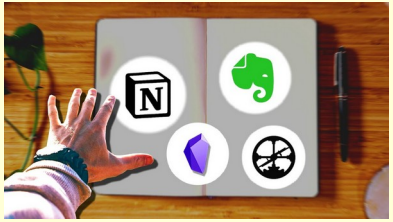

Notes
Quick Start


My Day
SECOND BRAIN
TASKS
NOTES
PROJECTS


Take smart notes!


PERSONAL
KNOWLEDGE
MANAGEMENT
NEVER STOP LEARNING


Zettelkasten Method:
Personal Knowledge Mangement
Simply Explained




Building
a Second
Brain
A Proven Method to Organize
Your Digital Life and Unlock
Your Creative Potential
TIAGO
FORTE


MINIMALIST
+ SETUP
GUIDE


ZETTELKASTEN
STEP BY STEP

Information overload

Calendar

Obsidian

Journal

Contacts

Second Brain

Zettelkasten

High performer

Notion

Tasks

Project Management

* basic idea *

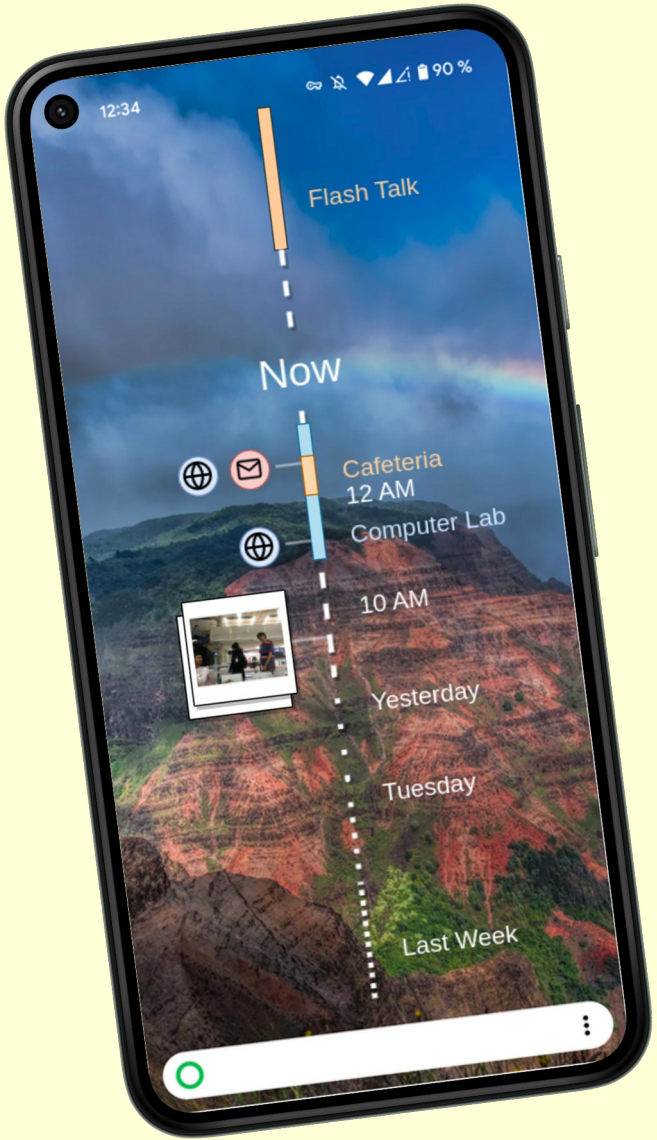
*Use data already contained in our day to day experiences to provide a **mental frame**.*

Then make use of the frame to manage your thougths.

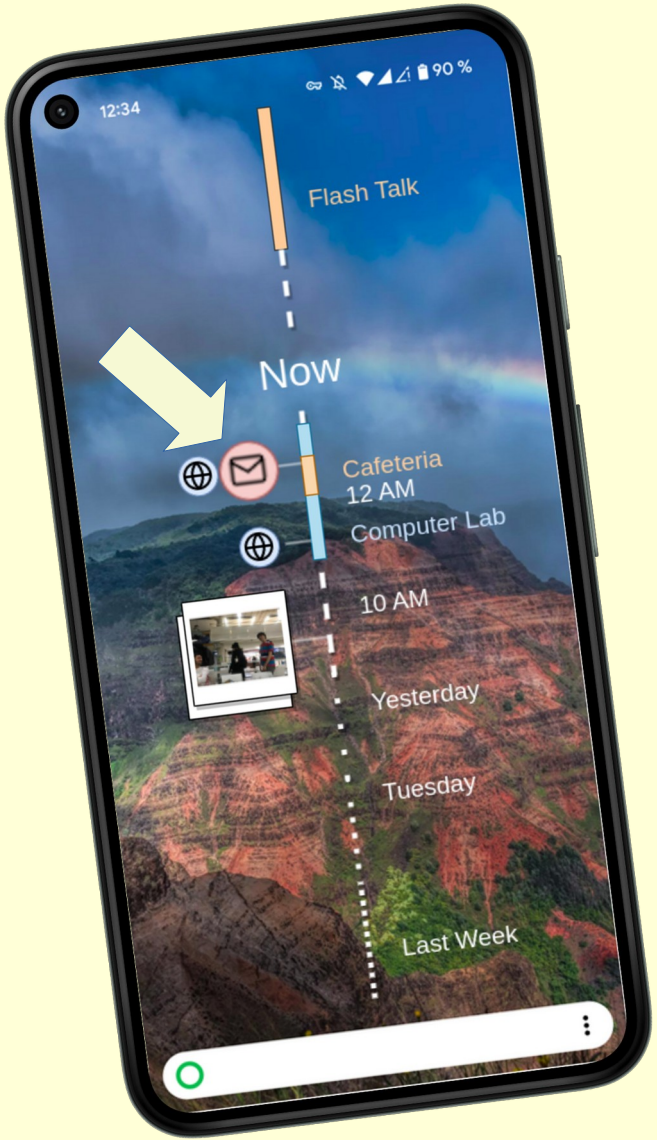

Use my own personal data for research
CHANGE MY MIND

demo

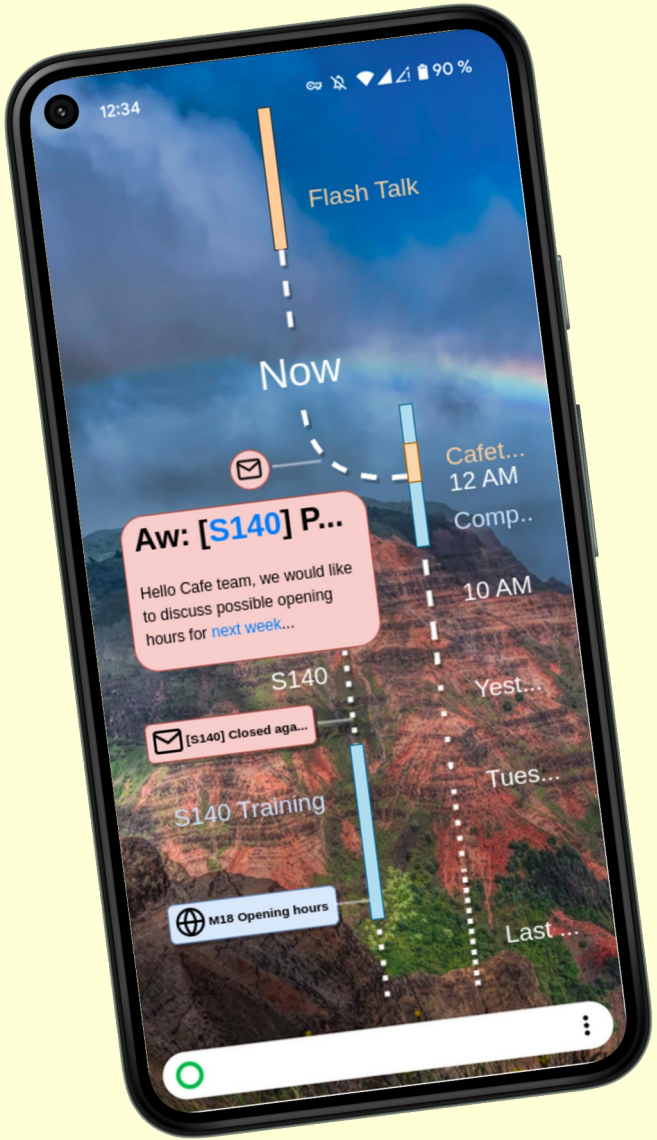
Navigate and search timeline



Select records



Traverse keyword specific timeline



MANIPULATING EMBEDDINGS OF STABLE DIFFUSION PROMPTS

-

MASTER THESIS

LEIPZIG, 10.03.23

JULIA PETERS

SUPERVISED BY NIKLAS DECKERS

TASK: GENERATE "PERFECT" IMAGE OF A FOX

Prompt:
"Fox, surrounded by red flowers, green landscape in the background, photorealistic"

Stable Diffusion

TASK: GENERATE "PERFECT" IMAGE OF A FOX II

➔ Problem: does not entirely satisfy the expectations

➔ Approaches:

- Trying different seeds
- Prompt adjustments

➔ Manual trial and error

PROBLEM WITH IMAGE REGENERATION

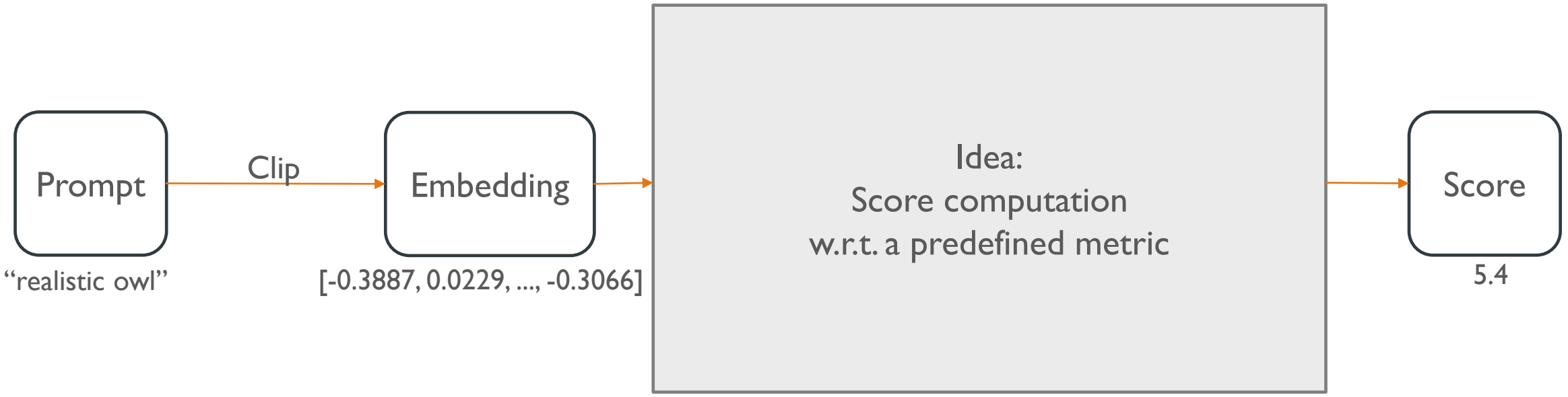
- Not a “pretty” image
- Low contrast in the background (landscape?)

➔ Endless image generation

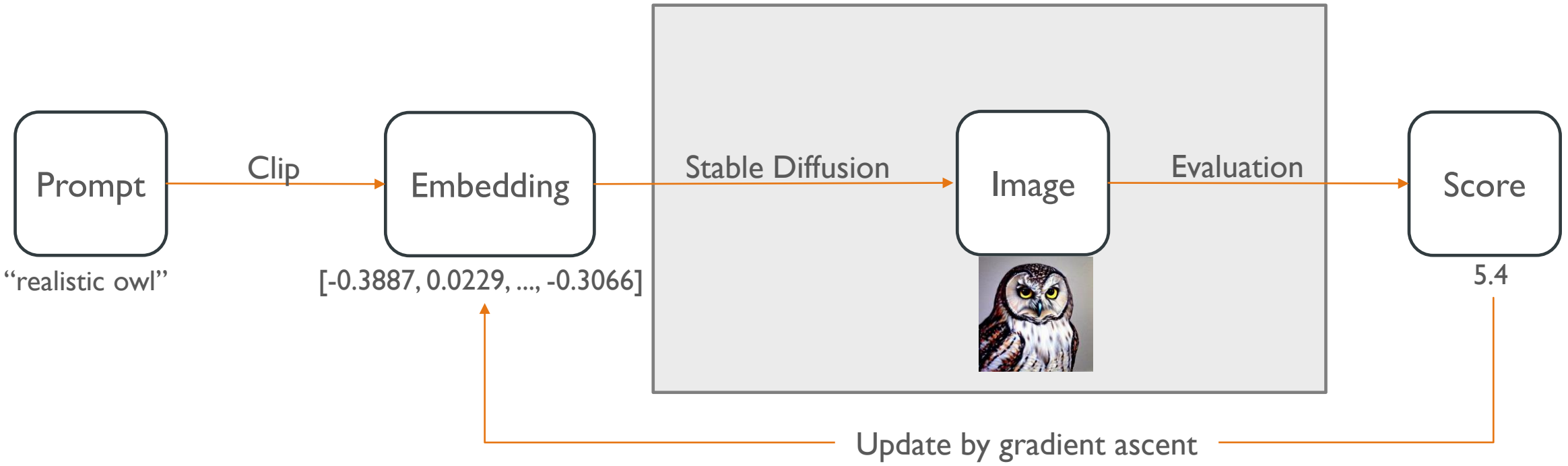
- Not realistic enough
- Too much light

- Small adjustments in prompt or usage of different seeds can lead to completely different images

PROPOSED PIPELINE FOR STEERING IMAGE GENERATION



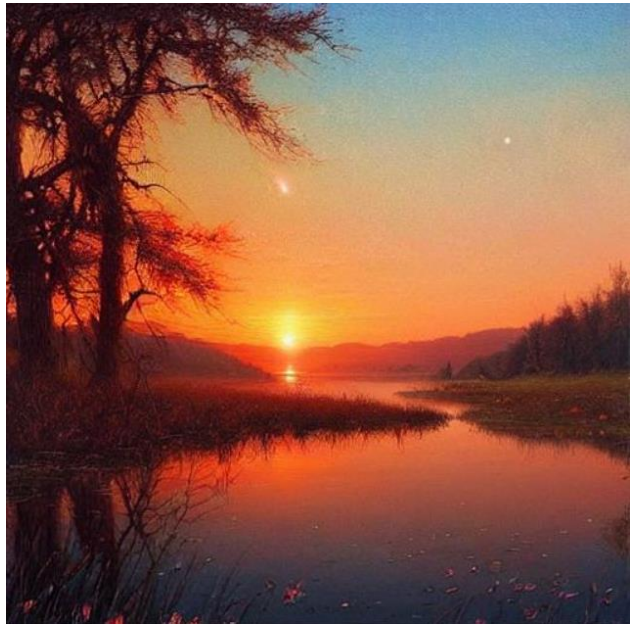
PROPOSED PIPELINE FOR STEERING IMAGE GENERATION



METRIC CHOICE FOR SCORE COMPUTATION

1. Simple Metric Ideas
   - Measurement of grayscale of the image
   - Blurriness of the Image

Blurriness decreased



Original Embedding

Blurriness increased

METRIC CHOICE FOR SCORE COMPUTATION

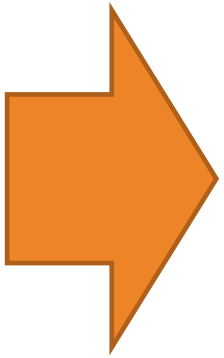
2. LAION Aesthetic Predictor V2
   - MLP trained on 2.37B image - rating pairs ranging from 1 – 10

Score: 2.3

Score: 5.8

LAION AESTHETIC PREDICTOR RESULTS

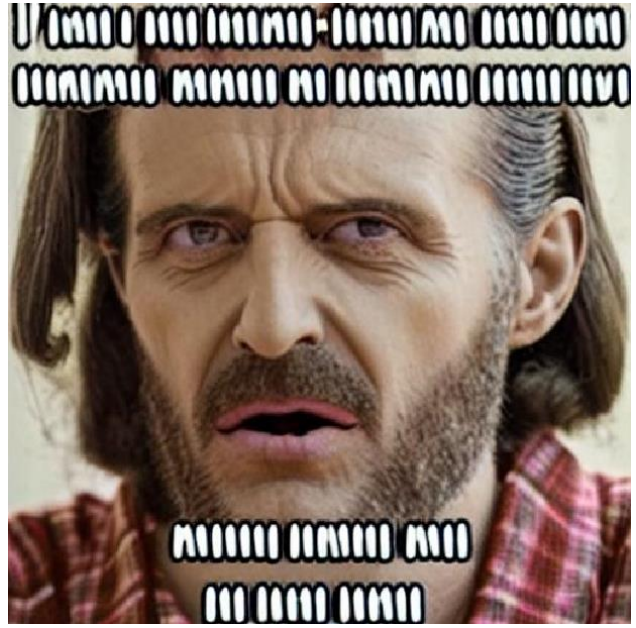
Score: 5.8



Score: 7.0

LAION AESTHETIC PREDICTOR RESULTS II

Score: 2.3



Score: 4.1

OUTLOOK

➔ Future work:

- Replacement of predefined metrics by user ratings (Active Learning)
- Reducing lack of control caused by random seed

OUTLOOK

➔ Future work:

- Replacement of predefined metrics by user ratings (Active Learning)
- Reducing lack of control caused by random seed

The Archive Query Log

Heinrich Sebastian Maik Lukas Harry Benno Matthias Martin

Query Logs

What are they good for?

- Query understanding
- Learning to rank
- SERP analysis

What's the issue?

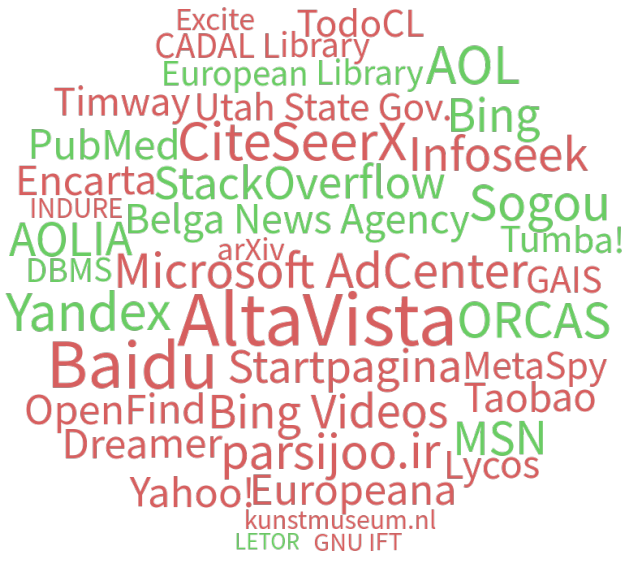
- Most are **private**
- Public ones **much smaller**

Previous Logs

Query Logs

What are they good for?

