

Improving Query Formulation for Systematic Review Literature Search

Guido Zuccon

g.zuccon@uq.edu.au

ielab, The University of Queensland, Australia

www.ielab.io

In this talk...

- **Query formulation** key step for systematic review creation
- Large scope for **automation and improvements** in query formulation in this context
- **Promising techniques** have emerged, but still **plenty of challenges**
- Exciting area of research
 - Neural models, especially generative
 - Prediction of performance
 - User understanding and interaction aspects

Systematic Reviews

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

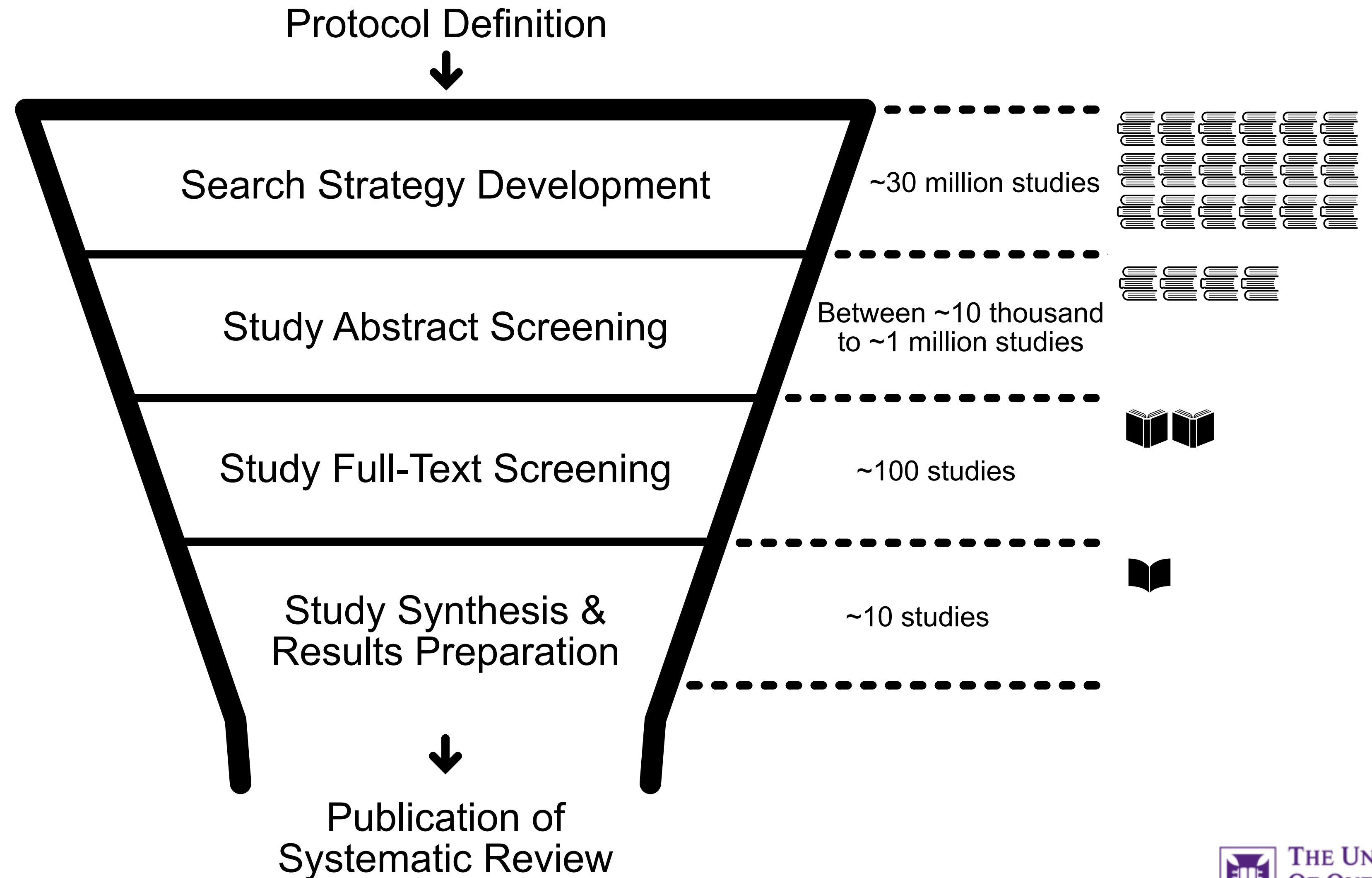


- Often adheres to strict guidelines and protocol, to ensure the review could be reproduced
- Cornerstone of evidence based medicine, guide clinical decisions
 - e.g. what actions clinicians should take to treat patients
- Inform institutional practice and policy
 - e.g. Banning smoking in public areas

Systematic Reviews are Costly and Take Long Time to Complete

- Increasing **number of publications** per year
 - Harder to find literature, **more studies to appraise**
- Reviews become **out of date** at time of publication (!)

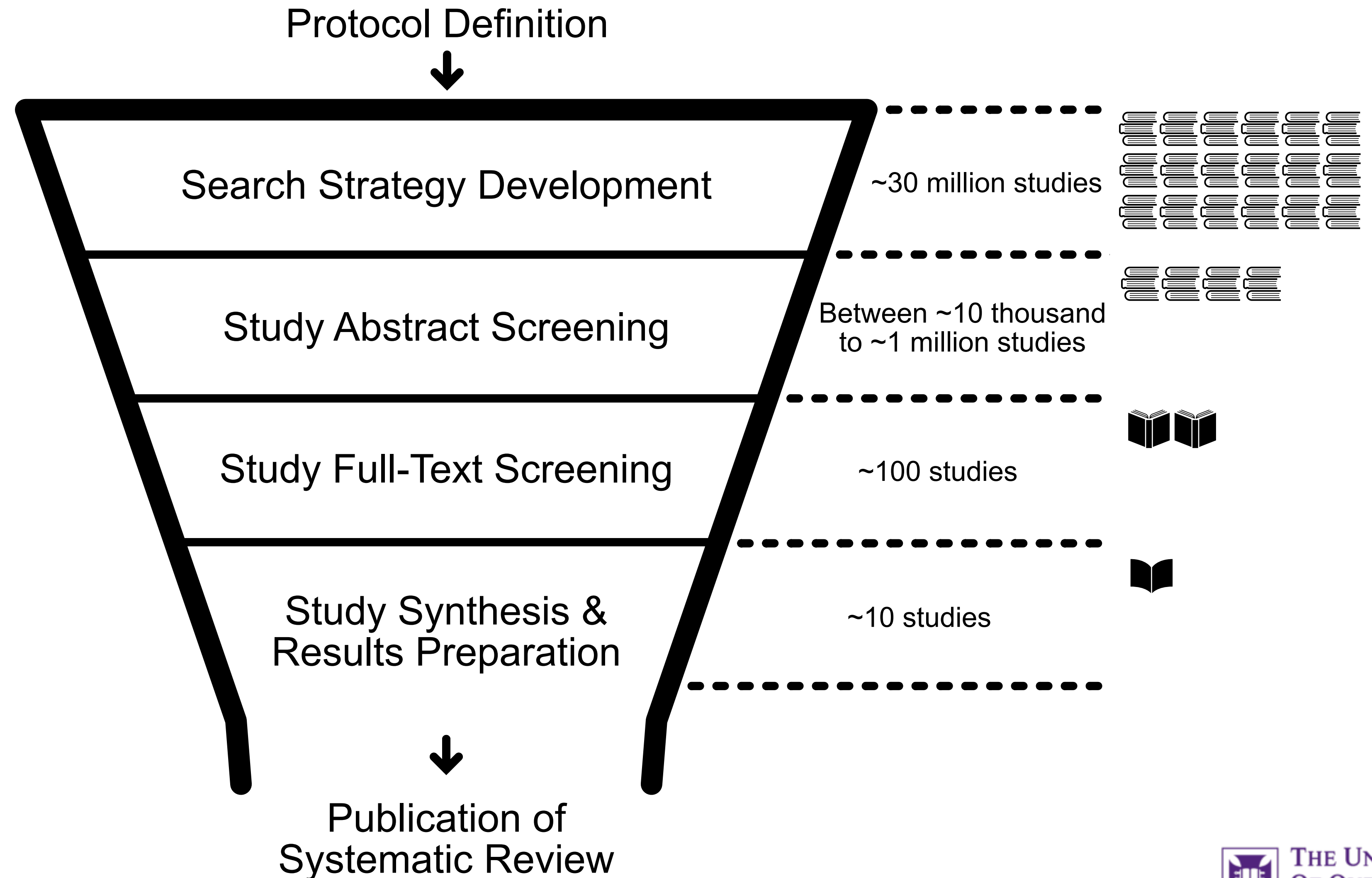
Phases of creation



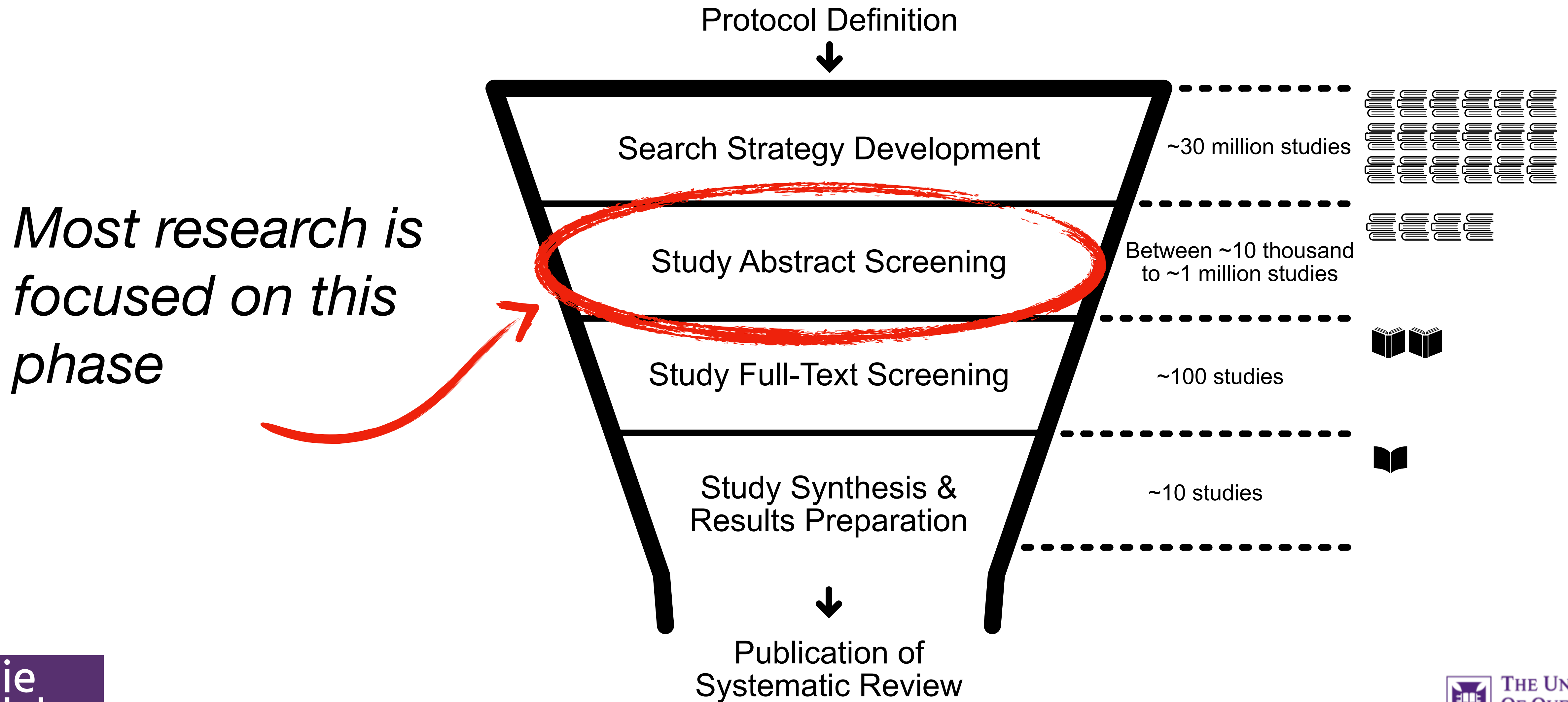
Phases of creation

Increasing amount of research aiming to automate part of the systematic review creation process:

- Reduce costs
- Speed up completion



Previous Research



Automation of Abstract Screening

- Train **classifiers** for a specific review; often using **Active Learning**:
 - Wallace, Small, Brodley, Trikalinos, 2010. *Active learning for biomedical citation screening*. SIGKDD.
 - Miwa, Thomas, O'Mara-Eves, Ananiadou, 2014. *Reducing systematic review workload through certainty-based screening*. Journal of biomedical informatics.
 - Molinari, Kanoulas, 2022. *Transferring knowledge between topics in systematic reviews*. Intelligent Systems with Applications

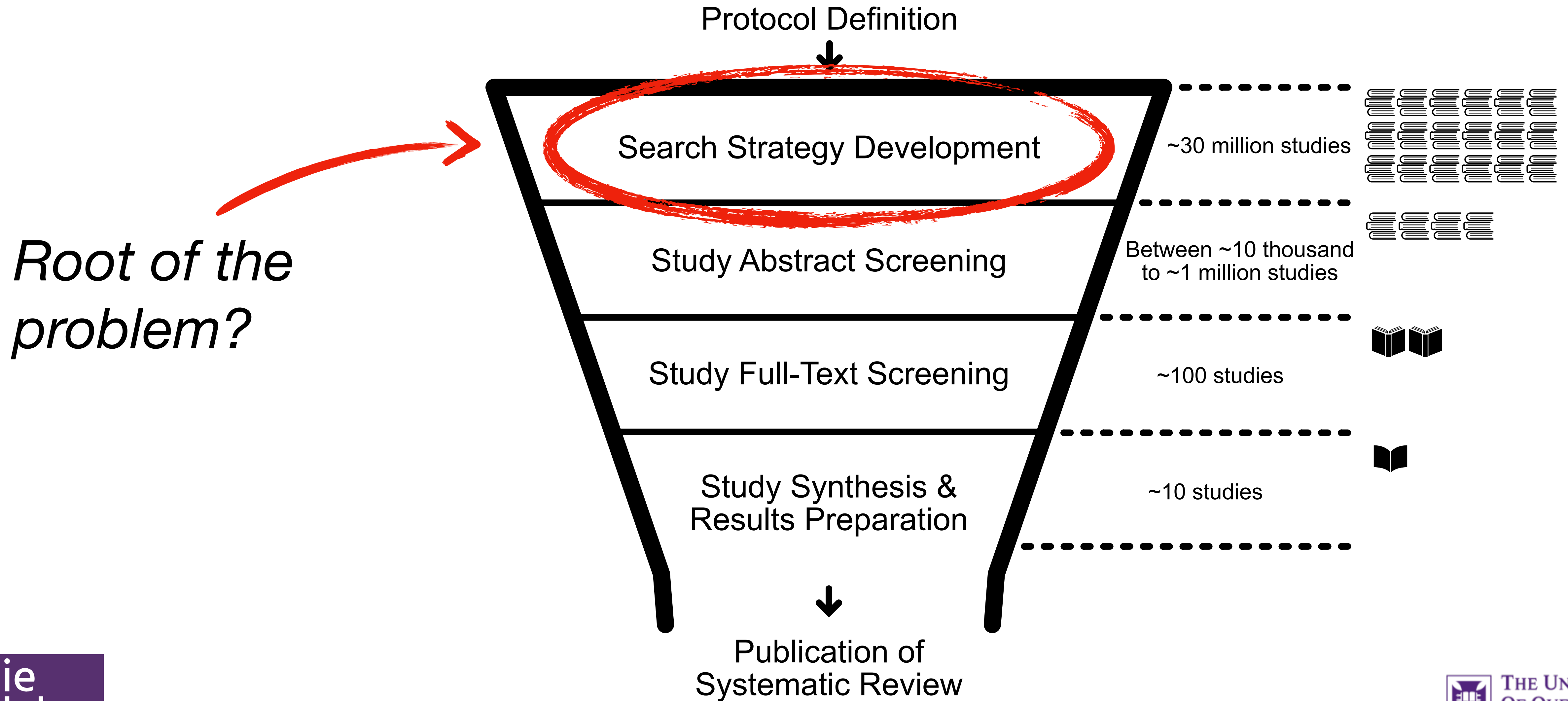
Automation of Abstract Screening

- Train classifiers for a specific review; often using Active Learning:
- **Screening prioritisation**
 - Lee, Sun, 2018. *Seed-driven document ranking for systematic reviews in evidence-based medicine*. SIGIR
 - Alharbi, Stevenson, 2019. *Improving ranking for systematic reviews using query adaptation*. CLEF
 - Scells, Zuccon, Koopman, 2020. *You can teach an old dog new tricks: Rank fusion applied to coordination level matching for ranking in systematic reviews*. ECIR
 - Wang, Scells, Koopman, Zuccon, 2022. *Neural Rankers for Effective Screening Prioritisation in Medical Systematic Review Literature Search*. ADCS

Automation of Abstract Screening

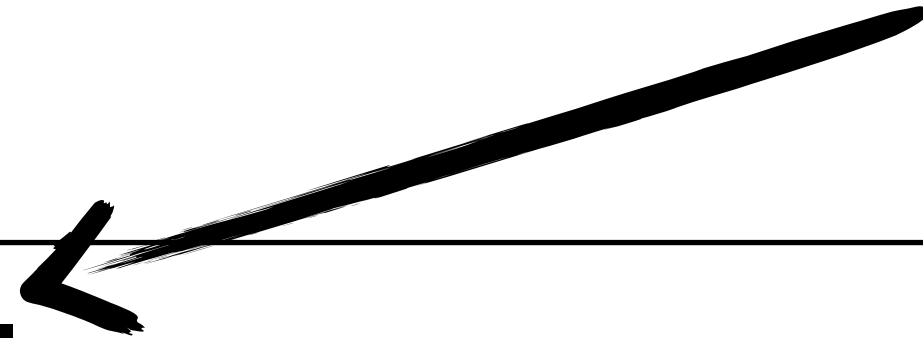
- Train classifiers for a specific review; often using Active Learning:
- Screening prioritisation
- Identify when to **stop screening**
 - Di Nunzio, 2018. *A study of an automatic stopping strategy for technologically assisted medical reviews*. ECIR
 - Li, Kanoulas, 2020. *When to stop reviewing in technology-assisted reviews: Sampling from an adaptive distribution to estimate residual relevant documents*. TOIS
 - Callaghan, Müller-Hansen, 2020. *Statistical stopping criteria for automated screening in systematic reviews*. Systematic Reviews.
 - Sneyd, Stevenson, 2021. *Stopping criteria for technology assisted reviews based on counting processes*. SIGIR
 - Yang, Lewis, Frieder, 2021. *Heuristic stopping rules for technology-assisted review*. DocEng

Search Strategies



1. (adrenergic* and antagonist*).tw.
2. (adrenergic* and block\$).tw.
3. (adrenergic* and beta-receptor*).tw.
4. (beta-adrenergic* and block*).tw.
5. (beta-blocker* and adrenergic*).tw.
6. (blockader*.tw. or Propranolol/ or Sotalol/)
7. or/1-6
8. Lung Diseases, Obstructive/
9. exp Pulmonary Disease, Chronic Obstructive/
10. emphysema*.tw.
11. (chronic* adj3 bronchiti*).tw.
12. (obstruct*.tw. adj3 (lung* or airway*).tw.)
13. COPD.tw.
14. COAD.tw.
15. COBD.tw.
16. AECEB.tw.
17. or/8-16
18. 7 and 17