- Query understanding
- Learning to rank
- SERP analysis

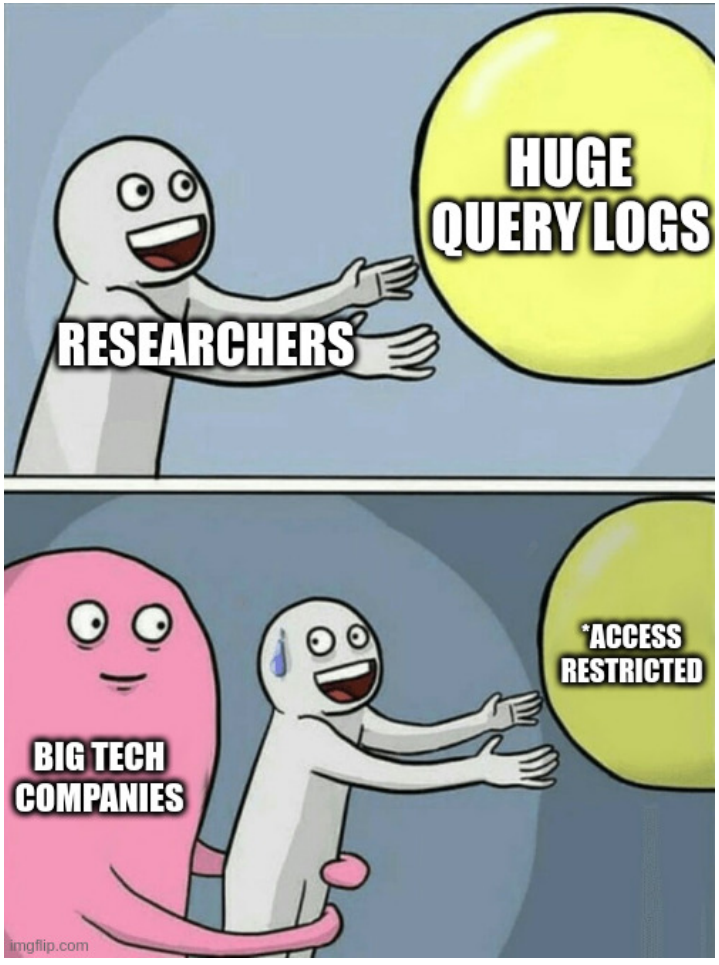


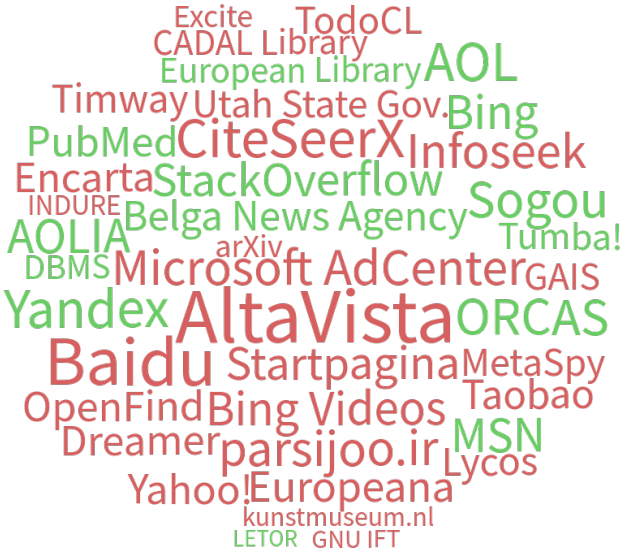
What's the issue?

- Most are **private**
- Public ones **much smaller**

Previous Logs

Query Logs





Previous Logs

Web Archive to the rescue!

But how?

1. List SERP URLs from Web Archive
2. Parse query
3. Download SERP HTML
4. Parse search results

Like this…

`https://google.com/search?q=covid+19+usa+map`

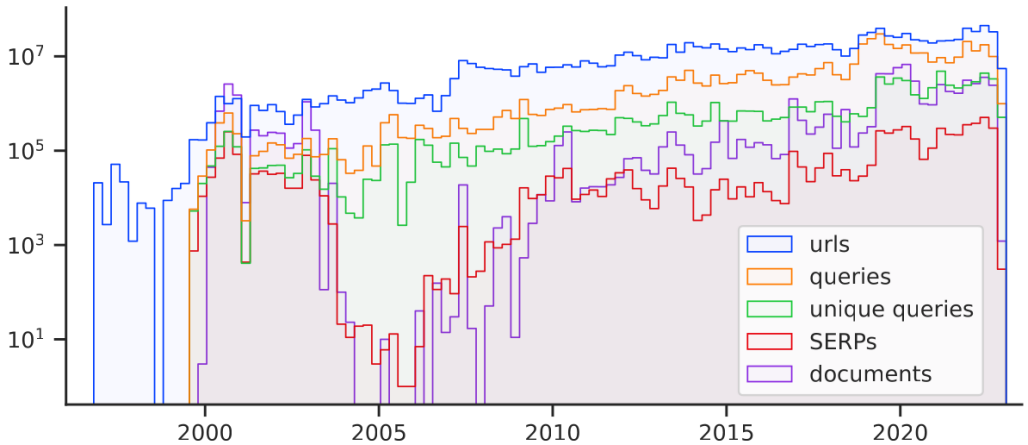
URL prefix — query

With the AQL

Meet the **Archive Query Log**

webis-de/archive-query-log

- 356 million queries
- 166 million SERPs
- 1.7 billion search results
- across 25 years
- 550 search providers



And now?

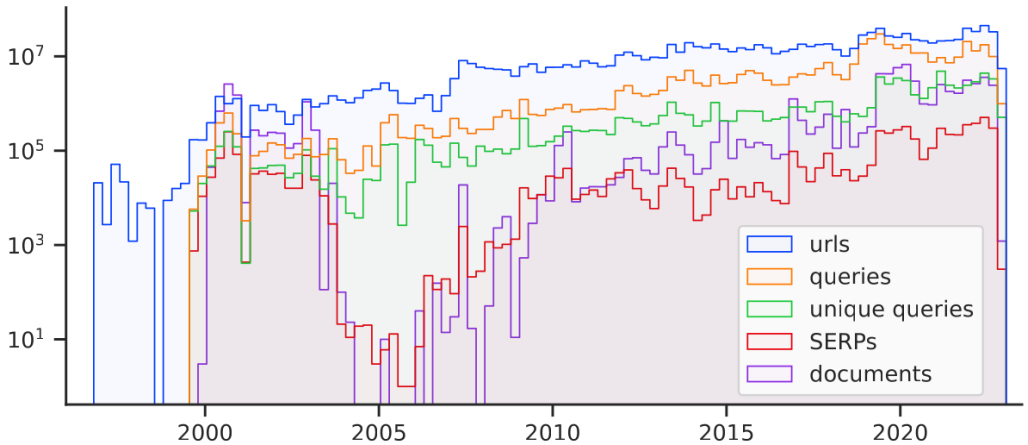
- high diversity: topic and language
- diachronic analyses
- publically accessible via TIRA

Thank you, stay tuned for more!

Meet the **Archive Query Log**

webis-de/archive-query-log

- 356 million queries
- 166 million SERPs
- 1.7 billion search results
- across 25 years
- 550 search providers



And now?

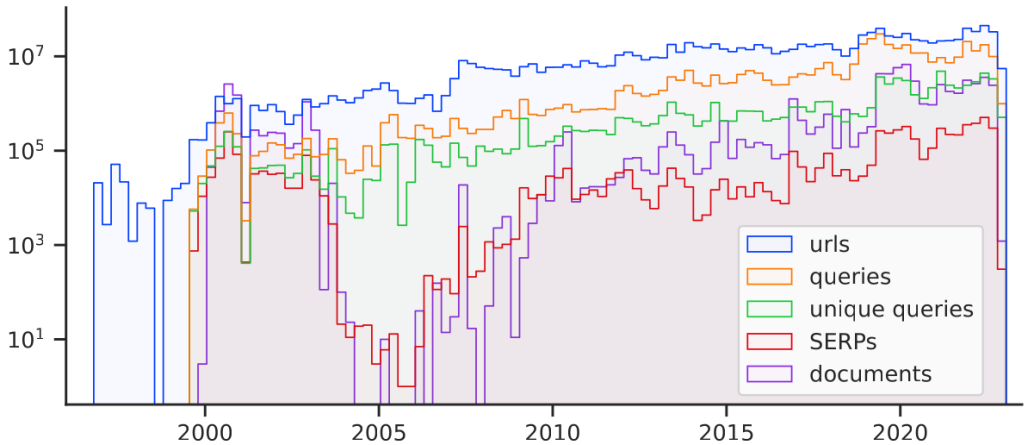
- high diversity: topic and language
- diachronic analyses
- publically accessible via TIRA

Thank you, stay tuned for more!

Meet the **Archive Query Log**

webis-de/archive-query-log

- 356 million queries
- 166 million SERPs
- 1.7 billion search results
- across 25 years
- 550 search providers



And now?

- high diversity: topic and language
- diachronic analyses
- publically accessible via TIRA

Thank you, stay tuned for more!

Quantifying the Effects of Real-World Conflicts on Wikipedia

Flash Talks 2023

Bauhaus University Weimar
Ali Saqallah
Master's Thesis

Vandalism in Wikipedia

What is Vandalism in Wikipedia?

Batman

From Wikipedia, the free encyclopedia

This is an old revision of this page, as edited by Guitar Godd23 (talk | contribs) at 20:55, 14 June 2007. It may differ significantly from the current revision.

(diff) ← Previous revision | Current revision (diff) | Newer revision → (diff)

DUN NUH NUH NUH NUH NUH DUH NUH NUH NUH NUH NUH
BATMAN! DUN NUH NUH NUH NUH NUH DUH NUH NUH NUH
NUH NUH BATMAN! DUN NUH NUH NUH NUH NUH DUH NUH
NUH NUH NUH NUH BATMAN! DUN NUH NUH NUH NUH NUH
DUH NUH NUH NUH NUH NUH BATMAN! BATMAN! BATMAN!
BATMAN! DUN NUH NUH NUH NUH NUH DUH NUH NUH NUH
NUH NUH **BATMAN**!!

Effects of Real-World Conflicts on Wikipedia

Before

Bucha (Ukrainian: Буча) is a city in Ukraine's Kyiv Oblast. Administratively, it is incorporated as a city of oblast significance. Its population is approximately 36,971 (2021 est.)[1] Bucha Day is celebrated in the city on 13 September.

History

The settlement arose with construction of the Kyiv–Kovel railway in 1898 around a small train stop by Bucha River [uk]. Bucha was a train stop of the Kyiv–Kovel railway similar to one in the modern city of Irpin. In close vicinity to the Bucha train stop there was a small village called Yablunka, where there used to be a brick factory. Yablunka is mentioned in the 19th century Polish Geographic dictionary as the village of Jabłonka 37 versts away from Kyiv.[2]

During World War II, before the liberation of Kyiv from Nazi forces in December 1943, Bucha was the location of the headquarters of the 1st Ukrainian Front commanded by General Vatutin.

Bucha was granted city status on February 9, 2006.[3] Before 1996, Bucha was a town within the Irpin city municipality.

Battle of Bucha

Main article: Battle of Bucha

During the 2022 Russian invasion of Ukraine, heavy fighting took place in Bucha as part of the Kyiv offensive, resulting in severe Russian losses. Russian forces attacked the town's Afghanistan War memorial, which they may have mistaken for a Ukrainian military vehicle.[4] The city was occupied by Russian forces on 12 March. Mayor Anatoliy Fedoruk announced the recapture of Bucha by Ukrainian forces on 31 March 2022. [5]

Bucha Massacre

Main article: Bucha massacre

On 2 April 2022, news reports and video emerged showing streets in Bucha covered with the bodies of men dressed in civilian clothes. Some of those found had their hands tied. According to one estimate, at least 20 bodies were found.[6] Later on, it was reported that almost 300 had been found buried in mass graves. The bodies included those of men, women, and children.[citation needed]

Places and people

There is a stadium in Bucha named Yuvileiny Stadium, where some matches were held in October 2016 for the 2017 UEFA European Under-19 Championship qualification.[7]

There is a glass factory in Bucha. Built in 1946, it was closed in 2016. There is a small train stop called

Bucha
Буча

City

Bucha city park

Flag

Coat of arms

Show map of Kyiv Oblast
Show map of Ukraine
Show all

Coordinates: 50°33'27.15"N 30°13'36.17"E

After

Bucha (Ukrainian: Буча) is a city in Ukraine's Kyiv Oblast. Administratively, it is incorporated as a city of oblast significance. Its population is approximately 36,971 (2021 est.)[1] Bucha Day is celebrated in the city on 13 September.

History

The settlement arose with construction of the Kyiv–Kovel railway in 1898 around a small train stop by Bucha River [uk]. Bucha was a train stop of the Kyiv–Kovel railway similar to one in the modern city of Irpin. In close vicinity to the Bucha train stop there was a small village called Yablunka, where there used to be a brick factory. Yablunka is mentioned in the 19th century Polish Geographic dictionary as the village of Jabłonka 37 versts away from Kyiv.[2]

During World War II, before the liberation of Kyiv from Nazi forces in December 1943, Bucha was the location of the headquarters of the 1st Ukrainian Front commanded by General Vatutin.

Bucha was granted city status on February 9, 2006.[3] Before 1996, Bucha was a town within the Irpin city municipality.

Battle of Bucha

Main article: Battle of Bucha

During the 2022 Russian invasion of Ukraine, heavy fighting took place in Bucha as part of the Kyiv offensive, resulting in severe Russian losses. Russian forces attacked the town's Afghanistan War memorial, which they may have mistaken for a Ukrainian military vehicle.[4] The city was occupied by Russian forces on 12 March. Mayor Anatoliy Fedoruk announced the recapture of Bucha by Ukrainian forces on 31 March 2022. [5]

Places and people

There is a stadium in Bucha named Yuvileiny Stadium, where some matches were held in October 2016 for the 2017 UEFA European Under-19 Championship qualification.[6]

There is a glass factory in Bucha. Built in 1946, it was closed in 2016. There is a small train stop called "Sklozavodska".

The town's main landmark is a 19th-century railway station located at the south edge of the city. Through the city runs a major highway M 07

Bucha
Буча

City

Bucha city park

Flag

Coat of arms

Steps and Methods

Mining Vandalism

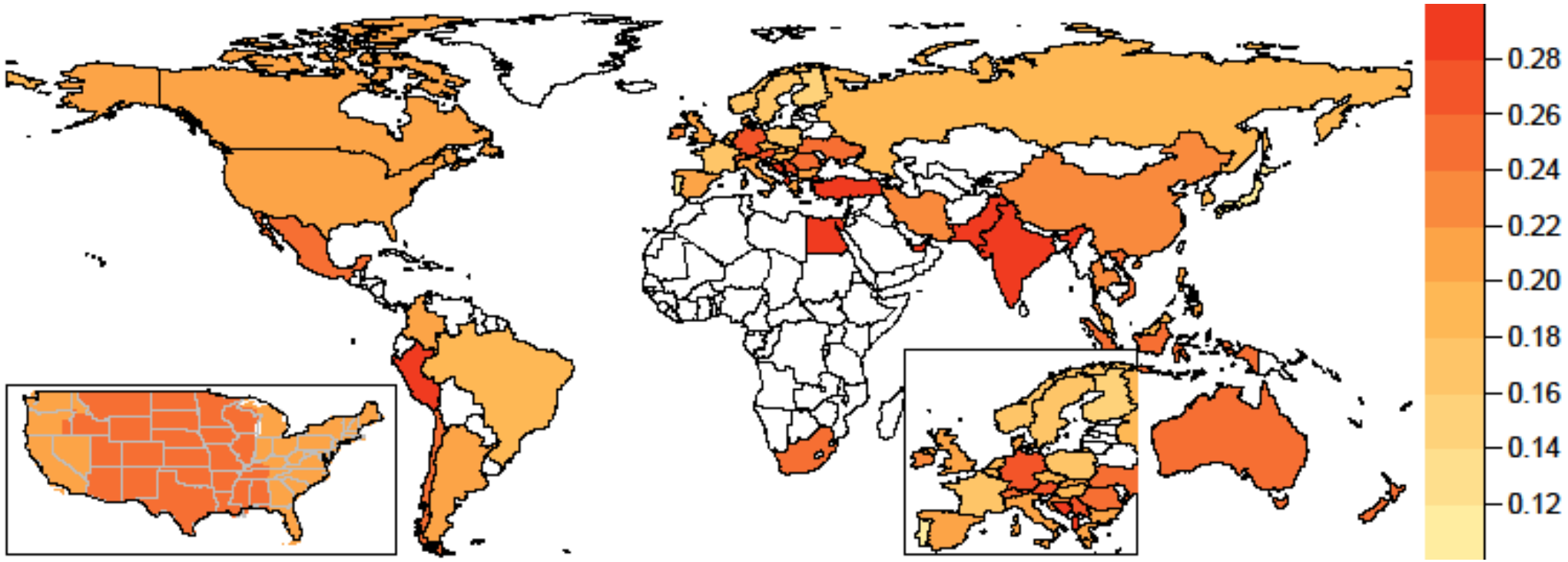
Geolocating Anonymous Wikipedia Editors with GeoDBs