Field Restrictions

- 
1. (adrenergic* and antagonist*).tw.
 2. (adrenergic* and block\$).tw.
 3. (adrenergic* and beta-receptor*).tw.
 4. (beta-adrenergic* and block*).tw.
 5. (beta-blocker* and adrenergic*).tw.
 6. (blockader*.tw. or Propranolol/ or Sotalol/)
 7. or/1-6
 8. Lung Diseases, Obstructive/
 9. exp Pulmonary Disease, Chronic Obstructive/
 10. emphysema*.tw.
 11. (chronic* adj3 bronchiti*).tw.
 12. (obstruct*.tw. adj3 (lung* or airway*).tw.)
 13. COPD.tw.
 14. COAD.tw.
 15. COBD.tw.
 16. AECEB.tw.
 17. or/8-16
 18. 7 and 17

Wildcard

Explicit Stemming

Field Restrictions

1. (adrenergic* and antagonist*).tw.
2. (adrenergic* and block\$).tw.
3. (adrenergic* and beta-receptor*).tw.
4. (beta-adrenergic* and block*).tw.
5. (beta-blocker* and adrenergic*).tw.
6. (blockader*.tw. or Propranolol/ or Sotalol/)
7. or/1-6
8. Lung Diseases, Obstructive/
9. exp Pulmonary Disease, Chronic Obstructive/
10. emphysema*.tw.
11. (chronic* adj3 bronchiti*).tw.
12. (obstruct*.tw. adj3 (lung* or airway*).tw.)
13. COPD.tw.
14. COAD.tw.
15. COBD.tw.
16. AECEB.tw.
17. or/8-16
18. 7 and 17

Wildcard

Explicit Stemming

Field Restrictions

1. (adrenergic* and antagonist*).tw.
2. (adrenergic* and block\$).tw.
3. (adrenergic* and beta-receptor*).tw.
4. (beta-adrenergic* and block*).tw.
5. (beta-blocker* and adrenergic*).tw.
6. (blockader*.tw. or Propranolol/ or Sotalol/)
7. or/1-6
8. Lung Diseases, Obstructive/
9. exp Pulmonary Disease, Chronic Obstructive/
10. emphysema*.tw.
11. (chronic* adj3 bronchiti*).tw.
12. (obstruct*.tw. adj3 (lung* or airway*).tw.)
13. COPD.tw.
14. COAD.tw.
15. COBD.tw.
16. AECB.tw.
17. or/8-16
18. 7 and 17

MeSH Heading

MeSH "Explosion"

Boolean Queries are a Difficult Beast

- Query formulation is hard
- Information experts (i.e. librarians) often assist [1]
- Difficult to determine effectiveness before appraisal
- Search is often largely unstructured and not methodical, it sometimes undergoes peer review

[1] Rethlefsen, M.L., Farrell, A.M., Trzasko, L.C.O. and Brigham, T.J., 2015. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. Journal of clinical epidemiology.

Boolean Queries are a Difficult Beast

- Query formulation is critical
 - Quality of sys. rev. ultimately decided by query
- Query takes significant resources (cost, time)
- Information experts = \$\$\$ [1]

[1] McGowan, J. and Sampson, M., 2005. Systematic reviews need systematic searchers. Journal of the Medical Library Association

Boolean Queries are a Difficult Beast

- What if we have better queries to begin with?
- What if we assist researchers formulate better queries?
- What if we automate the formulation of better queries?

Query Automation

Query Automation

Query Formulation

Automatic Conceptual Method

Scells,Zuccon,Koopman,Clark WWW 2020

Scells,Zuccon,Koopman IRJ 2020

Automatic Objective Method

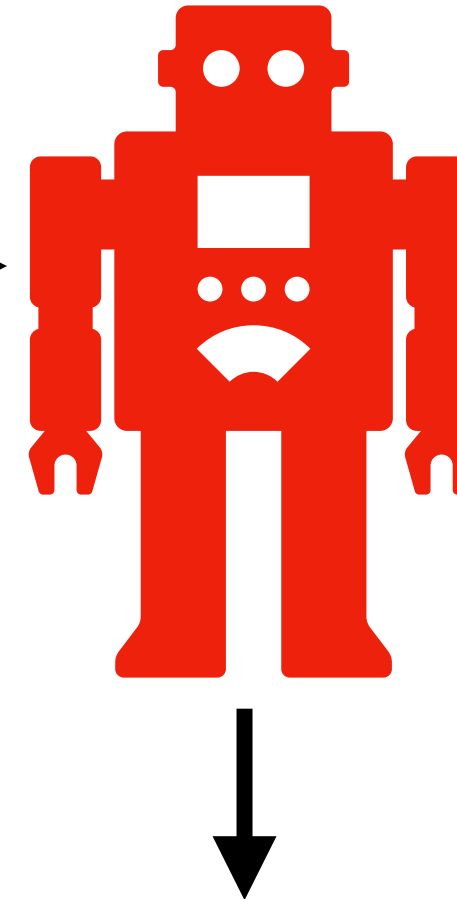
Scells,Zuccon,Koopman,Clark ECIR 2020

Formulating Boolean Queries with ChatGPT

Wang,Scells,Koopman,Zuccon 2023 arxiv

Query Automation

Specifications/Protocol
of systematic review →



1. (adrenergic* and antagonist*).tw.
2. (adrenergic* and block\$).tw.
- ...
18. 7 and 17

Query Formulation

Automatic Conceptual Method

Scells,Zuccon,Koopman,Clark WWW 2020

Scells,Zuccon,Koopman IRJ 2020

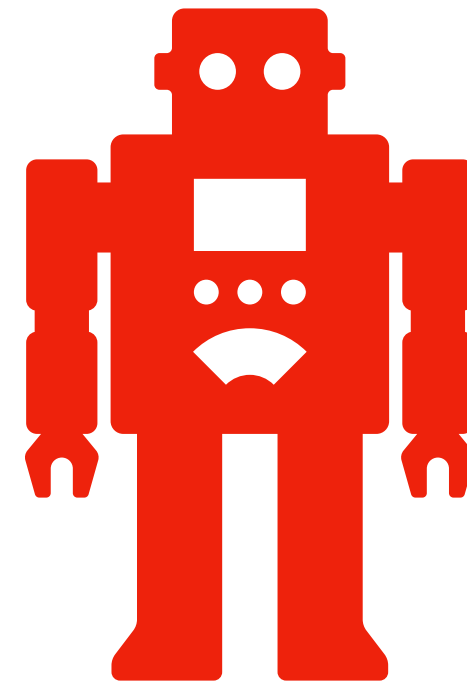
Automatic Objective Method

Scells,Zuccon,Koopman,Clark ECIR 2020

Formulating Boolean Queries with ChatGPT

Wang,Scells,Koopman,Zuccon 2023 arxiv

Query Automation



1. (adrenergic* and antagonist*).tw.

2. (adrenergic* and block\$).tw.

...

18. 7 and 17

1. (adrenergic\$ and antagonist*).tw.

2. (adrenergic\$ and block\$).tw.abs.

...

21. 7 and 17 and 20

Query Refinement

Query Transformations and Query Chains

Scells,Zuccon SIGIR 2018

Scells,Zuccon,Koopman WWW 2019

Scells,Zuccon,Sharaf,Koopman WWW 2020

Log Analysis of Refinements

Scells,Forbes,Clark,Koopman,Zuccon,
ICTIR 2022

MeSH Terms

Wang,Scells,Koopman,Zuccon,
Int.Sys. Appl. 2022

Wang,Li,Scells,Locke,Zuccon, ADCS 2021

Wang,Li,Zuccon, WSDM 2023

Query Formulation

Automatic Conceptual Method

Scells,Zucon,Koopman,Clark WWW 2020

Scells,Zucon,Koopman IRJ 2020

Automatic Objective Method

Scells,Zucon,Koopman,Clark ECIR 2020

Formulating Boolean Queries with ChatGPT

Wang,Scells,Koopman,Zucon 2023 arxiv

Query Automation

Query 1

1. (adrenergic* and antagonist*).tw.

2. (adrenergic* and block\$).tw.

...

18. 7 and 17

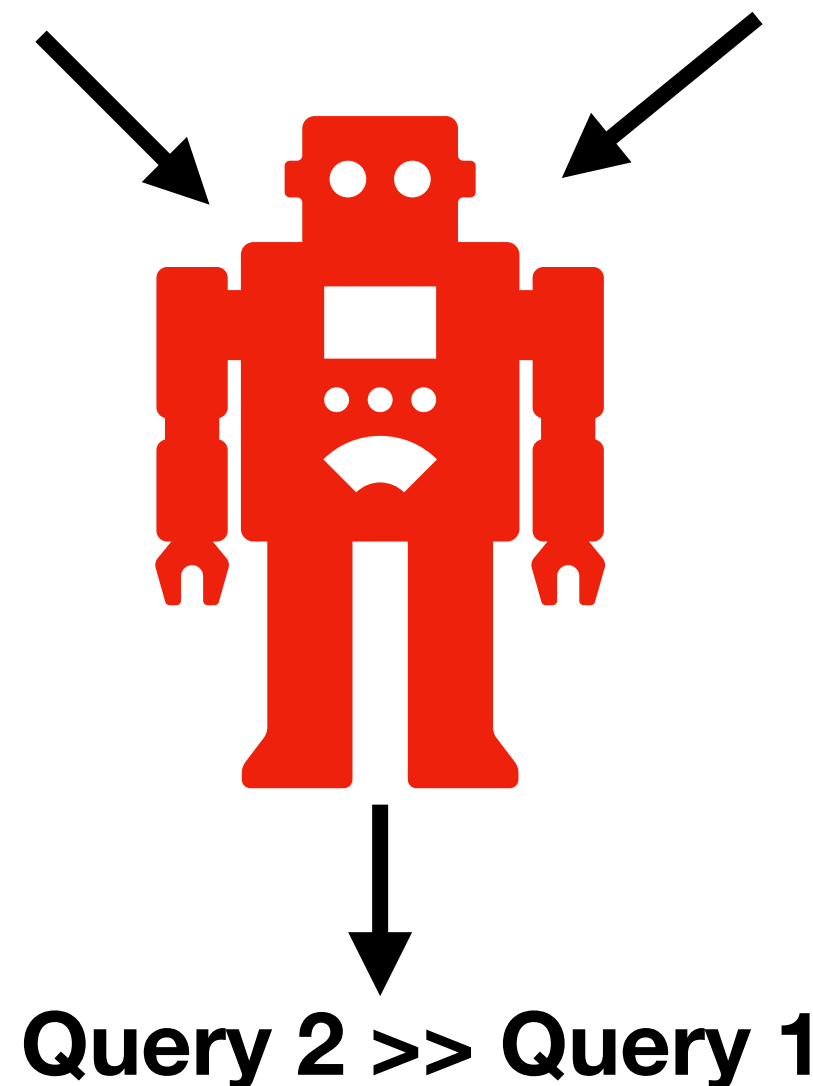
Query 2

1. (adrenergic\$ and antagonist*).tw.

2. (adrenergic\$ and block\$).tw.abs.

...

21. 7 and 17 and 20



Query Refinement

Query Transformations and Query Chains

Scells,Zucon SIGIR 2018

Scells,Zucon,Koopman WWW 2019

Scells,Zucon,Sharaf,Koopman WWW 2020

Log Analysis of Refinements

Scells,Forbes,Clark,Koopman,Zucon, ICTIR 2022

MeSH Terms

Wang,Scells,Koopman,Zucon, Int.Sys. Appl. 2022

Wang,Li,Scells,Locke,Zucon, ADCS 2021

Wang,Li,Zucon, WSDM 2023

QPP

Scells,Azzopardi,Zucon,Koopman SIGIR 2018

Query Formulation

Automatic Conceptual Method

Scells,Zucon,Koopman,Clark WWW 2020

Scells,Zucon,Koopman IRJ 2020

Automatic Objective Method

Scells,Zucon,Koopman,Clark ECIR 2020

Formulating Boolean Queries with ChatGPT

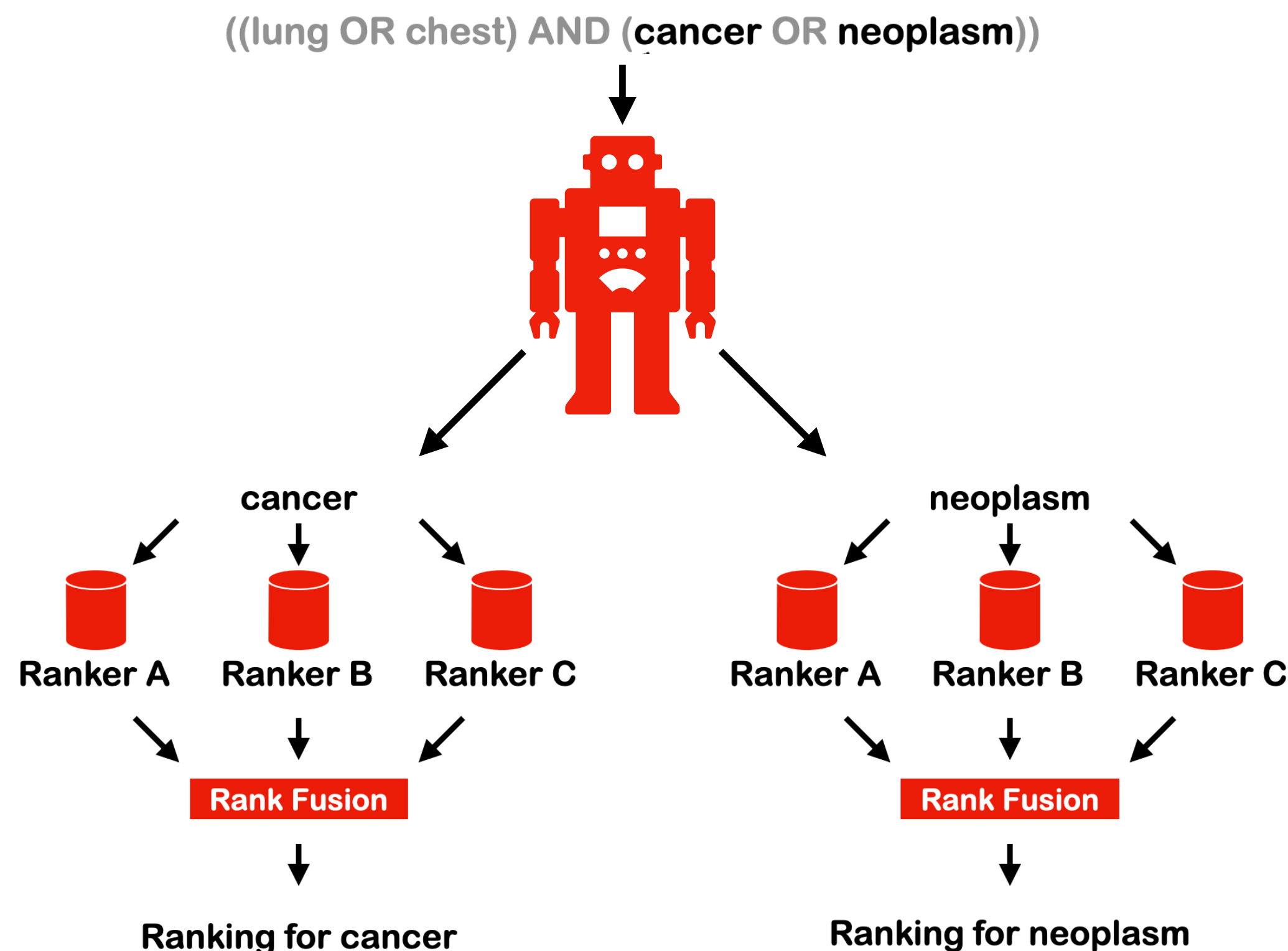
Wang,Scells,Koopman,Zucon 2023 arxiv

Query Exploitation

Scells,Zucon,Koopman,
Deacon,Azzopardi,Geva CIKM 2017

Scells,Zucon,Koopman ECIR 2020

Query Automation



Query Refinement

Query Transformations and Query Chains

Scells,Zucon SIGIR 2018

Scells,Zucon,Koopman WWW 2019

Scells,Zucon,Sharaf,Koopman WWW 2020

Log Analysis of Refinements

Scells,Forbes,Clark,Koopman,Zucon,
ICTIR 2022

MeSH Terms

Wang,Scells,Koopman,Zucon,
Int.Sys. Appl. 2022

Wang,Li,Scells,Locke,Zucon, ADCS 2021

Wang,Li,Zucon, WSDM 2023

QPP

Scells,Azzopardi,Zucon,Koopman
SIGIR 2018

Query Formulation

Automatic Conceptual Method

Scells,Zuccon,Koopman,Clark WWW 2020

Scells,Zuccon,Koopman IRJ 2020

Automatic Objective Method

Scells,Zuccon,Koopman,Clark ECIR 2020

Formulating Boolean Queries with ChatGPT

Wang,Scells,Koopman,Zuccon 2023 arxiv

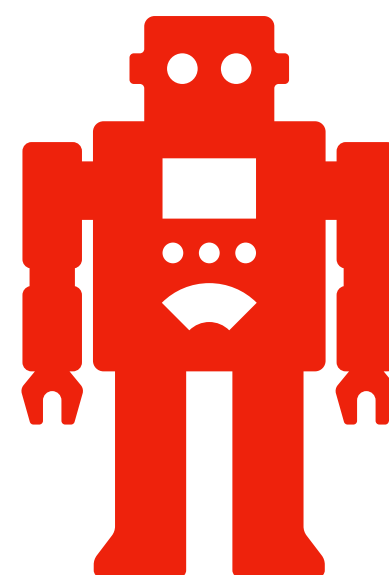
Query Exploitation

Scells,Zuccon,Koopman,
Deacon,Azzopardi,Geva CIKM 2017

Scells,Zuccon,Koopman ECIR 2020

Query Automation

Seeds



1. (adrenergic* and antagonist*).tw.
2. (adrenergic* and block\$).tw.
- ...
18. 7 and 17