Analyzing Vandalism Patterns through Spatio-Temporal Analysis

Choosing Regions, Times, and Events for Detailed Vandalism Analysis

Analyzing Vandalism in Wikipedia

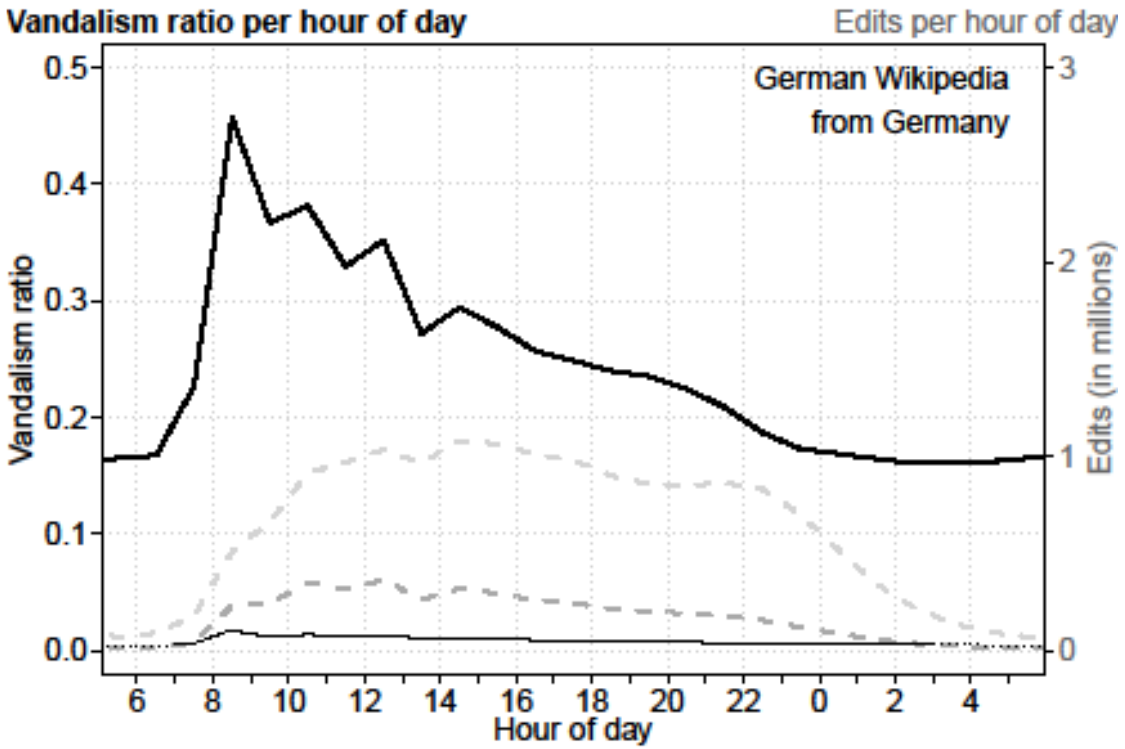
By country

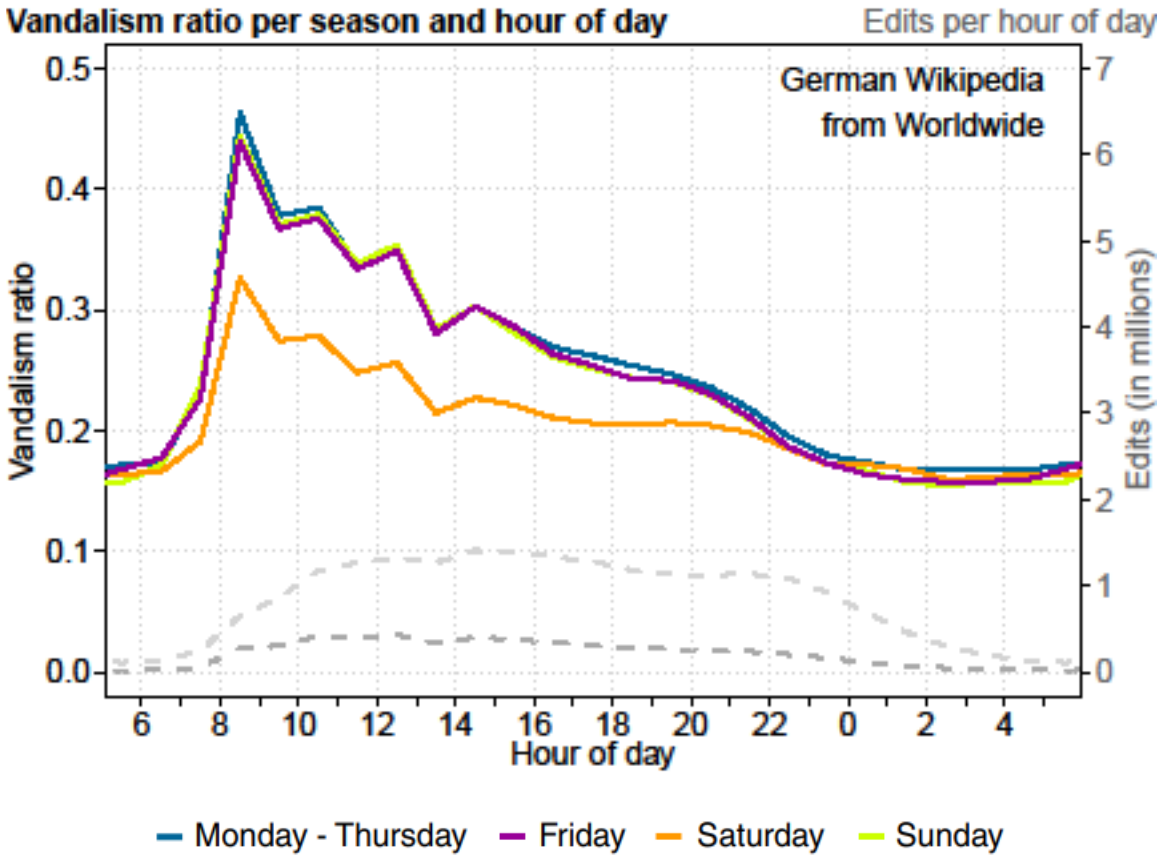


Ratio of vandalism to all edits in the German Wikipedia by country.

Analyzing Vandalism in Wikipedia

By time





Ratio of vandalism to all edits in the German Wikipedia by per hour of day.

Resources

- Kiesel, J., Potthast, M., Hagen, M., & Stein, B. (2017). Spatio-Temporal Analysis of Reverted Wikipedia Edits. Proceedings of the International AAAI Conference on Web and Social Media, 11(1), 122-131. https://doi.org/10.1609/icwsm.v11i1.14900

- Code from Spatio-Temporal Analysis of Reverted Wikipedia Edits Paper https://github.com/webis-de/ICWSM-17

Thank You

Efficient Cross-Encoder Re-ranking

A token's perspective

Cross-Encoder Re-reranking

Query: What is the best programming language?

Cross-Encoder Re-reranking

Query: What is the best programming language?

Document:
As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages.

Cross-Encoder Re-reranking

Query: What is the best programming language?

Document:
As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages.

Cross-Encoder Re-reranking

Query: What is the best programming language?

Document:
As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages.

[CLS] Query [SEP] Document [SEP] ➜

Cross-Encoder Re-reranking

Query: What is the best programming language?

Document:
As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages.

[CLS] Query [SEP] Document [SEP] ➜

Token's Perspective

Input:

[CLS] What is the best programming language? [SEP] As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages. [SEP]

Token's Perspective

Input:

[CLS] What is the best programming language? [SEP] As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages. [SEP]

Token's Perspective

Input:

[CLS] What is the best programming language? [SEP] As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages. [SEP]

Token's Perspective

Input:

[CLS] What is the best programming language? [SEP] As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages. [SEP]

We want to report to our boss how relevant we are to the query

Token's Perspective

Input:

[CLS] What is the best programming language? [SEP] As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages. [SEP]

We want to report to our boss how relevant we are to the query

What kind of python are we?

Token's Perspective

Input:

[CLS] What is the best programming language? [SEP] As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages. [SEP]

We want to report to our boss how relevant we are to the query

What kind of python are we? ➜ We need to talk to other tokens

Token's Perspective

Input:

[CLS] What is the best programming language? [SEP] As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages. [SEP]

We want to report to our boss how relevant we are to the query

What kind of python are we? ➜ We need to talk to other tokens

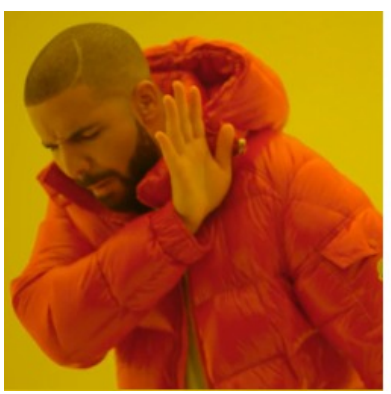
Token's Perspective

Input:

[CLS] What is the best programming language? [SEP] As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages. [SEP]

We want to report to our boss how relevant we are to the query

What kind of python are we? ➜ We need to talk to other tokens



Attend to all tokens

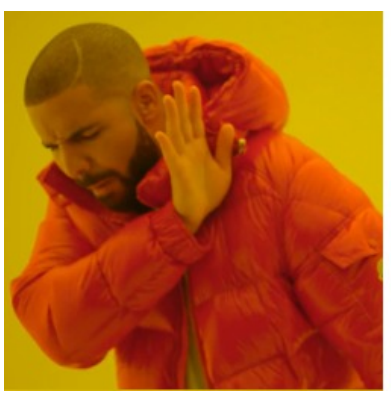
Token's Perspective

Input:

[CLS] What is the best programming language? [SEP] As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages. [SEP]

We want to report to our boss how relevant we are to the query

What kind of python are we? ➜ We need to talk to other tokens



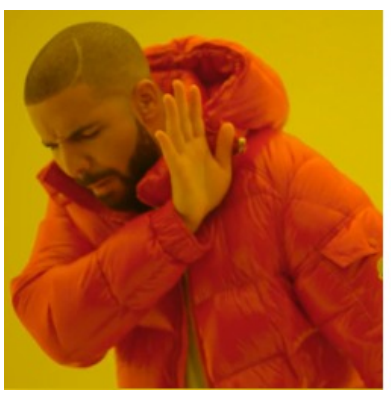
Token's Perspective

Input:

[CLS] What is the best programming language? [SEP] As far as web and software development goes, Python, followed by Java, JavaScript, and C++, are among the most popular programming languages. [SEP]

We want to report to our boss how relevant we are to the query

What kind of python are we? ➜ We need to talk to other tokens

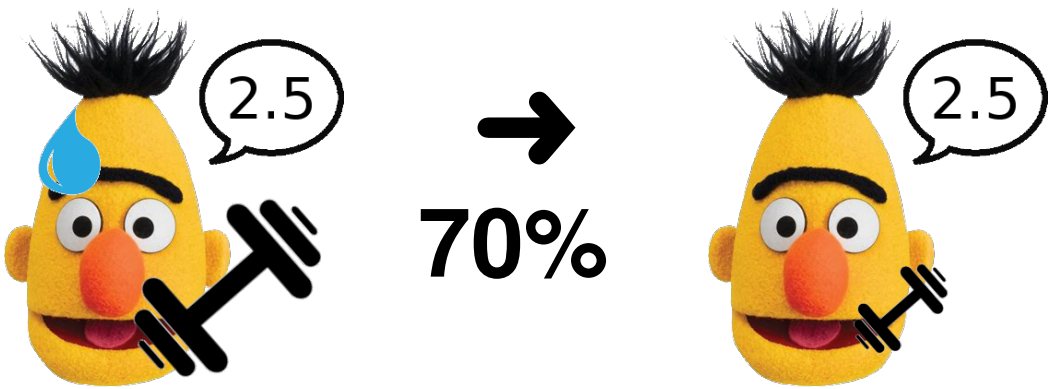


Attend to all tokens

Attend to tokens close to you

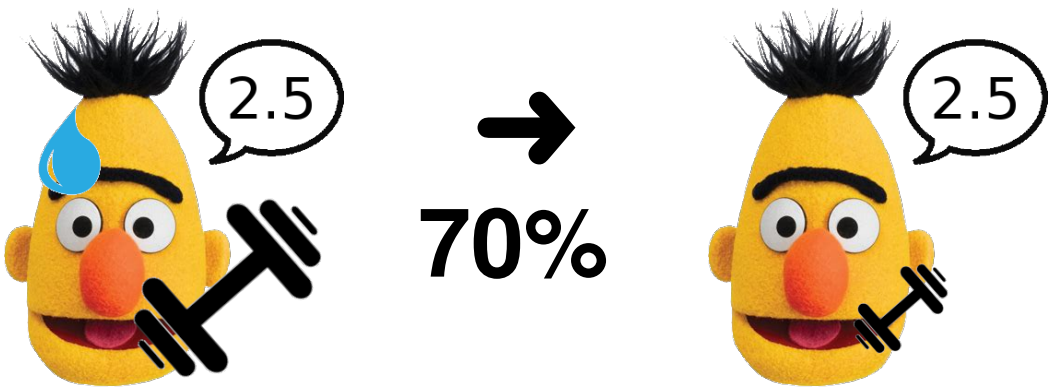
Results

- 70% lower memory consumption without significant drop in performance
- Short paper submitted to SIGIR'23



Results

- 70% lower memory consumption without significant drop in performance
- Short paper submitted to SIGIR'23



Thanks!

Structuring Discussions

Dominik Schwabe

Motivation

- Structure a discussion into consistent groups
- Find a comprehensive label for each group

Grouping

- Embed sentences with SBERT
- Cluster with HDBSCAN

Grouping

- Embed sentences with SBERT
- Cluster with HDBSCAN

Demo

Labeling

- T0 (11B)
- GPT NeoX (20B)
- OPT (66B)
- GPT-3.5 (175B)
- BLOOM (176B)

```
Generate a single descriptive phrase that
describes the following debate in very
simple language, without talking about
the debate or the author.
Debate: """{text}"""

Protecting non-smokers from second-hand
smoke.
```

AI Techniques for More Effective Systematic Review Literature Search

Shuai Wang

shuai.wang2@uq.edu.au
The University of Queensland, Australia

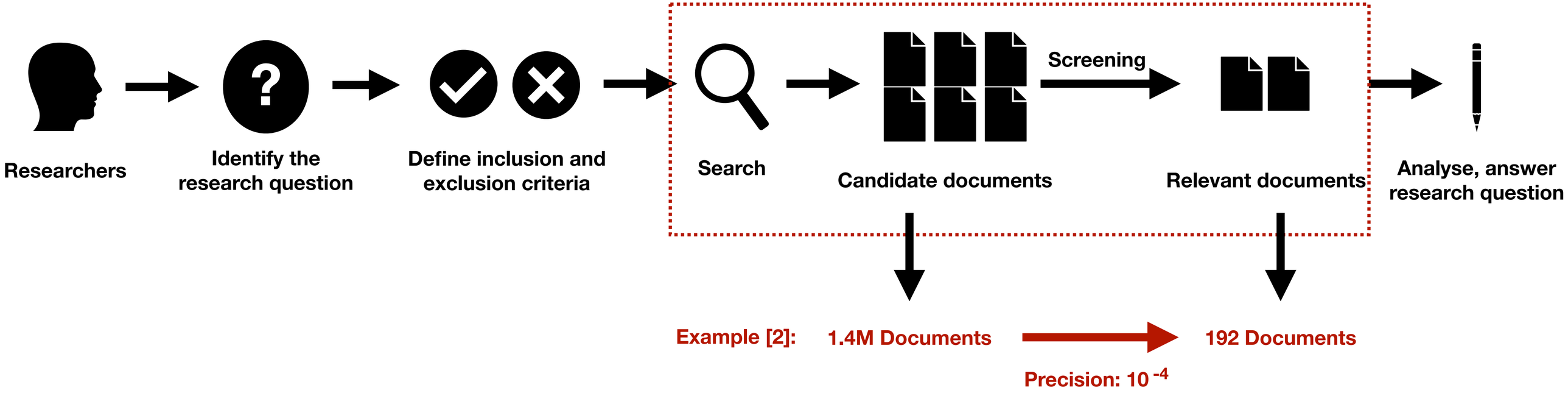
Systematic Reviews

A medical systematic review is a comprehensive review of literature for a highly focused research question.

- Time: >1,100 hours / >2 years
- Spend: Around $350k [1]

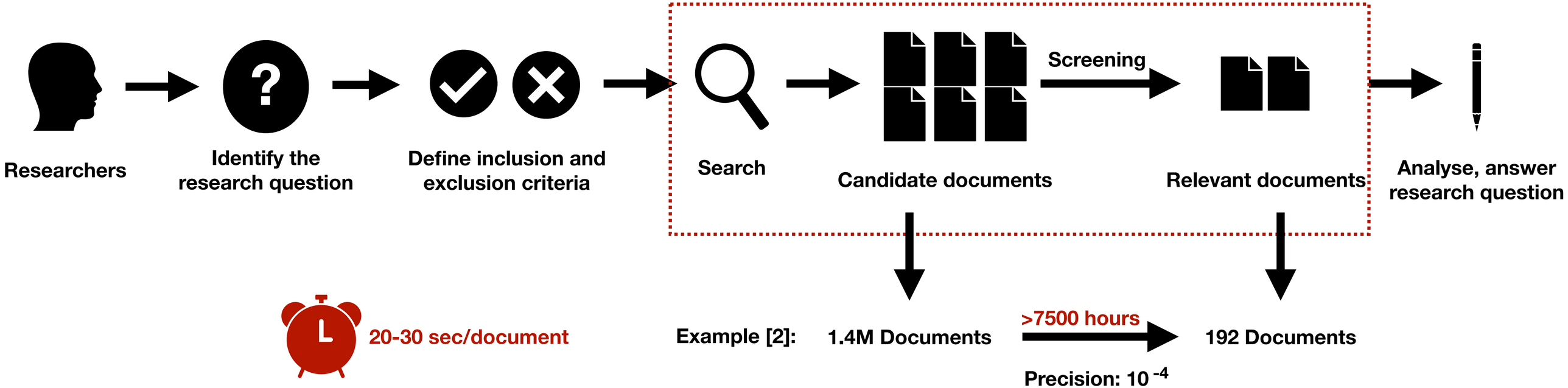
[1] McGowan, et al. Systematic reviews need systematic searchers, JMIR, 2005

The systematic review process



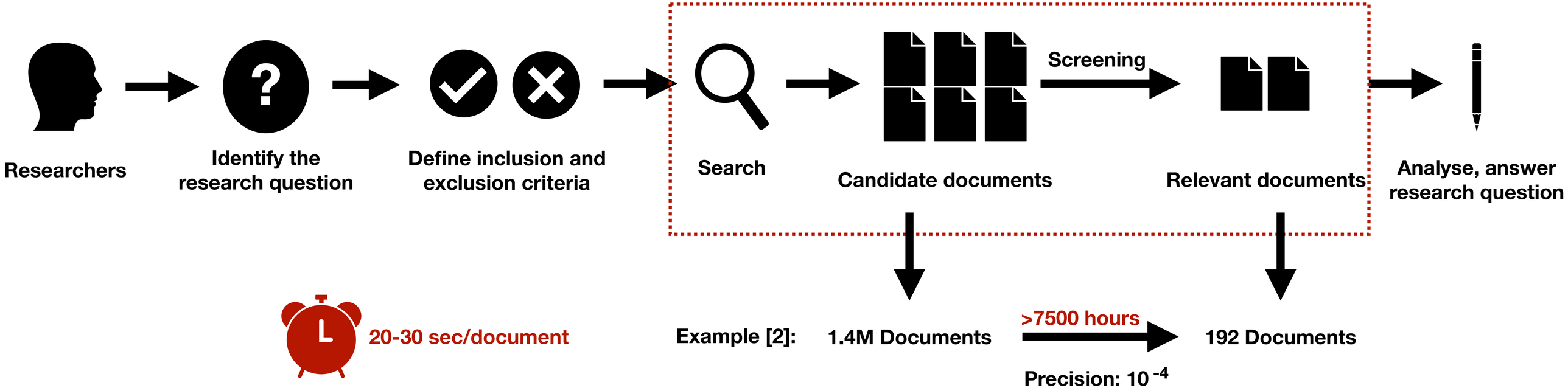
[2] Shemilt, et al. Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews, SR, 2016

The systematic review process



[2] Shemilt, et al. Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews, SR, 2016

The systematic review process



Can we build automation tools to minimise the costs of screening?