Seed Studies

Scells,Zuccon,Koopman,Clark ECIR 2020

Scells,Zuccon,Koopman IRJ 2020

Wang,Scells,Clark,Koopman,Zuccon,
SIGIR 2022

Wang,Scells,Koopman,Zuccon 2023 arxiv

Query Refinement

Query Transformations and Query Chains

Scells.Zuccon SIGIR 2018

Scells,Zuccon,Koopman WWW 2019

Scells,Zuccon,Sharaf,Koopman WWW 2020

Log Analysis of Refinements

Scells,Forbes,Clark,Koopman,Zuccon,
ICTIR 2022

MeSH Terms

Wang,Scells,Koopman,Zuccon,
Int.Sys. Appl. 2022

Wang,Li,Scells,Locke,Zuccon, ADCS 2021

Wang,Li,Zuccon, WSDM 2023

QPP

Scells,Azzopardi,Zuccon,Koopman
SIGIR 2018

Query Formulation

Automatic Conceptual Method

Scells,Zuccon,Koopman,Clark WWW 2020

Scells,Zuccon,Koopman IRJ 2020

Automatic Objective Method

Scells,Zuccon,Koopman,Clark ECIR 2020

Formulating Boolean Queries with ChatGPT

Wang,Scells,Koopman,Zuccon 2023 arxiv

Query Exploitation

Scells,Zuccon,Koopman,
Deacon,Azzopardi,Geva CIKM 2017

Scells,Zuccon,Koopman ECIR 2020

Query Automation

Systems & Integrations

Scells,Zuccon CIKM 2018

Scells,Zuccon SIGIR 2018

Li,Scells,Zuccon, SIGIR 2020

Wang,Li,Zuccon, WSDM 2023

Seed Studies

Scells,Zuccon,Koopman,Clark ECIR 2020

Scells,Zuccon,Koopman IRJ 2020

Wang,Scells,Clark,Koopman,Zuccon,
SIGIR 2022

Wang,Scells,Koopman,Zuccon 2023 arxiv

Query Refinement

Query Transformations and Query Chains

Scells.Zuccon SIGIR 2018

Scells,Zuccon,Koopman WWW 2019

Scells,Zuccon,Sharaf,Koopman WWW 2020

Log Analysis of Refinements

Scells,Forbes,Clark,Koopman,Zuccon,
ICTIR 2022

MeSH Terms

Wang,Scells,Koopman,Zuccon,
Int.Sys. Appl. 2022

Wang,Li,Scells,Locke,Zuccon, ADCS 2021

Wang,Li,Zuccon, WSDM 2023

QPP

Scells,Azzopardi,Zuccon,Koopman
SIGIR 2018

Query Formulation

Automatic Conceptual Method

Scells,Zuccon,Koopman,Clark WWW 2020

Scells,Zuccon,Koopman IRJ 2020

Automatic Objective Method

Scells,Zuccon,Koopman,Clark ECIR 2020

Formulating Boolean Queries with ChatGPT

Wang,Scells,Koopman,Zuccon 2023 arxiv

Query Exploitation

Scells,Zuccon,Koopman,
Deacon,Azzopardi,Geva CIKM 2017

Scells,Zuccon,Koopman ECIR 2020

Query Automation

Systems & Integrations

Scells,Zuccon CIKM 2018

Scells,Zuccon SIGIR 2018

Li,Scells,Zuccon, SIGIR 2020

Wang,Li,Zuccon, WSDM 2023

Seed Studies

Scells,Zuccon,Koopman,Clark ECIR 2020

Scells,Zuccon,Koopman IRJ 2020

Wang,Scells,Clark,Koopman,Zuccon,
SIGIR 2022

Wang,Scells,Koopman,Zuccon 2023 arxiv

Query Refinement

Query Transformations and Query Chains

Scells.Zuccon SIGIR 2018

Scells,Zuccon,Koopman WWW 2019

Scells,Zuccon,Sharaf,Koopman WWW 2020

Log Analysis of Refinements

Scells,Forbes,Clark,Koopman,Zuccon,
ICTIR 2022

MeSH Terms

Wang,Scells,Koopman,Zuccon,
Int.Sys. Appl. 2022

Wang,Li,Scells,Locke,Zuccon, ADCS 2021

Wang,Li,Zuccon, WSDM 2023

QPP

Scells,Azzopardi,Zuccon,Koopman
SIGIR 2018

Query Formulation

Automatic Conceptual Method

Scells,Zuccon,Koopman,Clark WWW 2020

Scells,Zuccon,Koopman IRJ 2020

Automatic Objective Method

Scells,Zuccon,Koopman,Clark ECIR 2020

Formulating Boolean Queries with ChatGPT

Wang,Scells,Koopman,Zuccon 2023 arxiv

Query Exploitation

Scells,Zuccon,Koopman,
Deacon,Azzopardi,Geva CIKM 2017

Scells,Zuccon,Koopman ECIR 2020

Query Automation



Harry Scells, now at Universität Leipzig



Systems & Integrations

Scells,Zuccon CIKM 2018

Scells,Zuccon SIGIR 2018

Li,Scells,Zuccon, SIGIR 2020

Wang,Li,Zuccon, WSDM 2023

Seed Studies

Scells,Zuccon,Koopman,Clark ECIR 2020

Scells,Zuccon,Koopman IRJ 2020

Wang,Scells,Clark,Koopman,Zuccon,
SIGIR 2022

Wang,Scells,Koopman,Zuccon 2023 arxiv

Query Refinement

Query Transformations and Query Chains

Scells.Zuccon SIGIR 2018

Scells,Zuccon,Koopman WWW 2019

Scells,Zuccon,Sharaf,Koopman WWW 2020

Log Analysis of Refinements

Scells,Forbes,Clark,Koopman,Zuccon,
ICTIR 2022

MeSH Terms

Wang,Scells,Koopman,Zuccon,
Int.Sys. Appl. 2022

Wang,Li,Scells,Locke,Zuccon, ADCS 2021

Wang,Li,Zuccon, WSDM 2023

QPP

Scells,Azzopardi,Zuccon,Koopman
SIGIR 2018

Query Formulation

Automatic Conceptual Method

Scells,Zuccon,Koopman,Clark WWW 2020

Scells,Zuccon,Koopman IRJ 2020

Automatic Objective Method

Scells,Zuccon,Koopman,Clark ECIR 2020

Formulating Boolean Queries with ChatGPT

Wang,Scells,Koopman,Zuccon 2023 arxiv

Query Exploitation

Scells,Zuccon,Koopman,
Deacon,Azzopardi,Geva CIKM 2017

Scells,Zuccon,Koopman ECIR 2020

Query Automation



Shuai Wang
PhD @ UQ

Systems & Integrations

Scells,Zuccon CIKM 2018

Scells,Zuccon SIGIR 2018

Li,Scells,Zuccon, SIGIR 2020

Wang,Li,Zuccon, WSDM 2023

Seed Studies

Scells,Zuccon,Koopman,Clark ECIR 2020

Scells,Zuccon,Koopman IRJ 2020

Wang,Scells,Clark,Koopman,Zuccon,
SIGIR 2022

Wang,Scells,Koopman,Zuccon 2023 arxiv

Query Refinement

Query Transformations and Query Chains

Scells.Zuccon SIGIR 2018

Scells,Zuccon,Koopman WWW 2019

Scells,Zuccon,Sharaf,Koopman WWW 2020

Log Analysis of Refinements

Scells,Forbes,Clark,Koopman,Zuccon,
ICTIR 2022

MeSH Terms

Wang,Scells,Koopman,Zuccon,
Int.Sys. Appl. 2022

Wang,Li,Scells,Locke,Zuccon, ADCS 2021

Wang,Li,Zuccon, WSDM 2023

QPP

Scells,Azzopardi,Zuccon,Koopman
SIGIR 2018

Case Study

Comparing Manual to Automatic

- Query for same topic derived manually and automatically
- No tuning performed for automatic method
- Reduction in # studies retrieved results in approx. USD\$90,000 savings

	# Ret	NNR	Precision	Recall
Manual	78,913	6070.23	0.0002	1.000
Automatic	48,945	3496.14	0.0003	1.000

Comparison between query formulated using manual vs automatic methods.

1. prostate.ti,ab.

2. psa.ti,ab.

3. used.ti,ab.

4. either.ti,ab.

5. seed.ti,ab.

6. symptom.ti,ab.

7. ml.ti,ab.

8. toxicities.ti,ab.

9. prostatic.ti,ab.

10. Prostatic Neoplasms/

11. or/1-10

12. beam.ti,ab.

13. brachytherapy.ti,ab.

14. radical.ti,ab.

15. prostatectomy.ti,ab.

16. ebrt.ti,ab.

17. cox.ti,ab.

18. androgen.ti,ab.

19. implantation.ti,ab.

20. consensus.ti,ab.

21. pretreatment.ti,ab.

22. sexual.ti,ab.

23. neoadjuvant.ti,ab.

24. mailed.ti,ab.

25. implant.ti,ab.

26. curative.ti,ab.

27. or/12-26

29. 11 and 27

1. cancer.ti,ab,sh.

2. adenocarcinoma.ti,ab,sh.

3. 1 or 2

4. prostat*.ti,ab,sh.

5. 3 and 4

6. Prostatic Neoplasms/

7. 5 or 6

8. seed*.rs.

9. permanent*.ti,ab,sh.

10. 8 or 9

11. implant*.ti,ab,sh.

12. 10 and 11

13. Brachytherapy/

14. Brachytherapy.ti,ab,sh.

15. or/12-14

16. 7 and 15

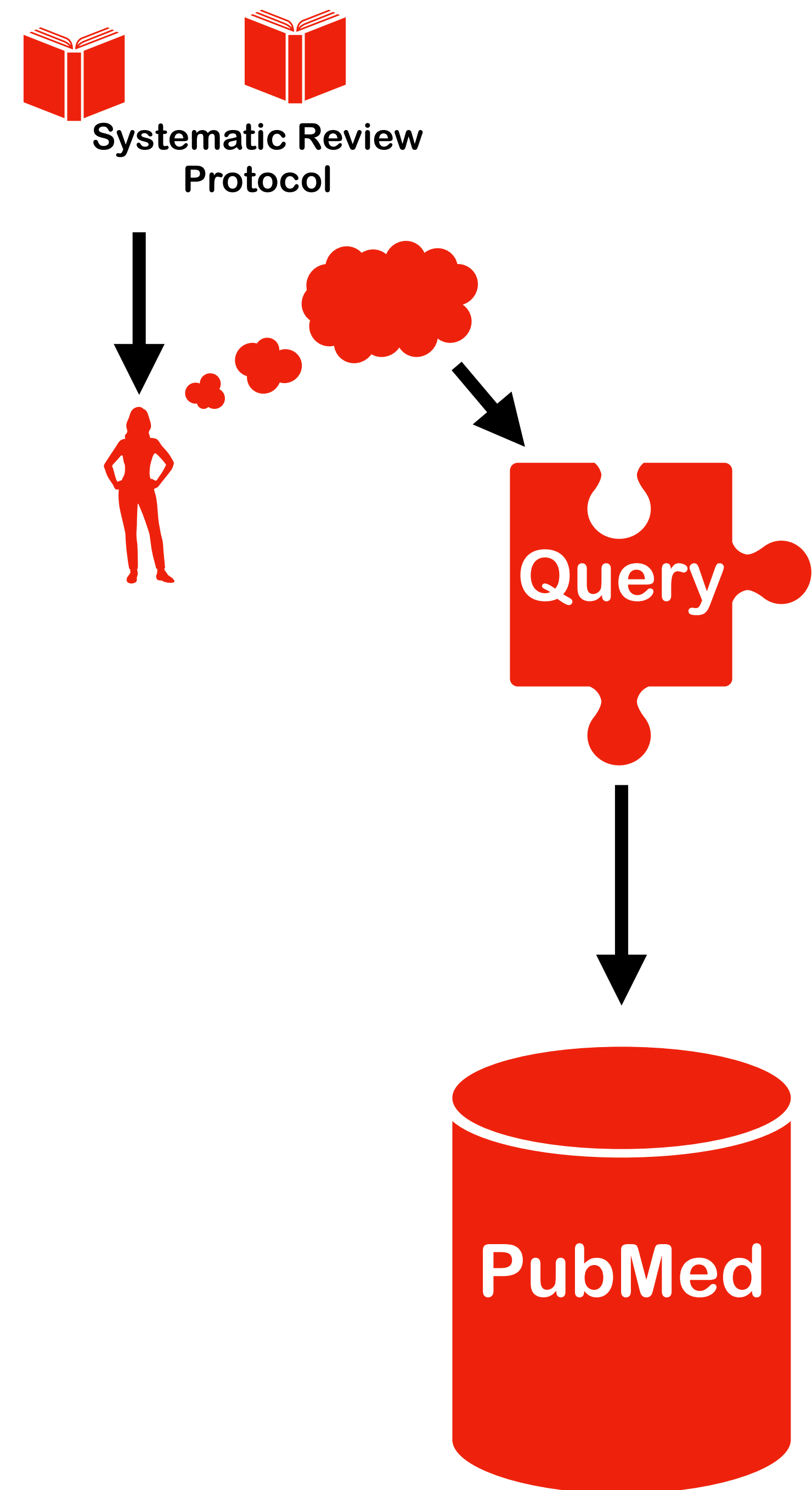
Left: computationally derived query.