[2] Shemilt, et al. Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews, SR, 2016

Current progress

Direction 1. Construction of high-quality systematic review Boolean Queries.

MeSH Term suggestion [14,15,16]

Direction 2. Systematic review screening prioritisation.

Screening prioritisation input & method [17,18,19]

[14] Wang, et al. “Mesh term suggestion for systematic review literature search,” ADCS, 2021.
[15] Wang, et al. “Automated mesh term suggestion for effective query formulation in systematic reviews literature search.”, ISWA, 2022.
[16] Wang, et al. “MeSH Suggester: A Library and System for MeSH Term Suggestion for Systematic Review Boolean Query Construction,” WSDM, 2023.
[17] Wang, et al. “From little things big things grow: A collection with seed studies for medical systematic review literature search,” SIGIR, 2022.
[18] Wang, et al. “Seed-driven document ranking for systematic reviews: A reproducibility study,” ECIR, 2022.
[19] Wang, et al. “Neural Rankers for Effective Screening Prioritization in Medical Systematic Review Literature Search,” ADCS, 2022.

Current progress

Direction 1. Construction of high-quality systematic review Boolean Queries.

MeSH Term suggestion [14,15,16]

Direction 2. Systematic review screening prioritisation.

Screening prioritisation input & method [17,18,19]

[14] Wang, et al. “Mesh term suggestion for systematic review literature search,” ADCS, 2021.
[15] Wang, et al. “Automated mesh term suggestion for effective query formulation in systematic reviews literature search.”, ISWA, 2022.
[16] Wang, et al. “MeSH Suggester: A Library and System for MeSH Term Suggestion for Systematic Review Boolean Query Construction,” WSDM, 2023.
[17] Wang, et al. “From little things big things grow: A collection with seed studies for medical systematic review literature search,” SIGIR, 2022.
[18] Wang, et al. “Seed-driven document ranking for systematic reviews: A reproducibility study,” ECIR, 2022.
[19] Wang, et al. “Neural Rankers for Effective Screening Prioritization in Medical Systematic Review Literature Search,” ADCS, 2022.

Boolean Search

("Aspergillus"[mesh] OR **"Aspergillosis"[mesh]** OR **"Pulmonary Aspergillosis"[mesh]** OR aspergill*[Title/Abstract] OR fungal infection[Text Word] OR (invasive[Title/Abstract] AND fungal[Title/Abstract]))

AND

(Platelia[Text Word] OR **"Mannans"[mesh]** OR galactomannan[Text Word])

AND

(**"Immunoassay"[mesh]** OR immunoassay[Title/Abstract] OR immunoassays[Title/Abstract] OR immuno assay[Title/Abstract] OR immuno assays[Title/Abstract] OR ELISA[Title/Abstract] OR ELISAs[Title/Abstract] OR EIA[Title/Abstract] OR EIAs[Title/Abstract] OR immunosorbent[Title/Abstract])

AND

(“**Serology"[mesh]** OR **Serology"[mesh]** OR serology[Title/Abstract] OR serodiagnosis[Title/Abstract] OR serologic[Title/Abstract])

MeSH: Medical Subject Headings (around 40,000 t.d)

Our Goal

aspergill*[Title/Abstract] OR fungal infection[Text Word] OR (invasive[Title/Abstract] AND fungal[Title/Abstract]))

AND

(Platelia[Text Word] OR galactomannan[Text Word])

AND

(immunoassay[Title/Abstract] OR immunoassays[Title/Abstract] OR immuno assay[Title/Abstract] OR immuno assays[Title/Abstract] OR ELISA[Title/Abstract] OR ELISAs[Title/Abstract] OR EIA[Title/Abstract] OR EIAs[Title/Abstract] OR immunosorbent[Title/Abstract])

AND

(serology[Title/Abstract] OR serodiagnosis[Title/Abstract] OR serologic[Title/Abstract])

Suggesting MeSH terms

("Aspergillus"[mesh] OR **"Aspergillosis"[mesh]** OR **"Pulmonary Aspergillosis"[mesh]** OR aspergill*[Title/Abstract] OR fungal infection[Text Word] OR (invasive[Title/Abstract] AND fungal[Title/Abstract]))

AND

(Platelia[Text Word] OR **"Mannans"[mesh]** OR galactomannan[Text Word])

AND

("Immunoassay"[mesh] OR immunoassay[Title/Abstract] OR immunoassays[Title/Abstract] OR immuno assay[Title/Abstract] OR immuno assays[Title/Abstract] OR ELISA[Title/Abstract] OR ELISAs[Title/Abstract] OR EIA[Title/Abstract] OR EIAs[Title/Abstract] OR immunosorbent[Title/Abstract])

AND

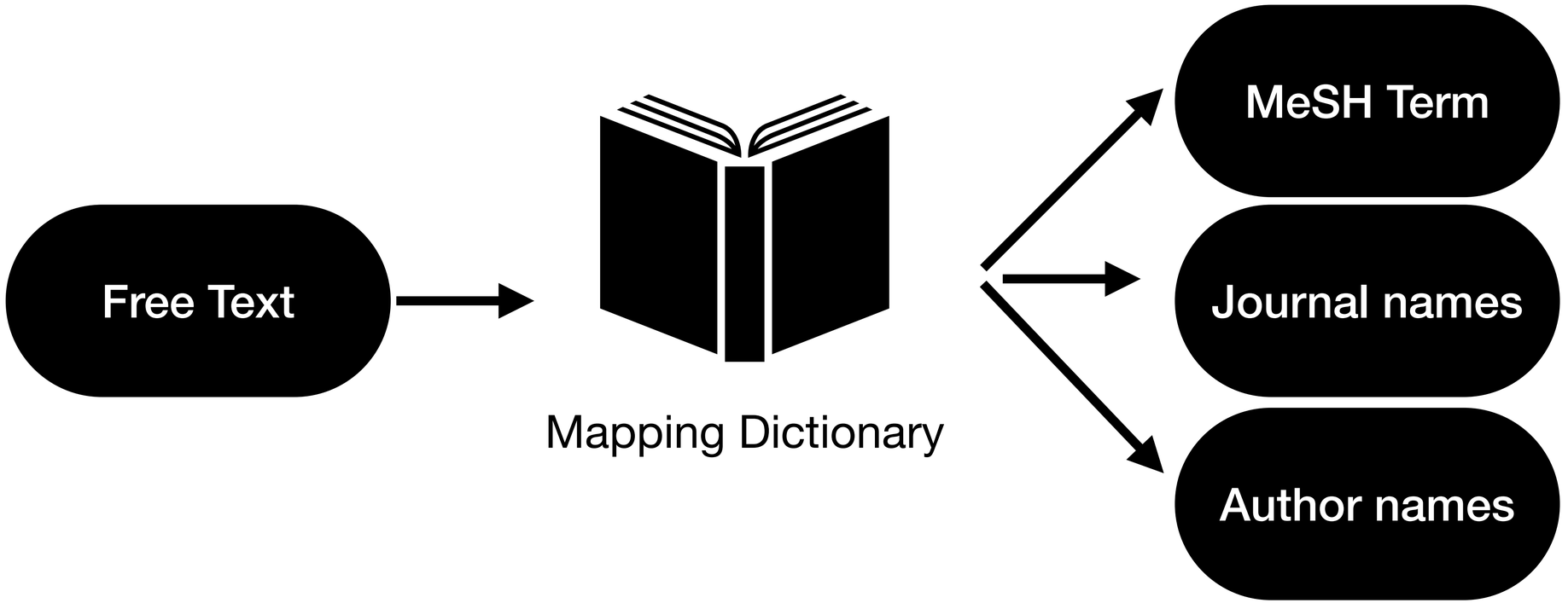
(“**Serology"[mesh]** OR **Serology"[mesh]** OR serology[Title/Abstract] OR serodiagnosis[Title/Abstract] OR serologic[Title/Abstract])

ie lab

The University Of Queensland AUSTRALIA

Current Method

Automated Term Mapping (ATM)



ie lab

The University Of Queensland AUSTRALIA

CREATE CHANGE

Current Method

ATM Limitations:

1. Resolution of acronyms can be inaccurate.

   Example: BE: barium enema

2. Confusion when synonymic (semantically matching) free-text terms are used

   Example: “liver neoplasms” vs “hepatic cancer”

3. Difficulties in disambiguating between MeSH terms and journal names.

   Example: Blood (Journal title)

Proposed Methods

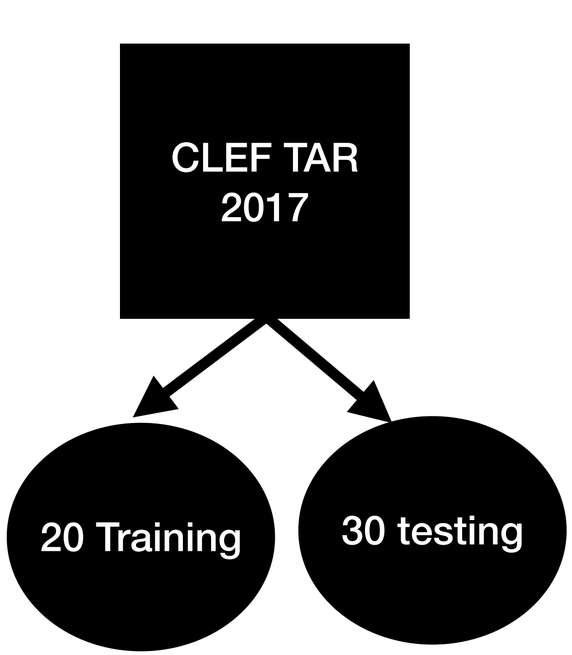
1. Lexical method

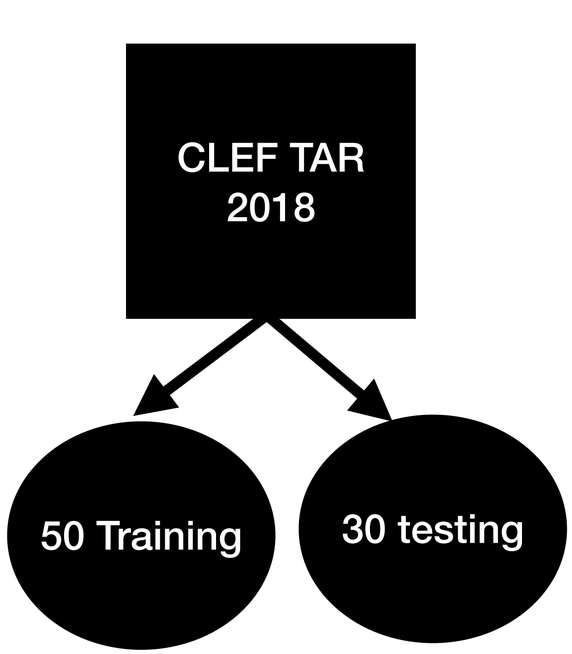
    - MeSH Term retrieval
    - MeSH Term ranking
    - MeSH Term refinement

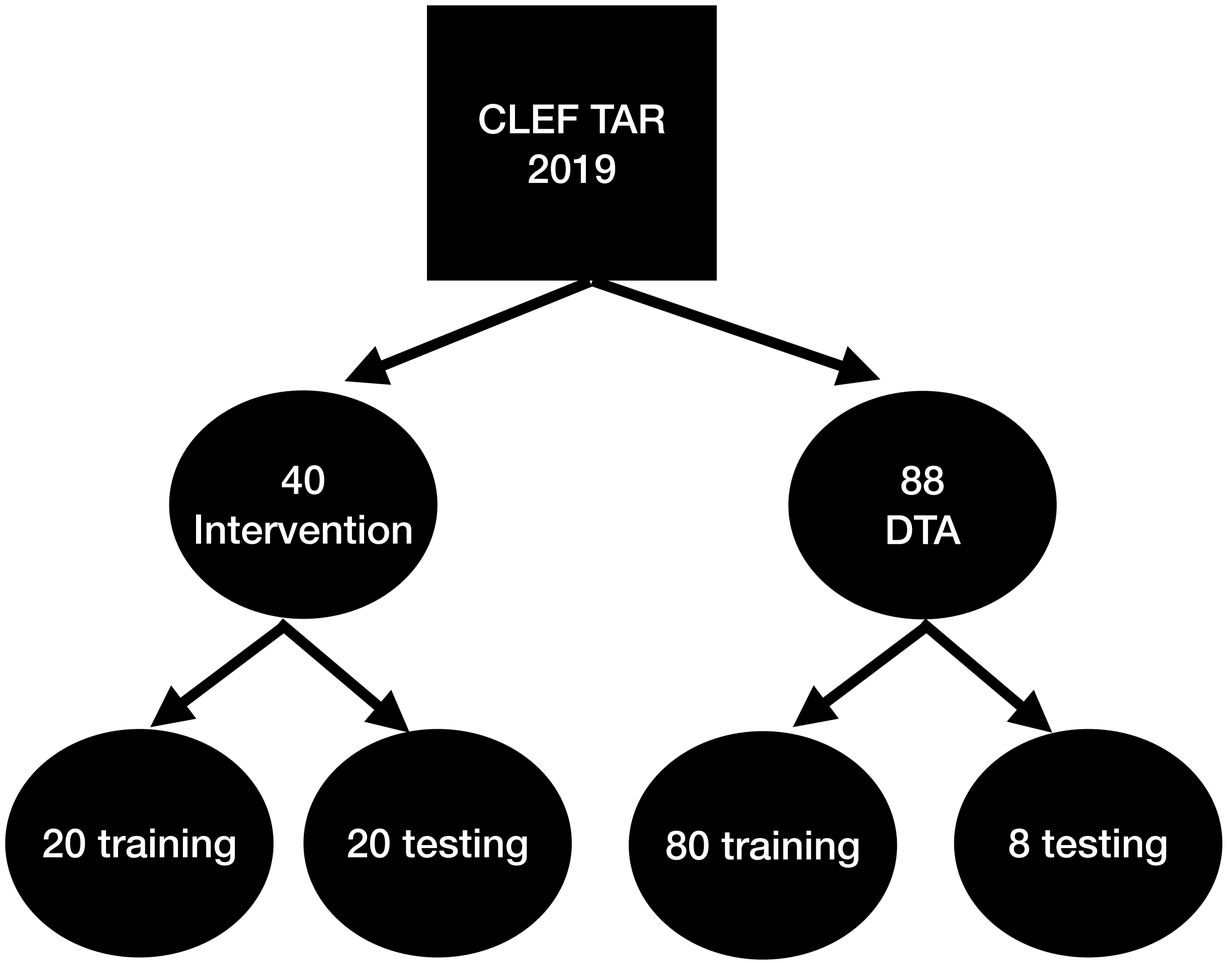
2. BERT method

    - Different ranking representation
    - Different cutoff strategy

CLEF Dataset



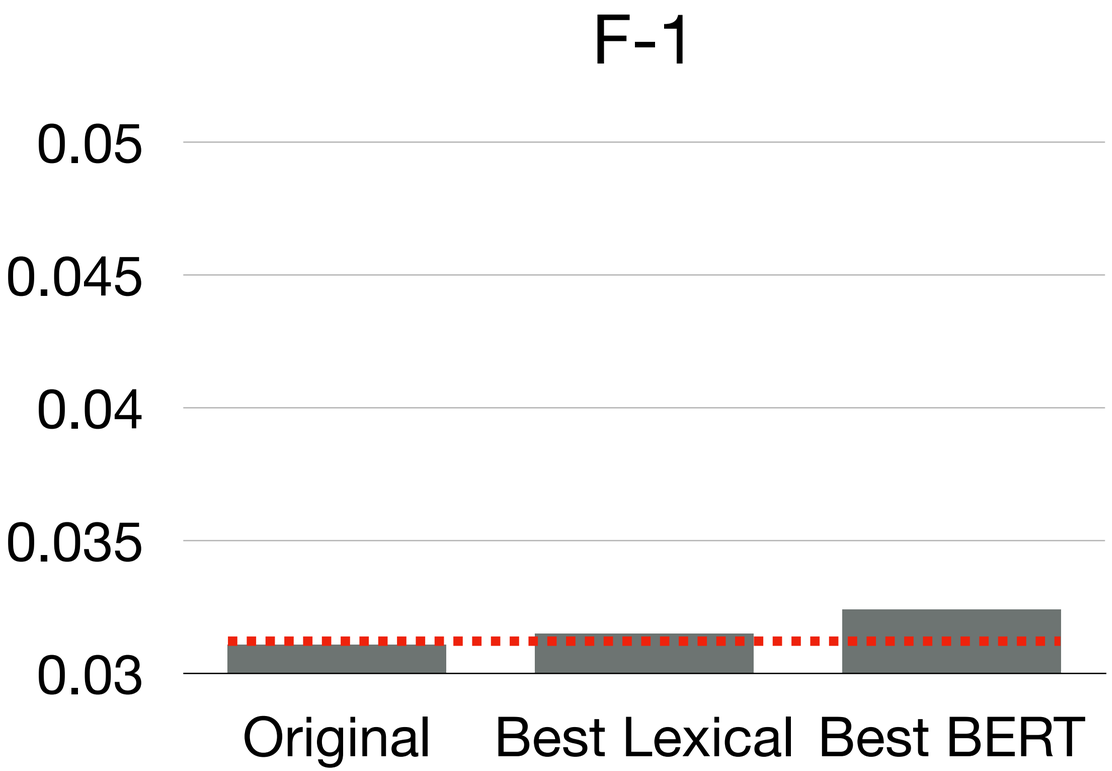


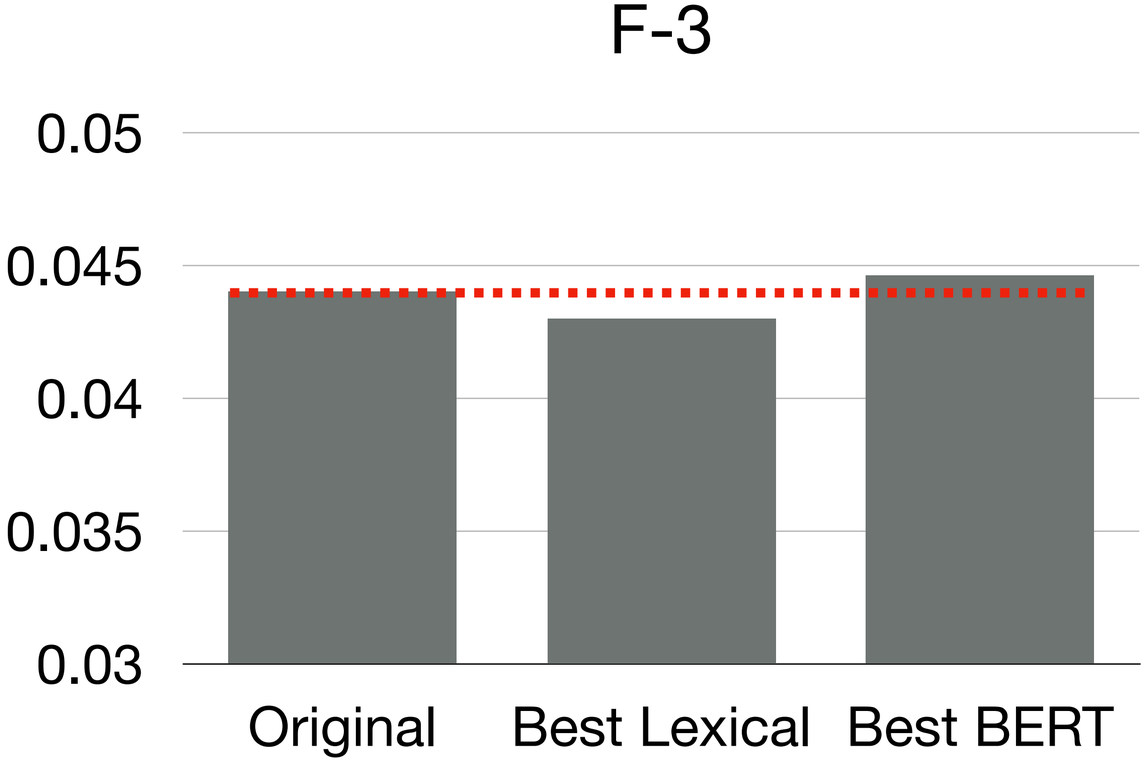


ie lab

The University Of Queensland
AUSTRALIA
CREATE CHANGE

Findings





Automatically suggested MeSH Terms can outperform MeSH Terms in the original query

Question

- Can we use ChatGPT to generate this complex systematic review Boolean Query?

ChatGPT Prompts

- Five query formulation Prompts
- Two query refinement Prompts
- One guided prompts (multiple iterations)

Query Formulation

	Prompt ID	Prompt
Simple	q1	For a systematic review titled "{review_title}", can you generate a systematic review Boolean query to find all included studies on PubMed for the review topic?
Detailed	q2	You are an information specialist who develops Boolean queries for systematic reviews. You have extensive experience developing highly effective queries for searching the medical literature. Your specialty is developing queries that retrieve as few irrelevant documents as possible and retrieve all relevant documents for your information need. Now you have your information need to conduct research on {review_title}. Please construct a highly effective systematic review Boolean query that can best serve your information need.
	q3	Imagine you are an expert systematic review information specialist; now you are given a systematic review research topic, with the topic title "{review_title}". Your task is to generate a highly effective systematic review Boolean query to search on PubMed (refer to the professionally made ones); the query needs to be as inclusive as possible so that it can retrieve all the relevant studies that can be included in the research topic; on the other hand, the query needs to retrieve fewer irrelevant studies so that researchers can spend less time judging the retrieved documents.
With Example	q4	You are an information specialist who develops Boolean queries for systematic reviews. You have extensive experience developing highly effective queries for searching the medical literature. Your specialty is developing queries that retrieve as few irrelevant documents as possible and retrieve all relevant documents for your information need. You are able to take an information need such as: "{example_review_title}" and generate valid pubmed queries such as: "{example_review_query}". Now you have your information need to conduct research on "{review_title}", please generate a highly effective systematic review Boolean query for the information need.
	q5	You are an information specialist who develops Boolean queries for systematic reviews. You have extensive experience developing highly effective queries for searching the medical literature. Your specialty is developing queries that retrieve as few irrelevant documents as possible and retrieve all relevant documents for your information need. A professional information specialist will extract PICO elements from information needs in a common practice in constructing a systematic review Boolean query. PICO means Patient/ Problem, Intervention, Comparison and Outcome. PICO is a format for developing a good clinical research question prior to starting one's research. It is a mnemonic used to describe the four elements of a sound clinical foreground question. You are able to take an information need such as: "{example_review_title}" and you generate valid pubmed queries such as: "{example_review_query}". Now you have your information need to conduct research on "{review_title}". First, extract PICO elements from the information needs and construct a highly effective systematic review Boolean query that can best serve your information need.

Table 1. Prompts for single prompt query formulation

Query Refinement

	Prompt ID	Prompt
Simple	q6	For a systematic review seed Boolean query: "{initial_query}", This query retrieves too many irrelevant documents and too few relevant documents about the information need: “{review_title}”, Please correct this query so that it can retrieve fewer irrelevant documents and more relevant documents.
With Example	q7	For a systematic review seed Boolean query: “{example_review_initial_query}" ,This query retrieves too many irrelevant documents and too few relevant documents about the information need: “{example_review_title}”, therefore it should be corrected to: “{example_review_refined_query}”. Now your task is to correct a systematic review Boolean query: "{initial_query}" for information need “{review_title}”, so it can retrieve fewer irrelevant documents and more relevant documents.

Table 2. Prompts for single prompt query refinement

ChatGPT Prompts

Step	Prompt	Example Answer
Step1	Follow my instructions precisely to develop a highly effective Boolean query for a medical systematic review literature search. Do not explain or elaborate. Only respond with exactly what I request. First, Given the following statement and text from a relevant study, please identify 50 terms or phrases that are relevant. The terms you identify should be used to retrieve more relevant studies, so be careful that the terms you choose are not too broad. You are not allowed to have duplicates in your list. statement: "Prevalence of Differentiated Thyroid Cancer in Autopsy Studies Over Six Decades: A Meta-Analysis" Text: Ten occult carcinomas of the thyroid gland were found in 274 unselected autopsies at the Pathology Service, Hospital José Joaquín Aguirre, between December 1980 and March 1983. This is the lowest incidence among the most recent published series. The smallest papillary microcarcinomas presented most frequently as nonencapsulated nonsclerosing tumors without inflammatory infiltrate or fibrosis, suggesting that they may represent the early stages of development. Our data show a relatively high and similar frequency of papillary microcarcinomas in surgical and autopsy series, but do not demonstrate risk factors for clinical evolution.	1. Differentiated thyroid cancer 2. Prevalence 3. Autopsy studies 4. Meta-analysis 5. Occult carcinomas 6. Thyroid gland 7. Incidence 8. Environmental factors 9. Etiology 10. Low risk ... 50. Parameters.
Step 2	For each item in the list you created in step 1, classify it into as of three categories: terms relating to health conditions (A), terms relating to a treatment (B), terms relating to types of study design (C). When an item does not fit one of these categories, mark it as (N/A). Each item needs to be categorised into (A), (B), (C), or (N/A).	1. (A) Differentiated thyroid cancer 2. (N/A) Prevalence 3. (C) Autopsy studies 4. (C) Meta-analysis 5. (A) Occult carcinomas 6. (A) Thyroid gland 7. (N/A) Incidence 8. (N/A) Environmental factors 9. (N/A) Etiology 10. (N/A) Low risk ... 50. (N/A) Parameters.
Step 3	Using the categorised list you created in step 2, create a Boolean query that can be submitted to PubMed which groups together items from each category. For example: ((itemA1[Title/Abstract] OR itemA2[Title/Abstract] or itemA2[Title/Abstract]) AND (itemB1[Title/Abstract] OR itemB2[Title/Abstract] OR itemB3[Title/Abstract]) AND (itemC1[Title/Abstract] OR itemC2[Title/Abstract] OR itemC3[Title/Abstract]))	(((differentiated[Title/Abstract] OR thyroid[Title/Abstract] OR carcinoma[Title/Abstract] OR papillary[Title/Abstract] OR microcarcinoma[Title/Abstract]) AND (cancer[Title/Abstract] OR incidence[Title/Abstract] OR etiology[Title/Abstract] OR risk[Title/Abstract] OR gender[Title/Abstract] OR hormonal[Title/Abstract] OR nodular[Title/Abstract] OR goiter[Title/Abstract] OR Hashimoto's[Title/Abstract] OR malignancy[Title/Abstract] OR concomitant[Title/Abstract] OR tumor[Title/Abstract] OR infiltrate[Title/Abstract] OR fibrosis[Title/Abstract] OR development[Title/Abstract] OR frequency[Title/Abstract])) AND (autopsy[Title/Abstract] OR surgical[Title/Abstract] OR material[Title/Abstract] OR series[Title/Abstract] OR specimens[Title/Abstract] OR cases[Title/Abstract])).
Step 4	Use your expert knowledge to refine the query, making it retrieve as many relevant documents as possible while minimising the total number of documents retrieved. Also add relevant MeSH terms into the query where necessary, e.g., MeSHTerm[MeSH]. Retain the general structure of the query, however, with each main clause of the query corresponding to a PICO element. The final query still needs to be executable on PubMed, so it should be a valid query.	(((differentiated thyroid cancer[MeSH] OR "differentiated thyroid"[All Fields] OR "thyroid carcinoma"[All Fields] OR "papillary microcarcinoma"[All Fields]) AND (prevalence[All Fields] OR incidence[MeSH] OR "etiology of"[All Fields] OR "risk factors"[All Fields] OR gender[All Fields] OR hormonal[All Fields] OR "nodular goiter"[All Fields] OR "Hashimoto's thyroiditis"[MeSH] OR malignancy[MeSH] OR "concomitant lesion"[All Fields] OR tumor[All Fields] OR infiltrate[All Fields] OR fibrosis[All Fields] OR "early stages of development"[All Fields] OR frequency[All Fields])) AND (autopsy[MeSH] OR surgical[All Fields] OR material[All Fields] OR series[All Fields] OR specimens[All Fields] OR cases[All Fields]))

Table 3. Example designed guided prompt for query formulation.

ChatGPT Prompts

- Five query formulation Prompts
- Two query refinement Prompts
- One guided prompts (multiple iterations)

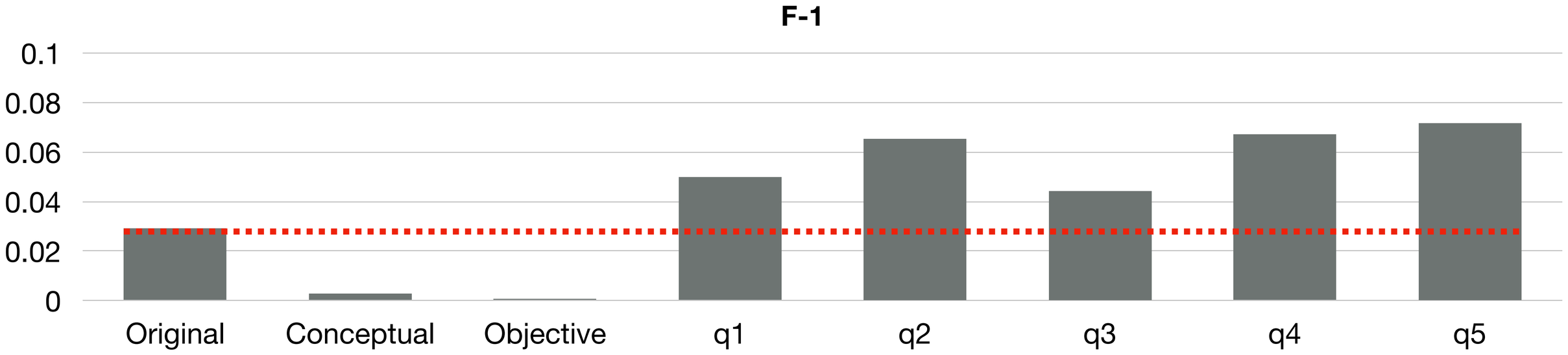
Tested on 72 SR topics on CLEF TAR 2017, 2018.
40 SR topics on Seed Collection [17]

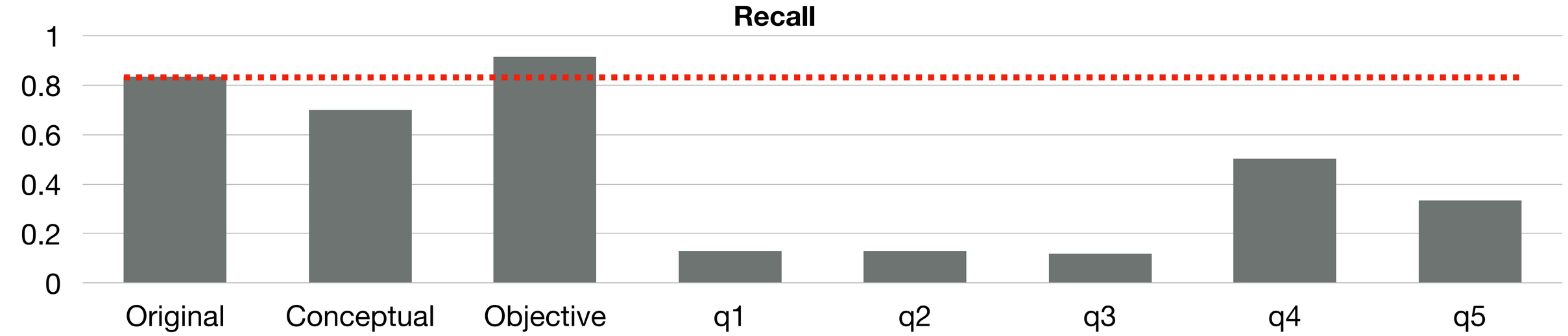
→

Involves around 5000 requests to ChatGPT

[17] Wang, et al. "From little things big things grow: A collection with seed studies for medical systematic review literature search," SIGIR, 2022.

Findings (query formulation)





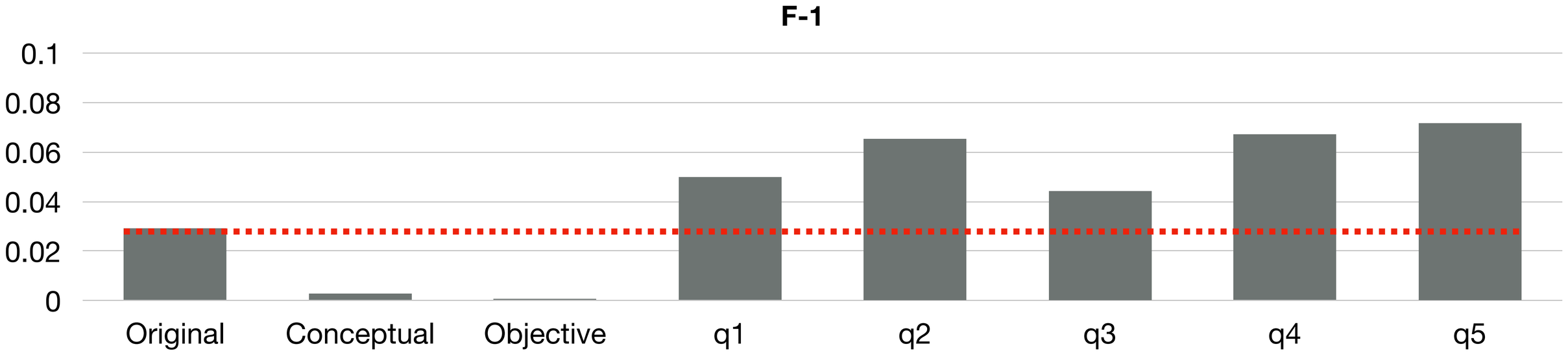
Query Formulation

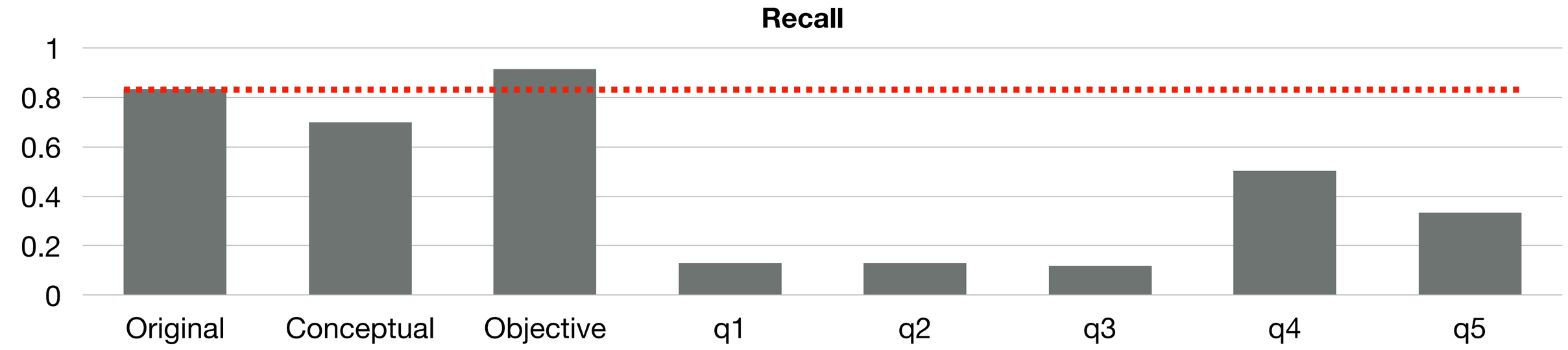
	Prompt ID	Prompt
Simple	q1	For a systematic review titled “{review_title}”, can you generate a systematic review Boolean query to find all included studies on PubMed for the review topic?
Detailed	q2	You are an information specialist who develops Boolean queries for systematic reviews. You have extensive experience developing highly effective queries for searching the medical literature. Your specialty is developing queries that retrieve as few irrelevant documents as possible and retrieve all relevant documents for your information need. Now you have your information need to conduct research on {review_title}. Please construct a highly effective systematic review Boolean query that can best serve your information need.
	q3	Imagine you are an expert systematic review information specialist; now you are given a systematic review research topic, with the topic title “{review_title}”. Your task is to generate a highly effective systematic review Boolean query to search on PubMed (refer to the professionally made ones); the query needs to be as inclusive as possible so that it can retrieve all the relevant studies that can be included in the research topic; on the other hand, the query needs to retrieve fewer irrelevant studies so that researchers can spend less time judging the retrieved documents.
With Example	q4	You are an information specialist who develops Boolean queries for systematic reviews. You have extensive experience developing highly effective queries for searching the medical literature. Your specialty is developing queries that retrieve as few irrelevant documents as possible and retrieve all relevant documents for your information need. You are able to take an information need such as: “{example_review_title}” and generate valid pubmed queries such as: “{example_review_query}". Now you have your information need to conduct research on “{review_title}”, please generate a highly effective systematic review Boolean query for the information need.
	q5	You are an information specialist who develops Boolean queries for systematic reviews. You have extensive experience developing highly effective queries for searching the medical literature. Your specialty is developing queries that retrieve as few irrelevant documents as possible and retrieve all relevant documents for your information need. A professional information specialist will extract PICO elements from information needs in a common practice in constructing a systematic review Boolean query. PICO means Patient/ Problem, Intervention, Comparison and Outcome. PICO is a format for developing a good clinical research question prior to starting one’s research. It is a mnemonic used to describe the four elements of a sound clinical foreground question. You are able to take an information need such as: “{example_review_title}" and you generate valid pubmed queries such as: “{example_review_query}". Now you have your information need to conduct research on “{review_title}”. First, extract PICO elements from the information needs and construct a highly effective systematic review Boolean query that can best serve your information need.