Right: Manually derived query.

Commonalities between queries have been highlighted.

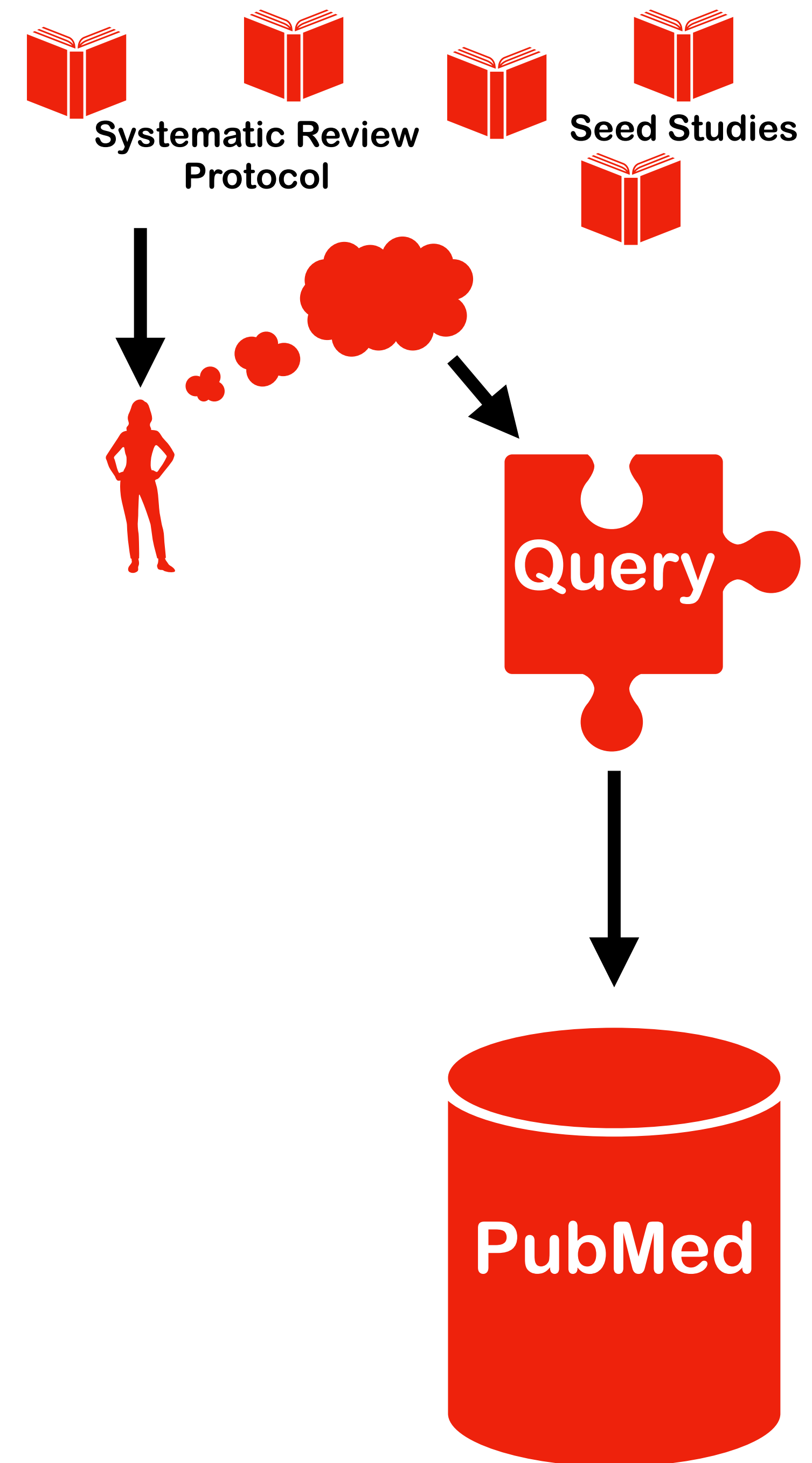
Query Formulation

- Development of highly complex Boolean queries by **information specialists**



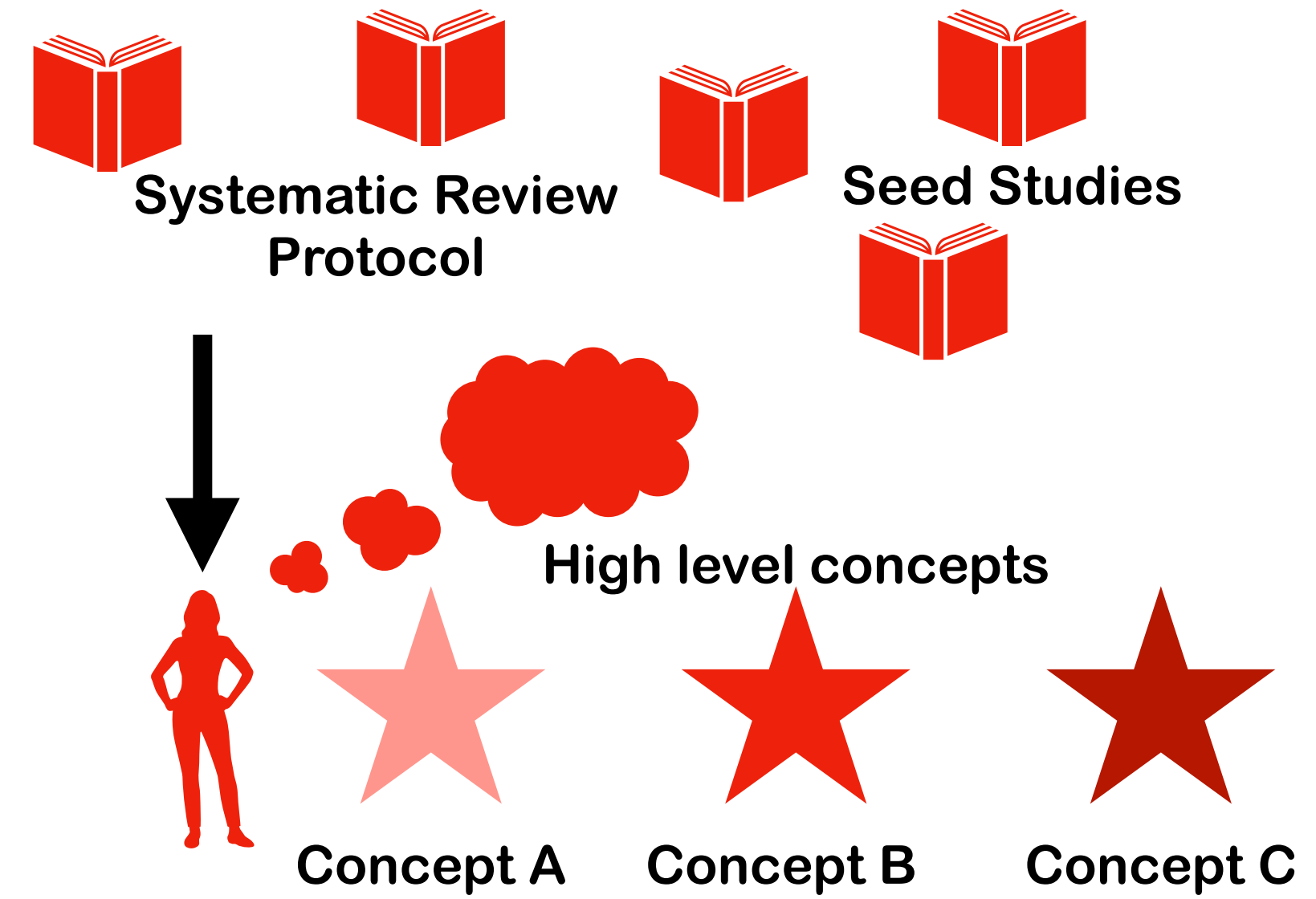
Query Formulation

- Development of highly complex Boolean queries by **information specialists**
- Often the development of Boolean query is “**seeded**” by a set of relevant studies provided to the information specialist a priori, as well as the protocol of the systematic review
- Information specialists gauge the effectiveness of their query using the seed studies, and use them to find terms