Table 1. Prompts for single prompt query formulation

ie lab

Findings (query formulation)



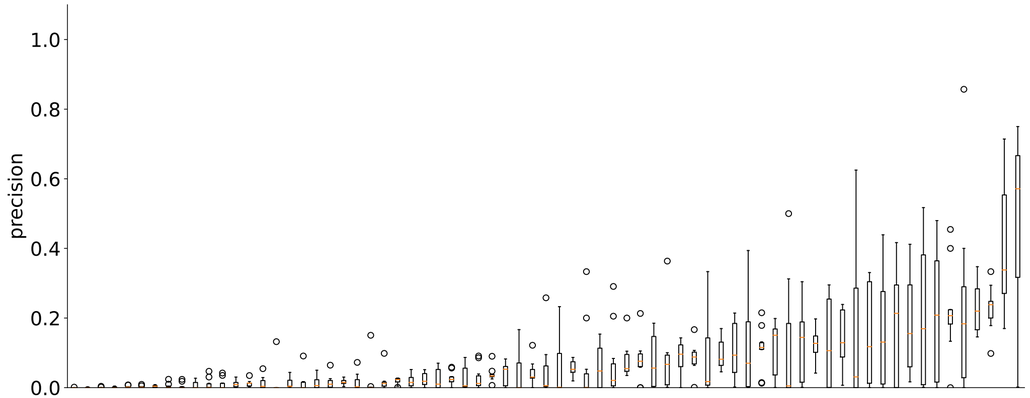


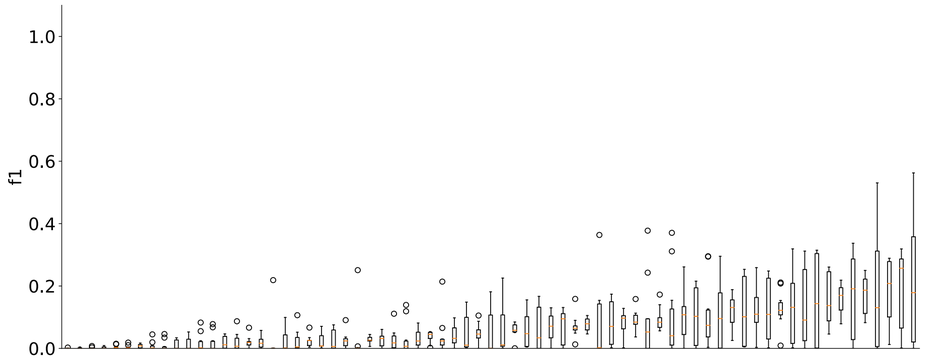
Findings

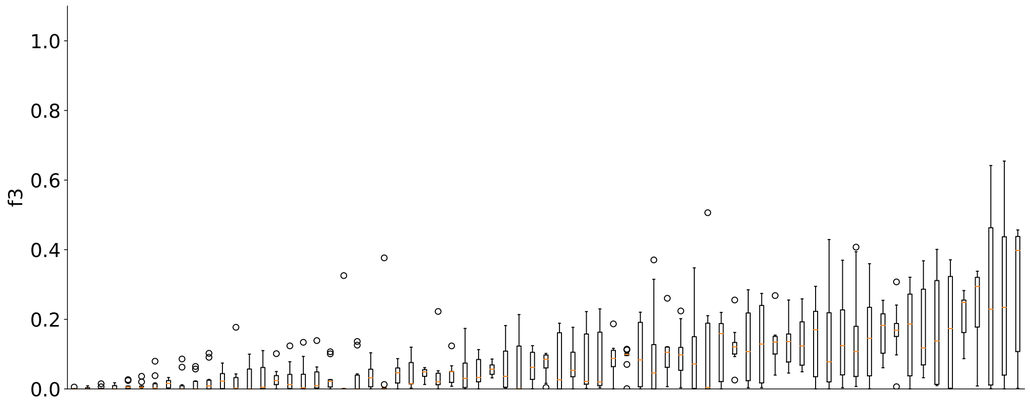
- ChatGPT can help to generate “more effective” queries; using appropriate prompts is critical to generate high-quality Boolean queries.

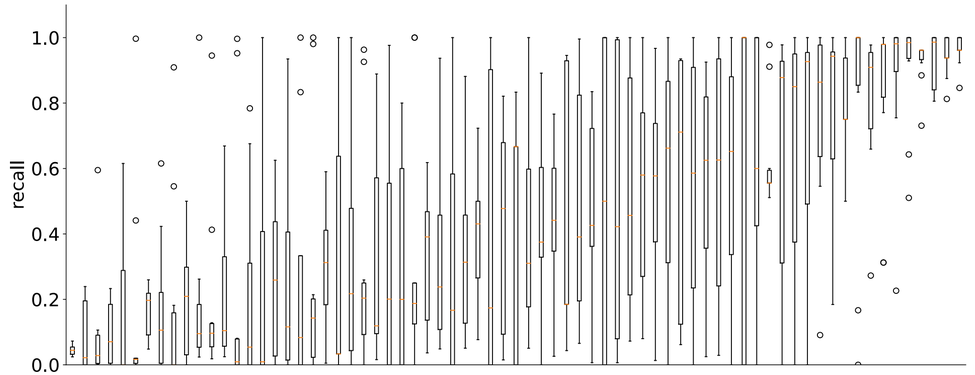
Findings (variability)

10 iteration of q4









ie lab

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Findings

- ChatGPT can help to generate “more effective” queries; using appropriate prompts is critical to generate high-quality Boolean queries.

- ChatGPT can not guarantee to generate a high-quality systematic review Boolean query; the resulting queries' effectiveness is often not stable.

Failing point

- “Poorly performing” queries obtain more un-judged documents than “good” queries

- 55% of MeSH terms generated from ChatGPT were not in the MeSH vocabulary.

What's Next

- Use a domain-specific GPT model to start training a Bio-ChatGPT?

- Fine-tune a BioGPT model to generate prompts for ChatGPT?

- Can we leverage the ChatGPT model to generate training data for our task?

Many other possibilities!

Thank you

Questions?

ie lab

For more information, please visit
https://ielab.io/people/shuai-wang

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Appendix

Papers published

Published:

1. **S. Wang**, H. Scells, J. Clark, B. Koopman, and G. Zuccon, "From little things big things grow: A collection with seed studies for medical systematic review literature search," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 3176–3186. [Online]. Available: https://doi.org/10.1145/3477495.3531748 **(SIGIR 2022)**
2. **S. Wang**, H. Li, H. Scells, D. Locke, and G. Zuccon, "Mesh term suggestion for systematic review literature search," in *Proceedings of the 25th Australasian Document Computing Symposium*, 2021, pp. 1–8. **(ADCS 2021, Best Student Paper award)**
3. **S. Wang**, H. Scells, B. Koopman, and G. Zuccon, " Automated mesh term suggestion for effective query formulation in systematic reviews literature search." *Intelligent Systems with Applications*, p. 200141, 2022. **(ISWA 2022)**
4. **S. Wang**, H. Scells, A. Mourad, and G. Zuccon, "Seed-driven document ranking for systematic reviews: A reproducibility study," in *European Conference on Information Retrieval*. Springer, 2022, pp. 686–700. **(ECIR 2022)**
5. **S. Wang**, S. Zhuang, and G. Zuccon. 2021. Bert-based dense retrievers require interpolation with bm25 for effective passage retrieval. In Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval. 317–324. **(ICTIR 2021)**
6. H. Li*, **S. Wang***, S. Zhuang, A. Mourad, X. Ma, J. Lin, and G. Zuccon. 2022. To Interpolate or Not to Interpolate: PRF, Dense and Sparse Retrievers. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (Madrid, Spain) (SIGIR '22). Association for Computing Machinery, New York, NY, USA, 2495–2500. https://doi.org/10.1145/3477495.3531884 **(SIGIR 2022)**
7. **S. Wang**, H. Scells, B. Koopman, and G. Zuccon, "Neural Rankers for Effective Screening Prioritization in Medical Systematic Review Literature Search," accepted in *Proceedings of the 26th Australasian Document Computing Symposium*, 2022. **(ADCS 2022)**
8. **S. Wang**, H. Li, B. Koopman, and G. Zuccon, "MeSH Suggester: A Library and System for MeSH Term Suggestion for Systematic Review Boolean Query Construction," Under review in *WSDM*, 2023. **(WSDM 2023)**

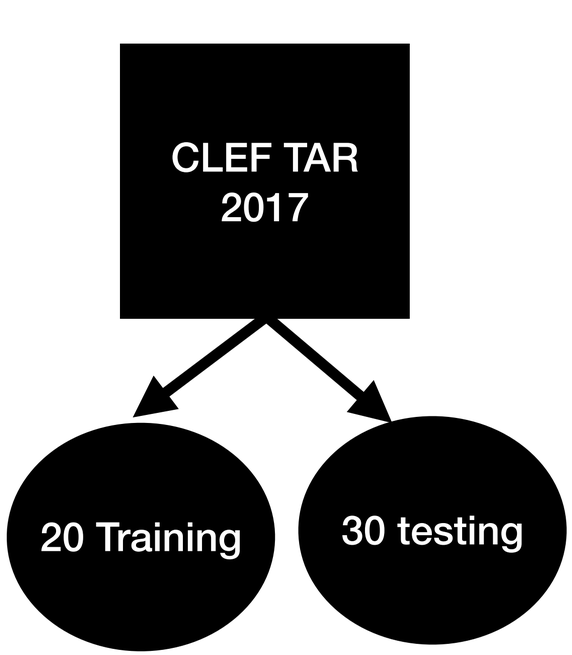
References

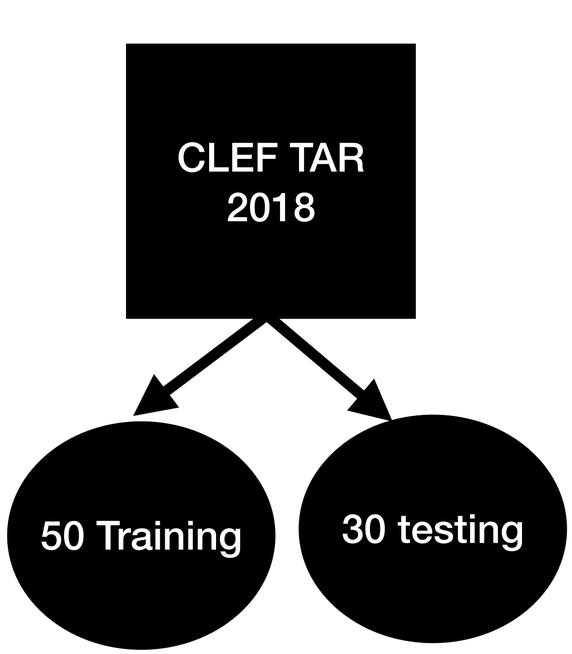
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018).
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019).
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Mi- chael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res. 21, 140 (2020), 1–67.
- I. Shemilt, N. Khan, S. Park, and J. Thomas, "Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews," *Systematic reviews*, vol. 5, no. 1, p. 140, 2016.
- H. Scells, G. Zuccon, B. Koopman, and J. Clark, "A computational approach for objectively derived systematic review search strategies," in *European conference on information retrieval*. Springer, 2020, pp. 385–398.
- H. Scells, G. Zuccon, and B. Koopman, "Automatic boolean query refinement for systematic review literature search," in *The Web Conference*, ser. WebConf '19, 2019, pp. 1646–1656.
- A. Alharbi, W. Briggs, and M. Stevenson, "Retrieving and ranking studies for systematic re- views: University of sheffield's approach to clef ehealth 2018 task 2," in *CEUR Workshop Pro- ceedings*, vol. 2125. CEUR Workshop Proceedings, 2018.
- A. Alharbi and M. Stevenson, "Ranking abstracts to identify relevant evidence for systematic reviews: The university of sheffield's approach to clef ehealth 2017 task 2." in *CLEF (Working Notes)*, 2017.
- H. Scells, G. Zuccon, A. Deacon, and B. Koopman, "Qut ielab at clef ehealth 2017 technology assisted reviews track: initial experiments with learning to rank," in *Working Notes of CLEF 2017-Conference and Labs of the Evaluation Forum [CEUR Workshop Proceedings, Volume [1866]*. Sun SITE Central Europe, 2017, pp. 1–6.
- H. Wu, T. Wang, J. Chen, S. Chen, Q. Hu, and L. He, "Ecnu at 2018 ehealth task 2: Technolog- ically assisted reviews in empirical medicine," *Methods*, vol. 4, no. 5, p. 7, 2018.
- A. Alharbi and M. Stevenson, "Ranking studies for systematic reviews using query adaptation: University of sheffield's approach to clef ehealth 2019 task 2 working notes for clef 2019," in *Working Notes of CLEF 2019-Conference and Labs of the Evaluation Forum*, vol. 2380. CEUR Workshop Proceedings, 2019.

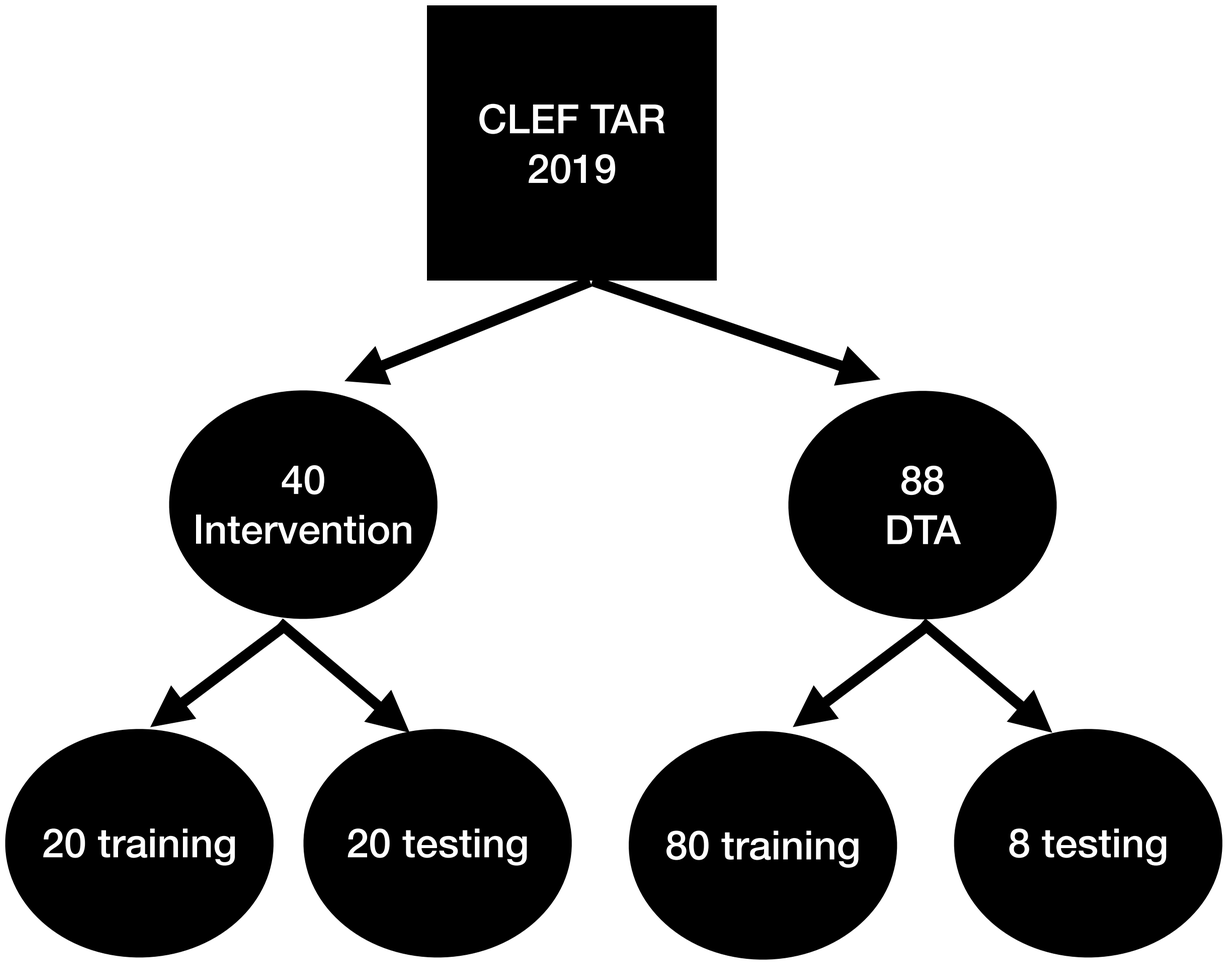
References

- H. Scells and G. Zuccon, “You can teach an old dog new tricks: Rank fusion applied to coordi- nation level matching for ranking in systematic reviews,” in *42nd European Conference on IR Research, ECIR 2020*, 2020.
- Alan R Aronson. 2001. AMIA Symposium. Effective mapping of biomedical text to the UMLS Meta- thesaurus: the MetaMap program.
- Olivier Bodenreider. 2004. Nucleic acids research 32, suppl_1. The unified medical language system (UMLS): integrating biomedical terminology.
- Balog. 2018. Springer. Entity-oriented search.
- E. Kanoulas, D. Li, L. Azzopardi, and R. Spijker. 2017. CLEF 2017 Technologically Assisted Reviews in Empirical Medicine Overview. In CLEF’17.
- Evangelos Kanoulas, Rene Spijker, Dan Li, and Leif Azzopardi. 2018. CLEF 2018 Technology Assisted Reviews in Empirical Medicine Overview. In CLEF 2018 Evaluation Labs and Workshop: Online Working Notes, CEUR-WS.
- Evangelos Kanoulas, Dan Li, Leif Azzopardi, and Rene Spijker. 2019. CLEF 2019 technology assisted reviews in empirical medicine overview. In CEUR Workshop Proceedings, Vol. 2380.
- I. Beltagy, M. E. Peters, and A. Cohan, “Longformer: The long-document transformer,” *arXiv preprint arXiv:2004.05150*, 2020.
- V. Zhong, C. Xiong, and R. Socher, “Seq2sql: Generating structured queries from natural lan- guage using reinforcement learning,” *arXiv preprint arXiv:1709.00103*, 2017.
- T. Scholak, R. Li, D. Bahdanau, H. de Vries, and C. Pal, “Duorat: Towards simpler text-to-sql models,” *arXiv preprint arXiv:2010.11119*, 2020.
- B. Qin, B. Hui, L. Wang, M. Yang, J. Li, B. Li, R. Geng, R. Cao, J. Sun, L. Si *et al.*, “A survey on text-to-sql parsing: Concepts, methods, and future directions,” *arXiv preprint arXiv:2208.13629*, 2022.