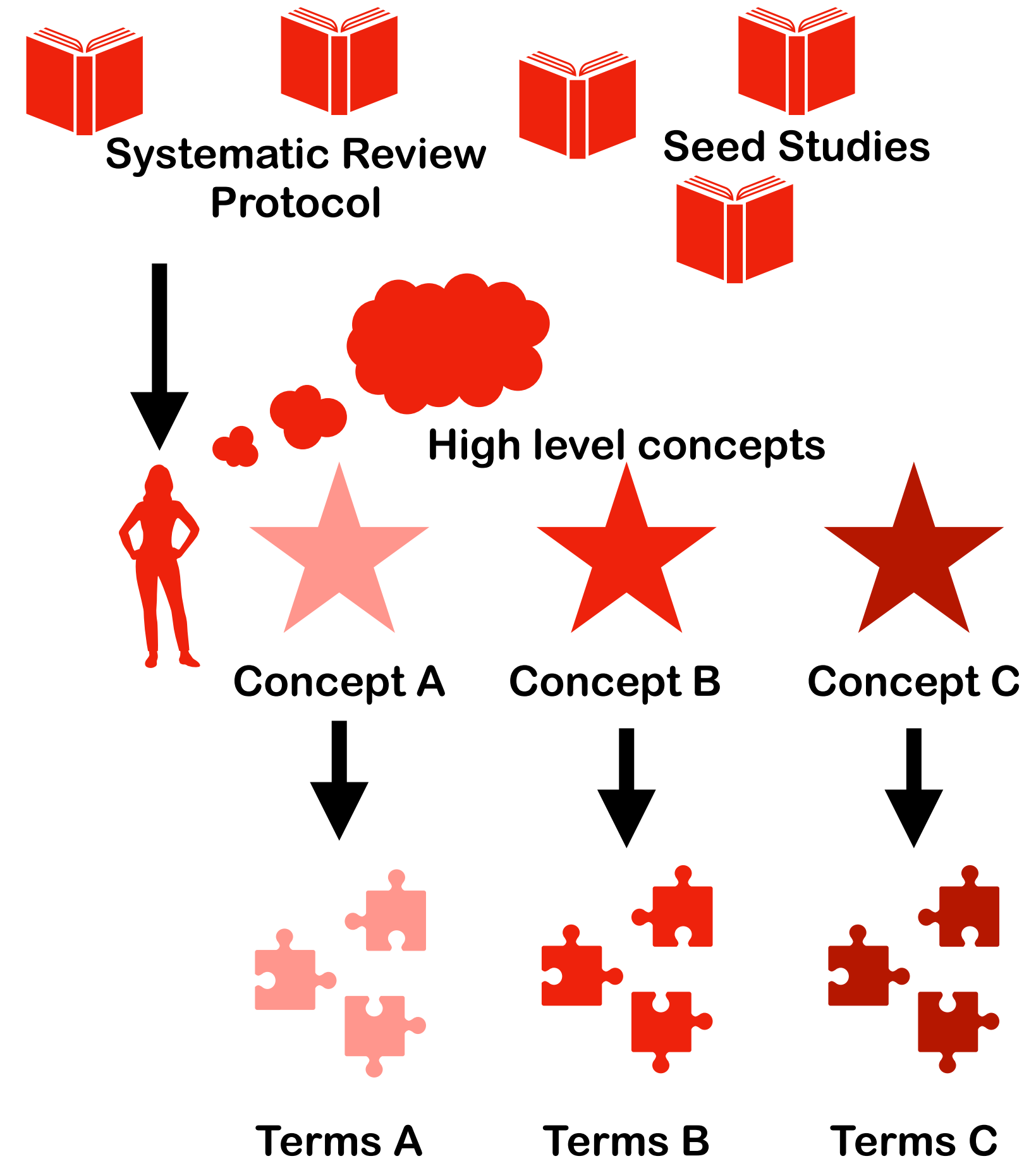
Conceptual Query Formulation

- Information specialist identifies key high-level concepts



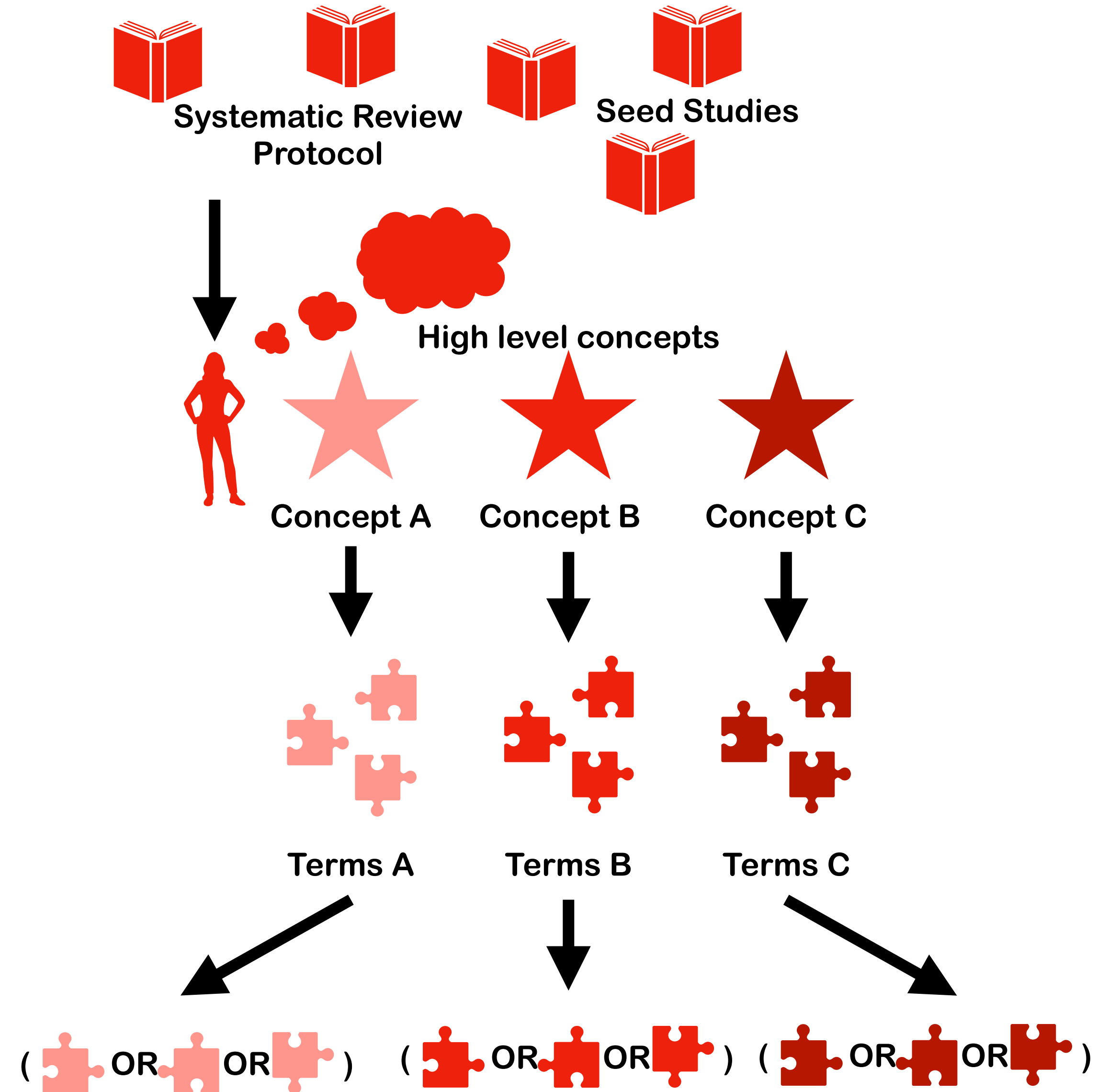
Conceptual Query Formulation

- Information specialist identifies key high-level concepts
- Terms are identified which relate to each concept



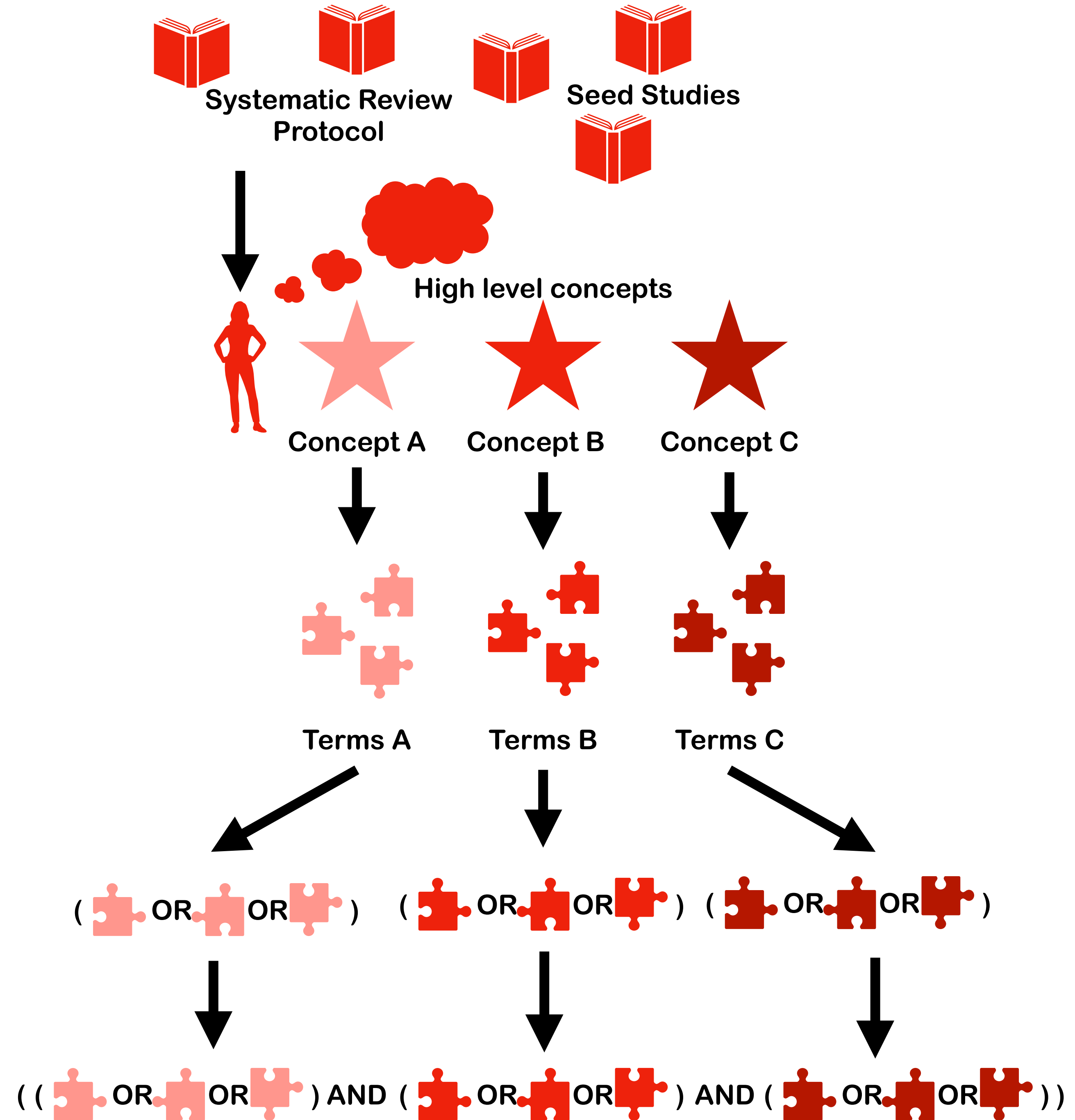
Conceptual Query Formulation

- Information specialist identifies key high-level concepts
- Terms are identified which relate to each concept
- Each group of related terms becomes an OR clause in a Boolean query