CLEF Dataset





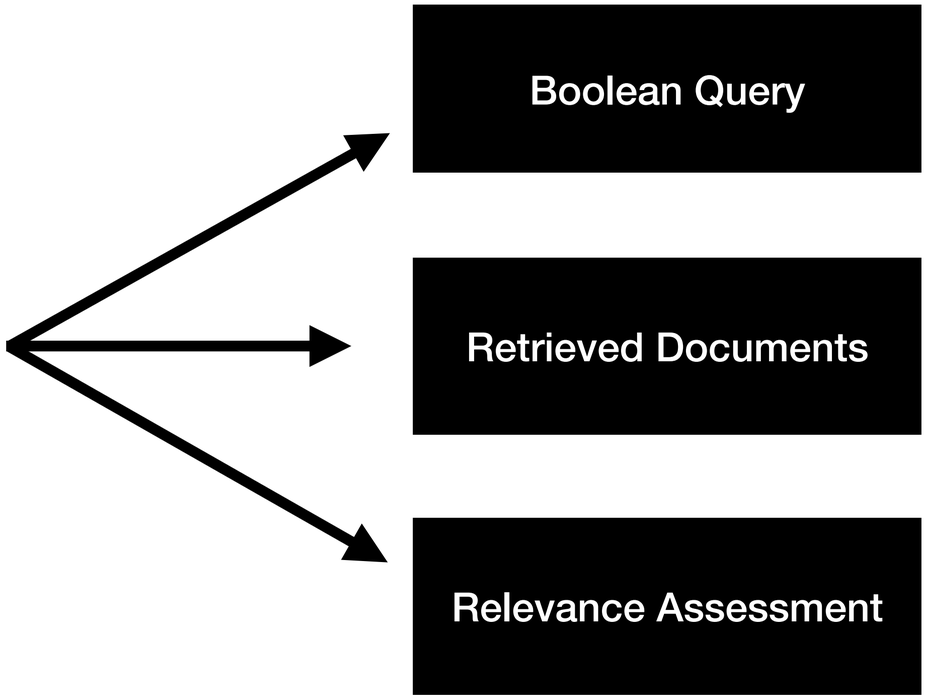


ielab

The University Of Queensland
AUSTRALIA
CREATE CHANGE

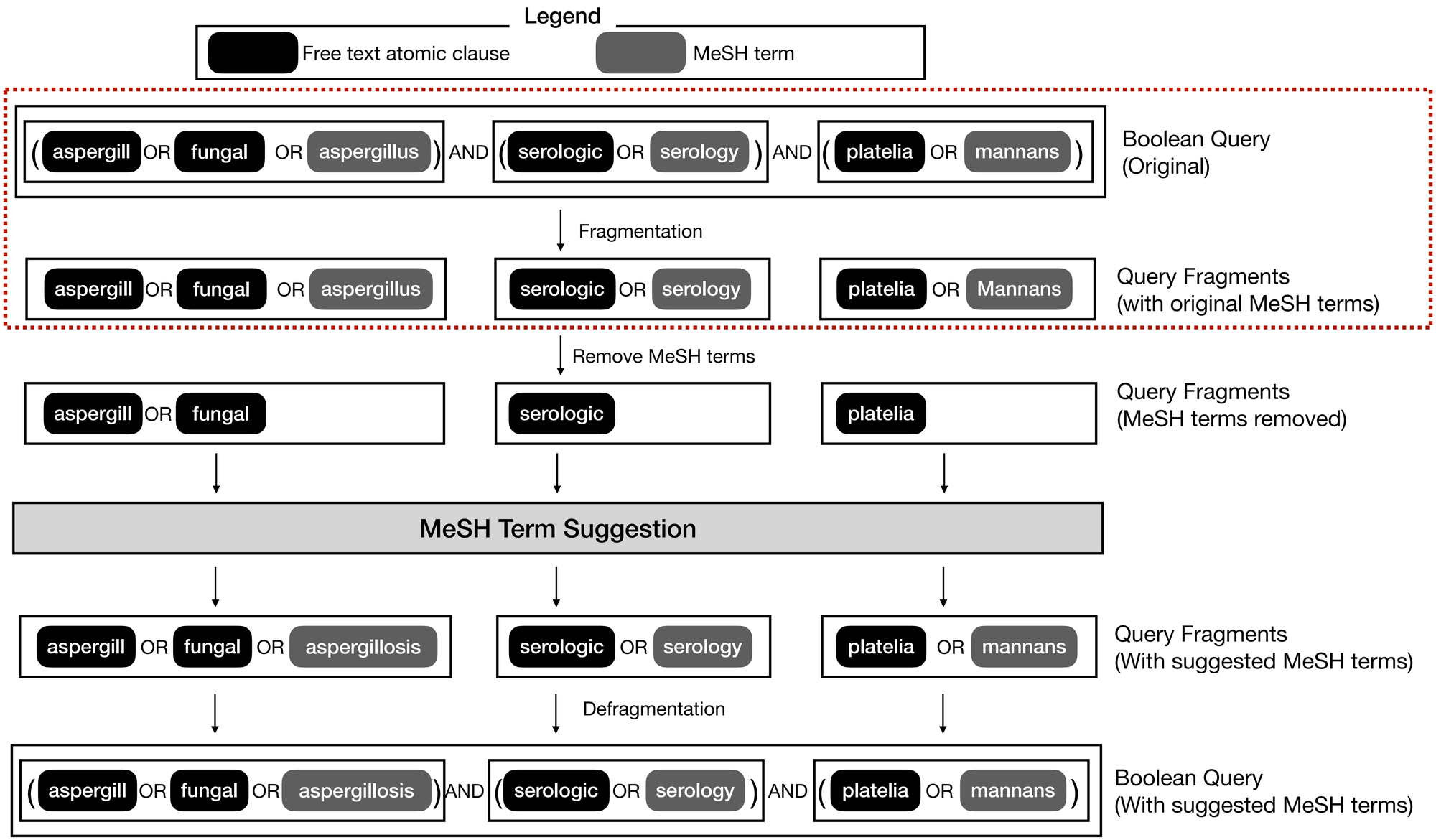
Dataset

CLEF TAR Datasets [23,24,25]

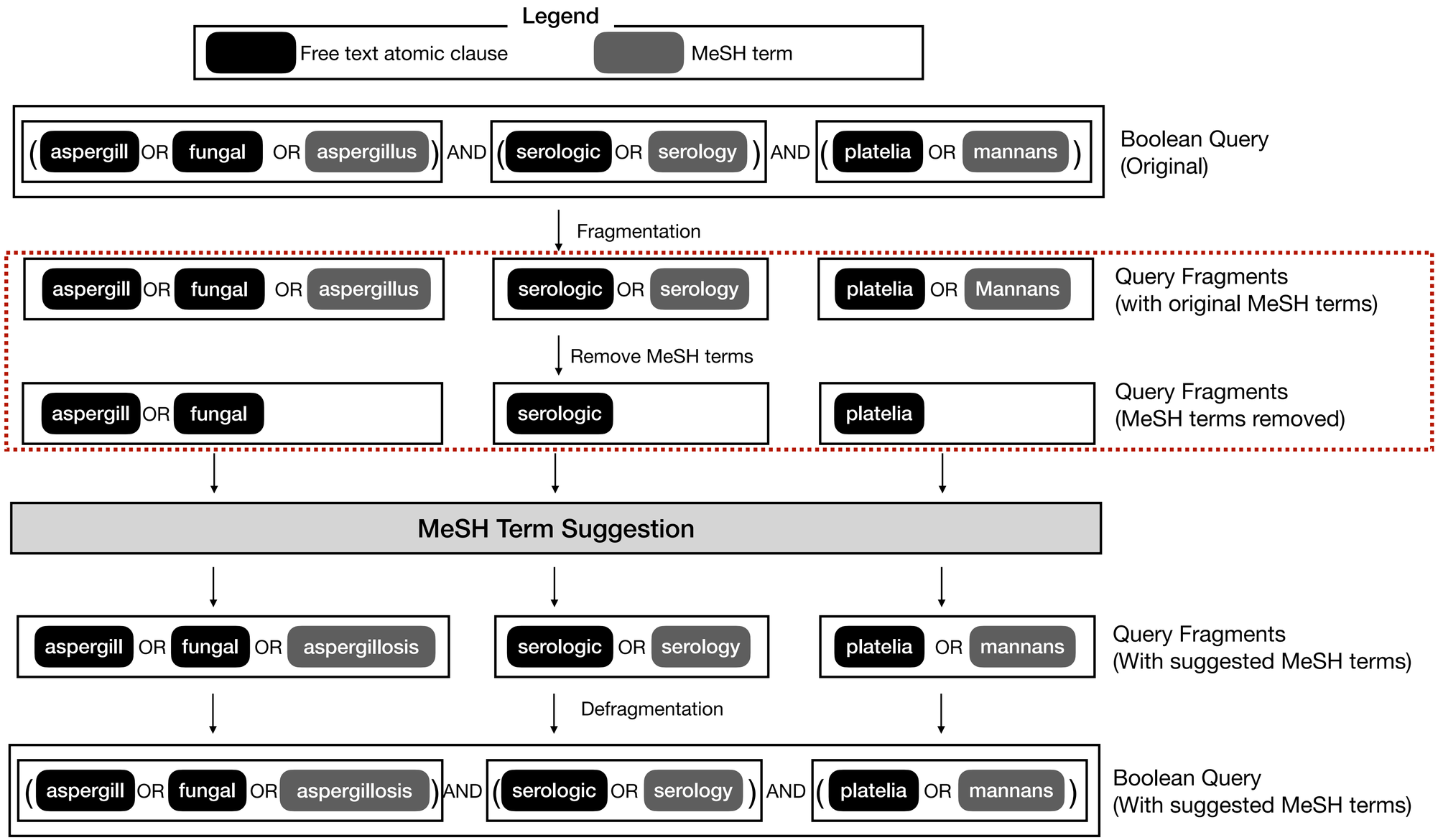


[23] Kanoulas, et al. CLEF 2017 Technologically Assisted Reviews in Empirical Medicine Overview, 2017.
[24] Kanoulas, et al. CLEF 2018 Technology Assisted Reviews in Empirical Medicine Overview, 2018.
[25] Kanoulas, et al. CLEF 2019 technology assisted reviews in empirical medicine overview, 2019.

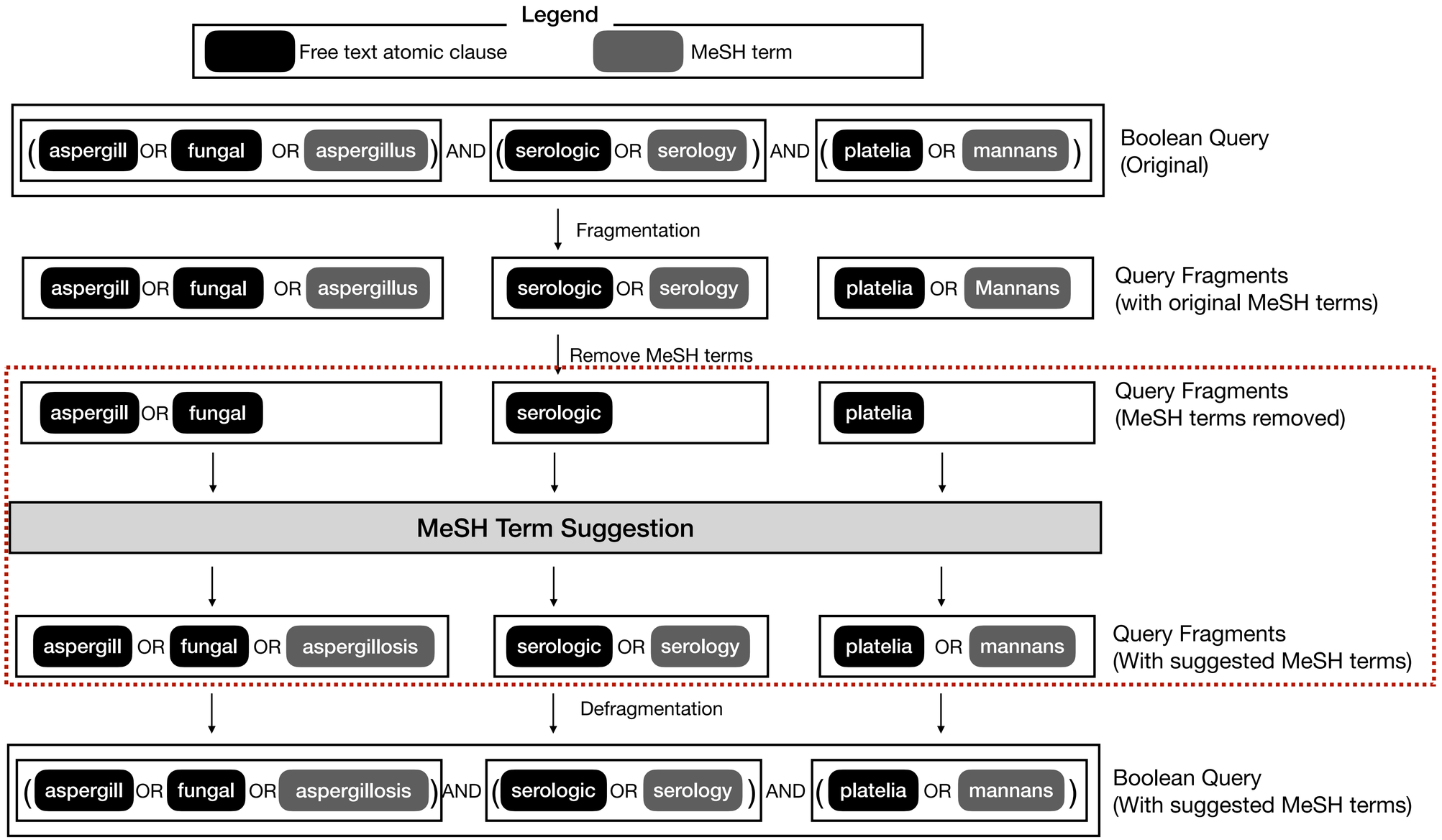
Pipeline


Legend
Free text atomic clause
MeSH term
(aspergill OR fungal OR aspergillus) AND (serologic OR serology) AND (platelia OR mannans)
Boolean Query (Original)
Fragmentation
aspergill OR fungal OR aspergillus
serologic OR serology
platelia OR Mannans
Query Fragments (with original MeSH terms)
Remove MeSH terms
aspergill OR fungal
serologic
platelia
Query Fragments (MeSH terms removed)
MeSH Term Suggestion
aspergill OR fungal OR aspergillosis
serologic OR serology
platelia OR mannans
Query Fragments (With suggested MeSH terms)
Defragmentation
(aspergill OR fungal OR aspergillosis) AND (serologic OR serology) AND (platelia OR mannans)
Boolean Query (With suggested MeSH terms)

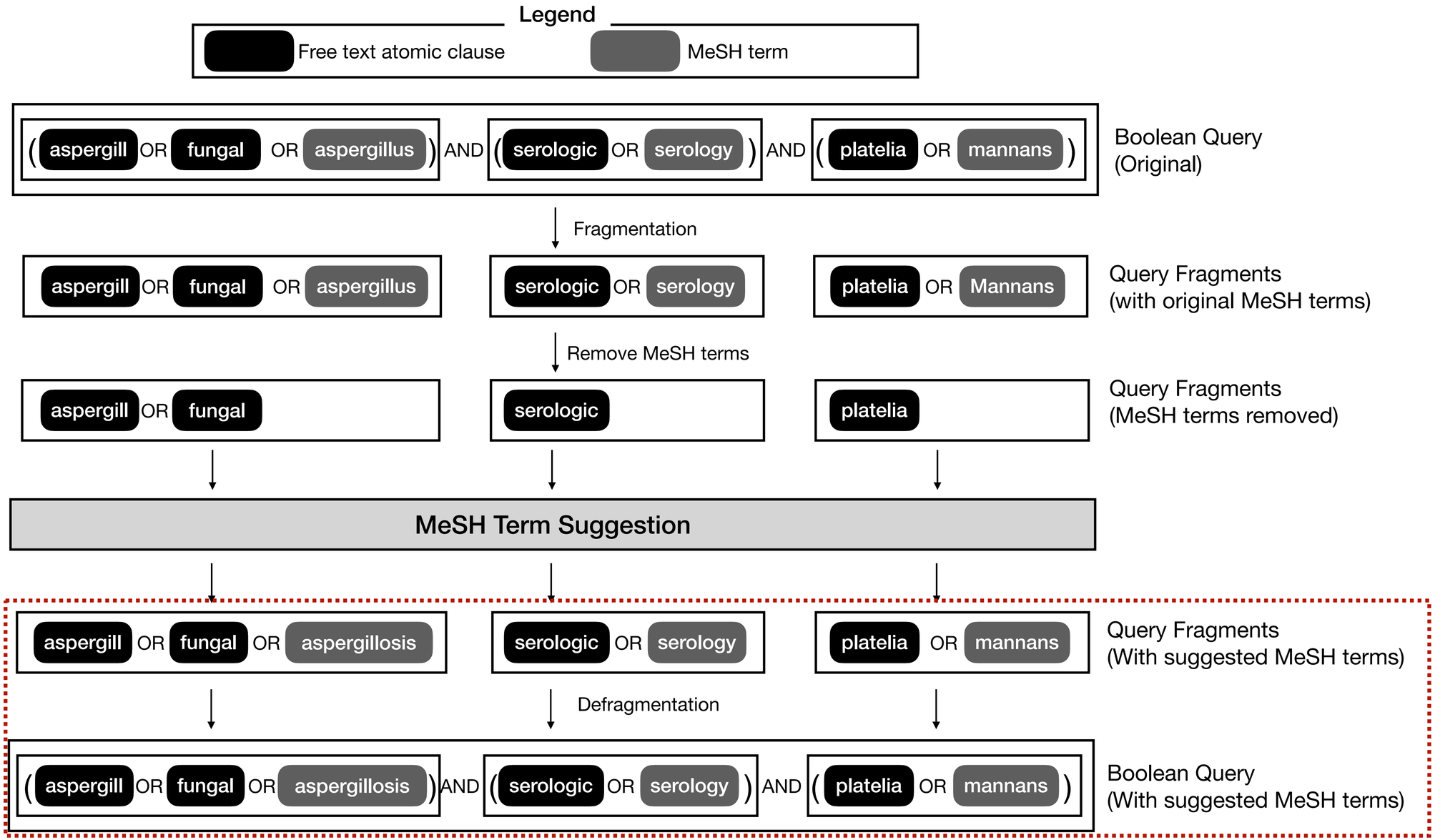
Pipeline


Legend
Free text atomic clause
MeSH term
(aspergill OR fungal OR aspergillus) AND (serologic OR serology) AND (platelia OR mannans)
Boolean Query (Original)
Fragmentation
aspergill OR fungal OR aspergillus
serologic OR serology
platelia OR Mannans
Query Fragments (with original MeSH terms)
Remove MeSH terms
aspergill OR fungal
serologic
platelia
Query Fragments (MeSH terms removed)
MeSH Term Suggestion
aspergill OR fungal OR aspergillosis
serologic OR serology
platelia OR mannans
Query Fragments (With suggested MeSH terms)
Defragmentation
(aspergill OR fungal OR aspergillosis) AND (serologic OR serology) AND (platelia OR mannans)
Boolean Query (With suggested MeSH terms)

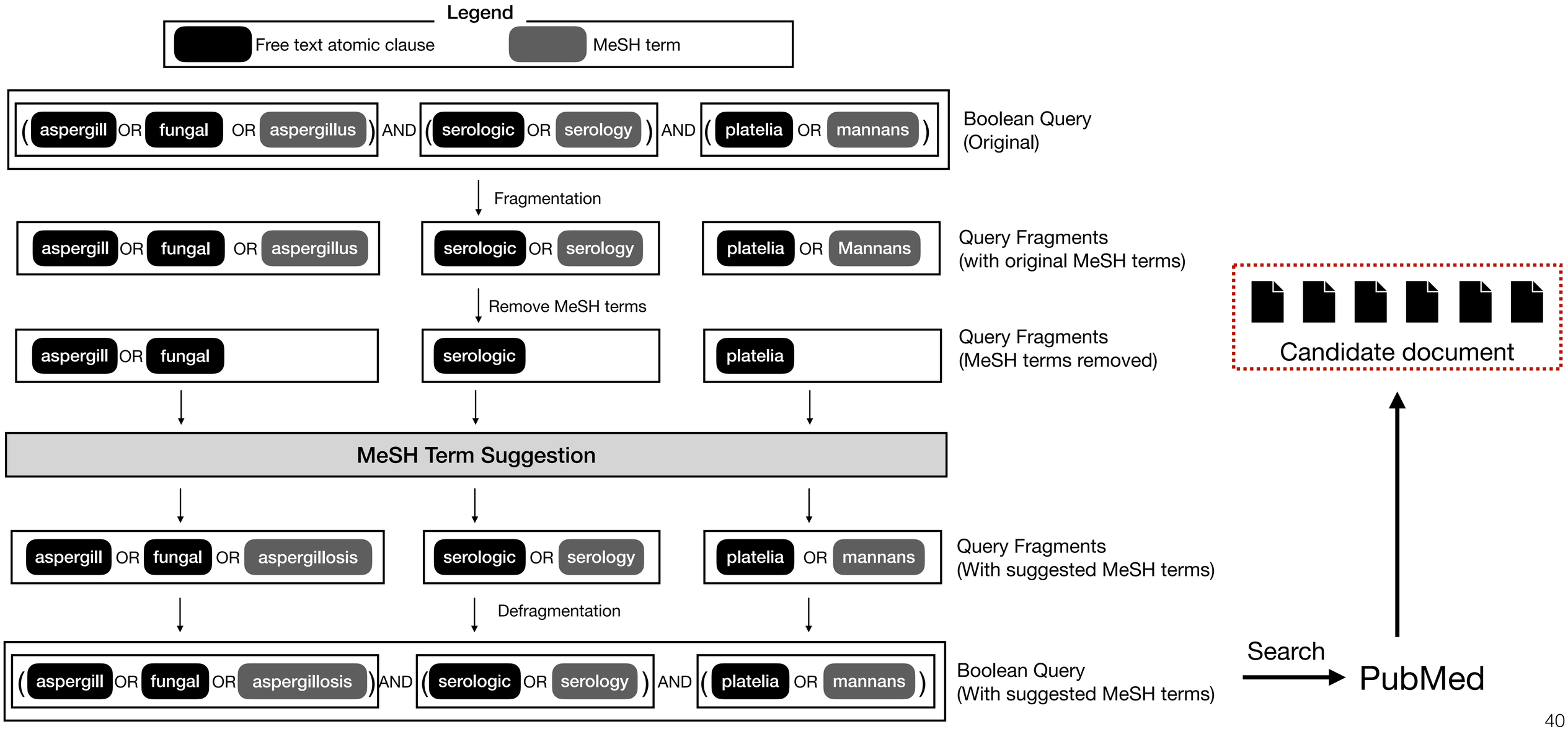
Pipeline


Legend
Free text atomic clause
MeSH term
(aspergill OR fungal OR aspergillus) AND (serologic OR serology) AND (platelia OR mannans)
Boolean Query (Original)
Fragmentation
aspergill OR fungal OR aspergillus
serologic OR serology
platelia OR Mannans
Query Fragments (with original MeSH terms)
Remove MeSH terms
aspergill OR fungal
serologic
platelia
Query Fragments (MeSH terms removed)
MeSH Term Suggestion
aspergill OR fungal OR aspergillosis
serologic OR serology
platelia OR mannans
Query Fragments (With suggested MeSH terms)
Defragmentation
(aspergill OR fungal OR aspergillosis) AND (serologic OR serology) AND (platelia OR mannans)
Boolean Query (With suggested MeSH terms)

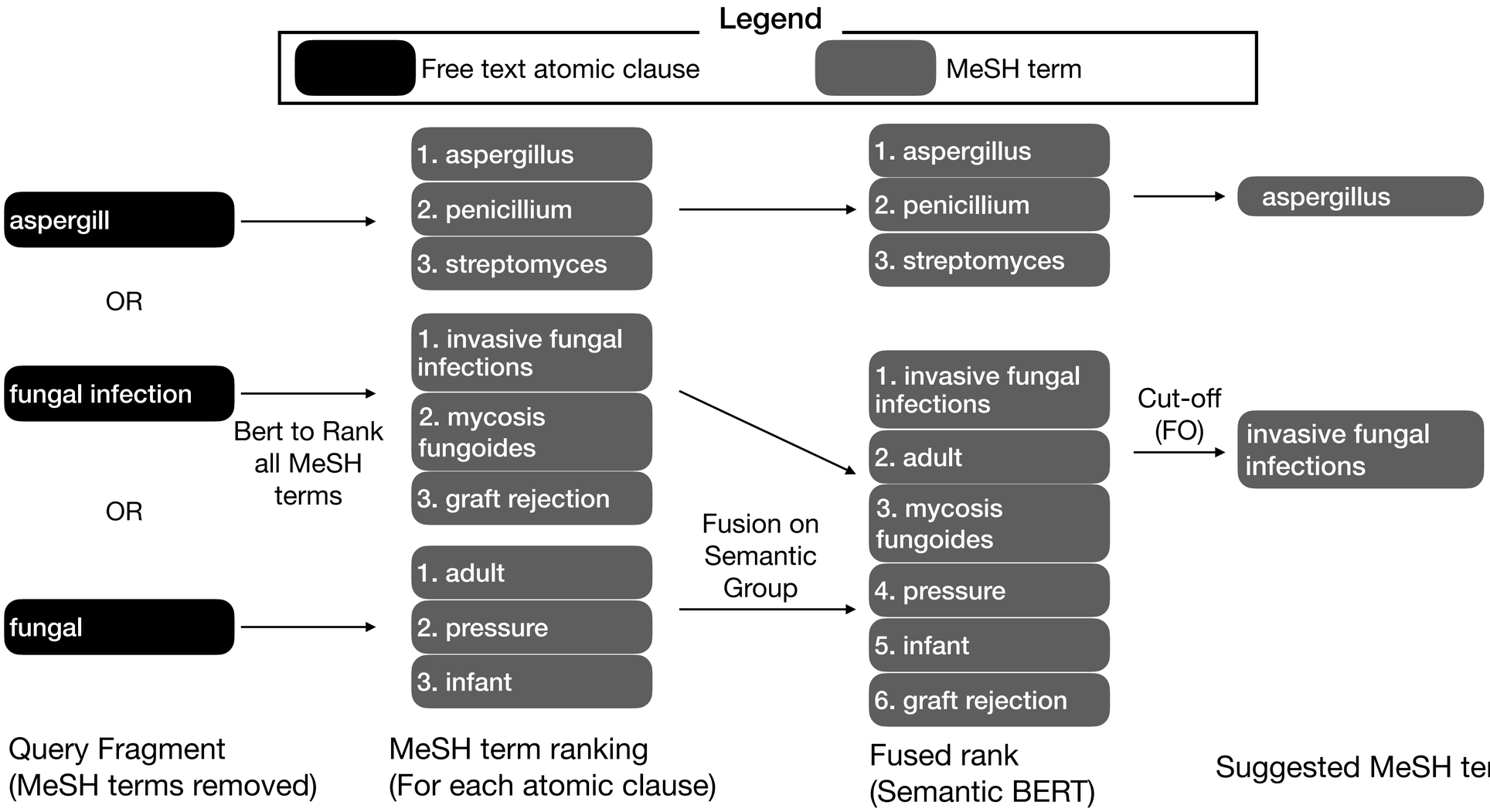
Pipeline


Legend
Free text atomic clause
MeSH term
(aspergill OR fungal OR aspergillus) AND (serologic OR serology) AND (platelia OR mannans)
Boolean Query (Original)
Fragmentation
aspergill OR fungal OR aspergillus
serologic OR serology
platelia OR Mannans
Query Fragments (with original MeSH terms)
Remove MeSH terms
aspergill OR fungal
serologic
platelia
Query Fragments (MeSH terms removed)
MeSH Term Suggestion
aspergill OR fungal OR aspergillosis
serologic OR serology
platelia OR mannans
Query Fragments (With suggested MeSH terms)
Defragmentation
(aspergill OR fungal OR aspergillosis) AND (serologic OR serology) AND (platelia OR mannans)
Boolean Query (With suggested MeSH terms)

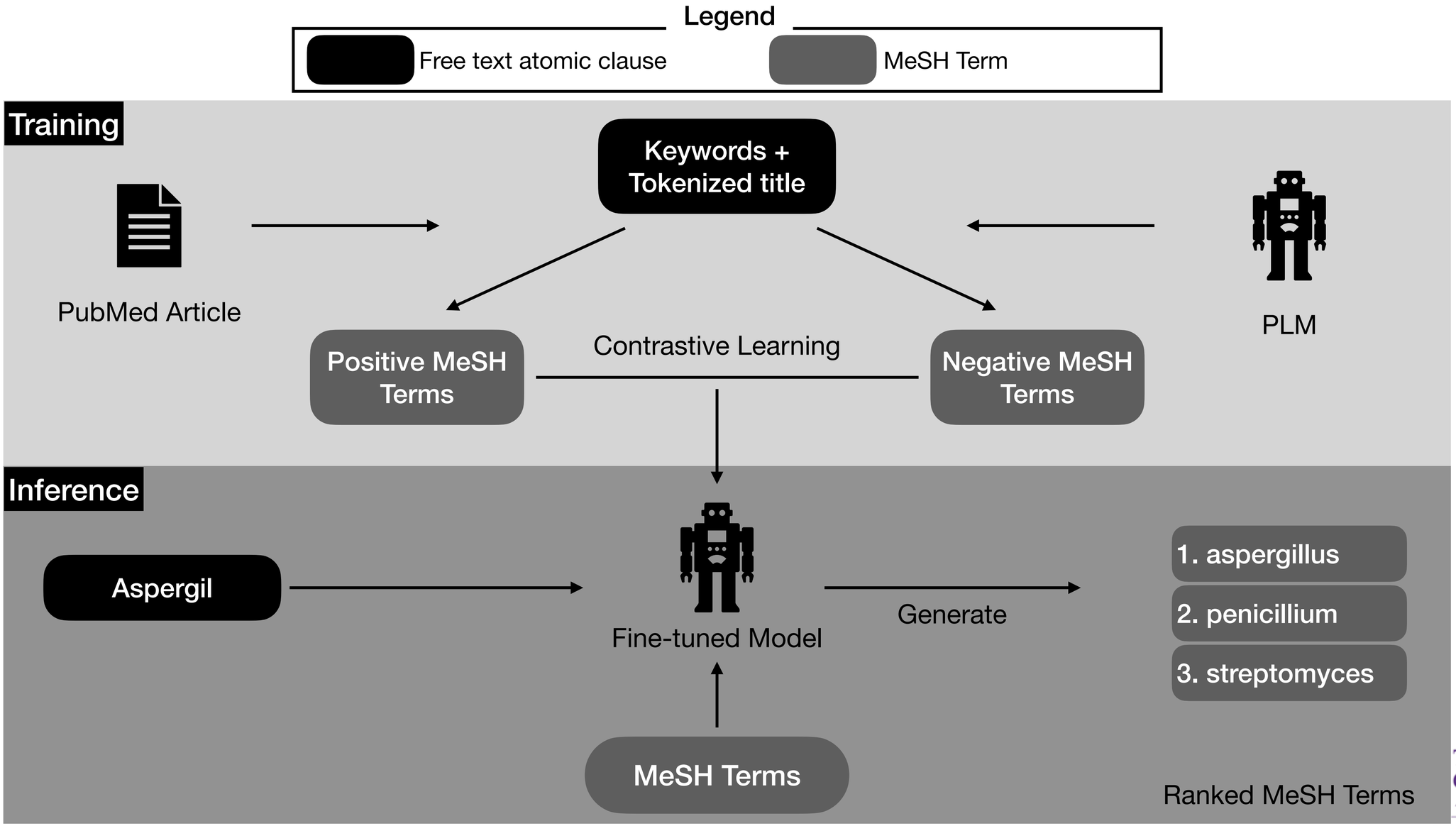
Evaluate


Legend
Free text atomic clause
MeSH term
(aspergill OR fungal OR aspergillus) AND (serologic OR serology) AND (platelia OR mannans)
Boolean Query (Original)
Fragmentation
aspergill OR fungal OR aspergillus
serologic OR serology
platelia OR Mannans
Query Fragments (with original MeSH terms)
Remove MeSH terms
aspergill OR fungal
serologic
platelia
Query Fragments (MeSH terms removed)
MeSH Term Suggestion
aspergill OR fungal OR aspergillosis
serologic OR serology
platelia OR mannans
Query Fragments (With suggested MeSH terms)
Defragmentation
(aspergill OR fungal OR aspergillosis) AND (serologic OR serology) AND (platelia OR mannans)
Boolean Query (With suggested MeSH terms)
Candidate document
Search
PubMed

BERT methods (MTS)



BERT methods train (MTS)



Improving Causal Relation Extraction From The Web

Presentation by Charly Zimmer
Scientific Assistant at Leipzig University

Causal Relation Extraction

Causal Relation Extraction

Earth quakes cause waves.

Causal Relation Extraction

Earth quakes cause waves.

cause entity

effect entity

Causal Relation Extraction

Earth quakes cause waves.

cause entity causal pattern effect entity

Causal Relation Extraction

Earth quakes cause waves.

cause entity causal pattern effect entity

Examples lead to a better understanding.

Causal Relation Extraction

Earth quakes cause waves.

cause entity — causal pattern — effect entity

Examples lead to a better understanding.

cause entity — effect entity

Causal Relation Extraction

Earth quakes cause waves.

cause entity — causal pattern — effect entity

Examples lead to a better understanding.

cause entity — causal pattern — effect entity

CauseNet

CauseNet

- paper published in 2020

CauseNet

- paper published in 2020

- authors:

  Stefan Heindorf
  Yan Scholten
  Henning Wachsmuth
  Axel-Cyrille Ngonga Ngomo
  Martin Potthast

CauseNet

- paper published in 2020

- authors:

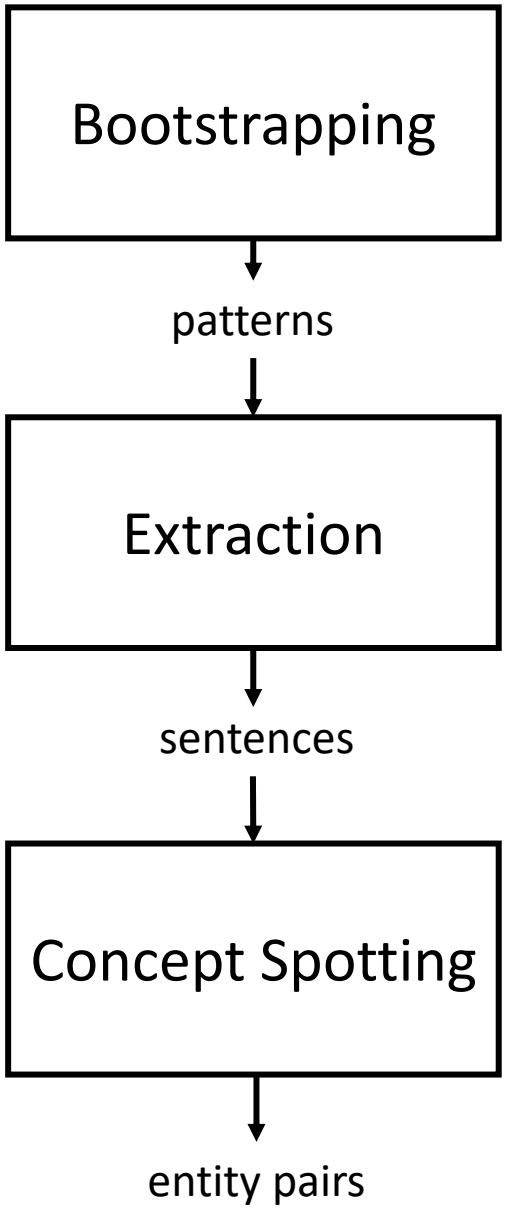
  Stefan Heindorf

  Yan Scholten

  Henning Wachsmuth

  Axel-Cyrille Ngonga Ngomo

  Martin Potthast



Improvement

Improvement

	# of causal relations	**precision**	**# of patterns**
CauseNet	11.6 M	84.5%	52
My thesis	~45.4 M	87.7%	57

Improvement

	# of causal relations	precision	# of patterns
CauseNet	11.6 M	84.5%	52
My thesis	~45.4 M	87.7%	57

- changed from ClueWeb12 to Common Crawl
- modified bootstrapping
- manual evaluation of patterns

Current Work

Current Work

- new Causal Concept Spotter

Current Work

- new Causal Concept Spotter
- transformer-based model

Dataset	# sentences
SemEval 2007 Train	73
SemEval 2007 Test	80
SemEval 2010 Train	1003
SemEval 2010 Test	328
TRex	327*
Paderborn	583*
Overall	2394

*… multiple relations per sentence

Thank you for your attention!