Conceptual Query Formulation

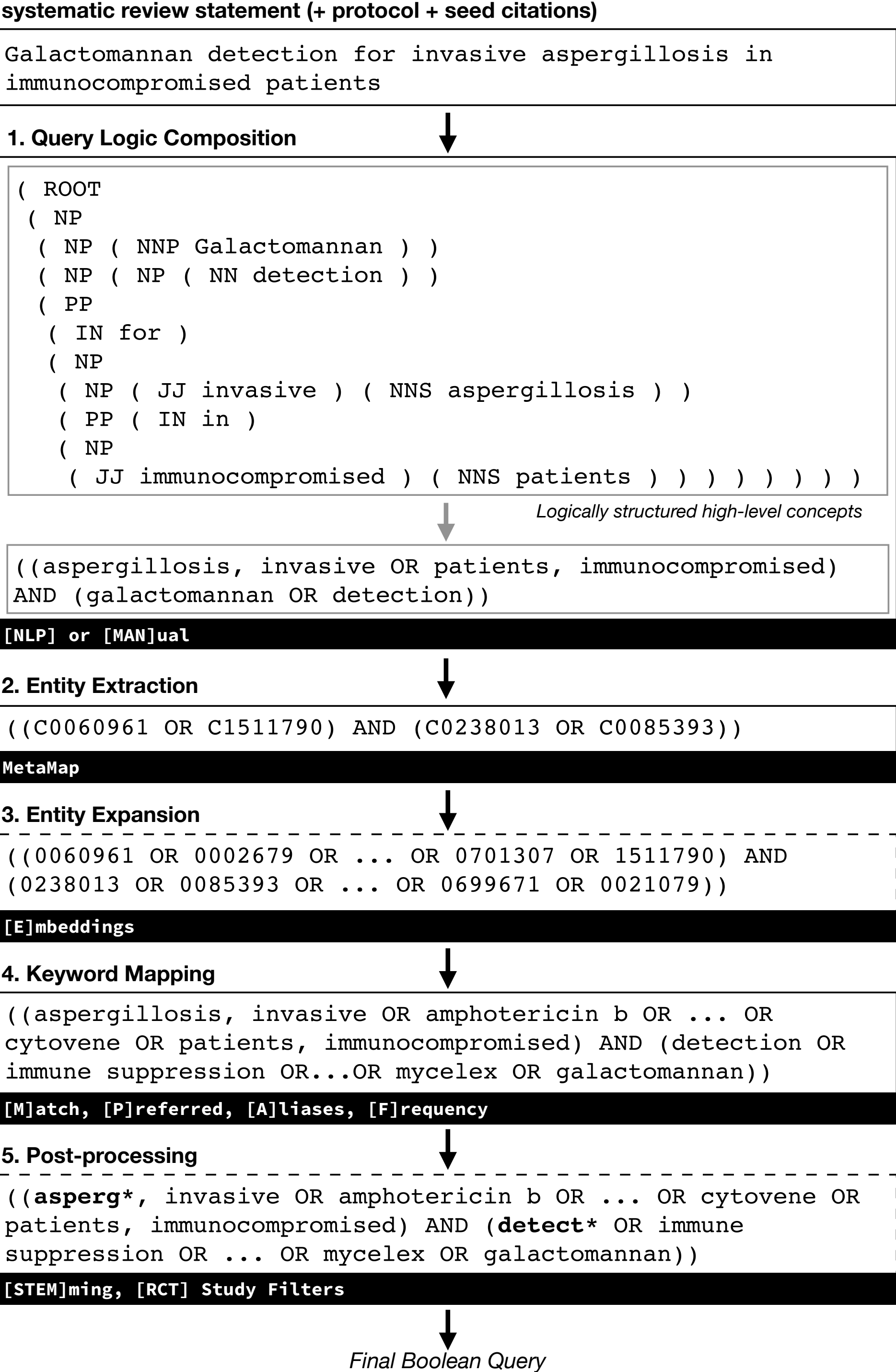
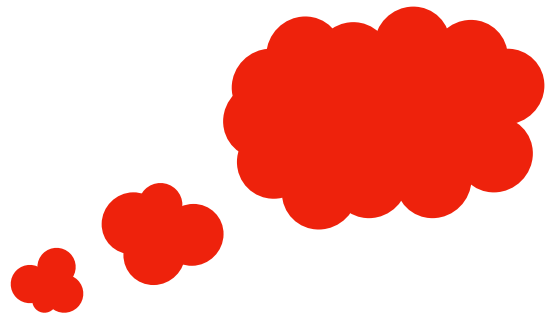
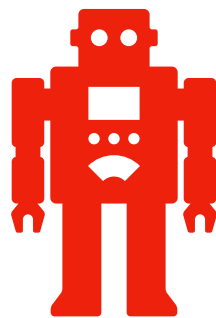
- Information specialist identifies key high-level concepts
- Terms are identified which relate to each concept
- Each group of related terms becomes an OR clause in a Boolean query
- Each clause is grouped by an AND clause



Automatic Conceptual Query Formulation

Scells, Zuccon, Koopman, Clark. Automatic boolean query formulation for systematic review literature search.
WWW 2020

- We propose a computational framework to **automatically formulate Boolean queries** for systematic review literature search
- Approximates the processes and intuitions of information specialists
- Five-step process which encapsulates construction, expansion, and refinement



Query Logic Composition

- Purpose: derive **high-level concepts** of the query
- Take a sentence describing the **topic** of the systematic review
- Create the **logical structure** of the query
- Two methods for constructing proto-query:
 - **[NLP]**: English PFCG parser -> noun phrases -> Boolean clauses
 - **[Manual]**: manually segment sentence -> Boolean clauses

systematic review statement (+ protocol + seed citations)

Galactomannan detection for invasive aspergillosis in immunocompromised patients

1. Query Logic Composition

```
( ROOT
  ( NP
    ( NP ( NNP Galactomannan ) )
    ( NP ( NP ( NN detection ) )
      ( PP
        ( IN for )
        ( NP
          ( NP ( JJ invasive ) ( NNS aspergillosis ) )
          ( PP ( IN in )
            ( NP
              ( JJ immunocompromised ) ( NNS patients ) ) ) ) ) ) ) ) ) )
```

Logically structured high-level concepts

```
((aspergillosis, invasive OR patients, immunocompromised)
AND (galactomannan OR detection))
```

[NLP] or [MAN]ual

2. Entity Extraction

```
((C0060961 OR C1511790) AND (C0238013 OR C0085393))
```

MetaMap

3. Entity Expansion

```
((0060961 OR 0002679 OR ... OR 0701307 OR 1511790) AND
(0238013 OR 0085393 OR ... OR 0698671 OR 0021079))
```

Entity Extraction

- **Entities extracted** from each clause in proto-query
- UMLS entities using **MetaMap**
- Models the **high-level concepts** an information specialist would conceptualise themselves
- Other entity extraction methods and terminologies could be used



systematic review statement (+ protocol + seed citations)

Galactomannan detection for invasive aspergillosis in immunocompromised patients

1. Query Logic Composition

```
( ROOT
  ( NP
    ( NP ( NNP Galactomannan ) )
    ( NP ( NP ( NN detection ) )
    ( PP
      ( IN for )
      ( NP
        ( NP ( JJ invasive ) ( NNS aspergillosis ) )
        ( PP ( IN in )
          ( NP
            ( JJ immunocompromised ) ( NNS patients ) ) ) ) ) ) ) ) )
```

Logically structured high-level concepts

((aspergillosis, invasive OR patients, immunocompromised)
AND (galactomannan OR detection))

[NLP] or [MAN]ual

2. Entity Extraction

((C0060961 OR C1511790) AND (C0238013 OR C0085393))

MetaMap

3. Entity Expansion

((0060961 OR 0002679 OR ... OR 0701307 OR 1511790) AND
(0238013 OR 0085393 OR ... OR 0699671 OR 0021079))

[E]mbeddings

4. Keyword Mapping

((aspergillosis, invasive OR amphotericin b OR ... OR

Entity Expansion

- Optional step which **broadens the scope** of the query
- If not applied, the query may be too narrow
- Method:
 - **[E]mbed** entity into high-dimensional vector space using word embedding techniques
 - Measure distance to similar embeddings
 - Take top-k entities with highest similarity

1. Query Logic Composition

```
( ROOT
  ( NP
    ( NP ( NNP Galactomannan ) )
    ( NP ( NP ( NN detection ) )
      ( PP
        ( IN for )
        ( NP
          ( NP ( JJ invasive ) ( NNS aspergillosis ) )
          ( PP ( IN in )
            ( NP
              ( JJ immunocompromised ) ( NNS patients ) ) ) ) ) ) ) ) ) )
```

Logically structured high-level concepts

```
((aspergillosis, invasive OR patients, immunocompromised)
AND (galactomannan OR detection))
```

[NLP] or [MAN]ual

2. Entity Extraction

```
((C0060961 OR C1511790) AND (C0238013 OR C0085393))
```

MetaMap

3. Entity Expansion

```
((0060961 OR 0002679 OR ... OR 0701307 OR 1511790) AND
(0238013 OR 0085393 OR ... OR 0699671 OR 0021079))
```

[E]mbeddings

4. Keyword Mapping

```
((aspergillosis, invasive OR amphotericin b OR ... OR
cytovene OR patients, immunocompromised) AND (detection OR
immune suppression OR...OR mycelex OR galactomannan))
```

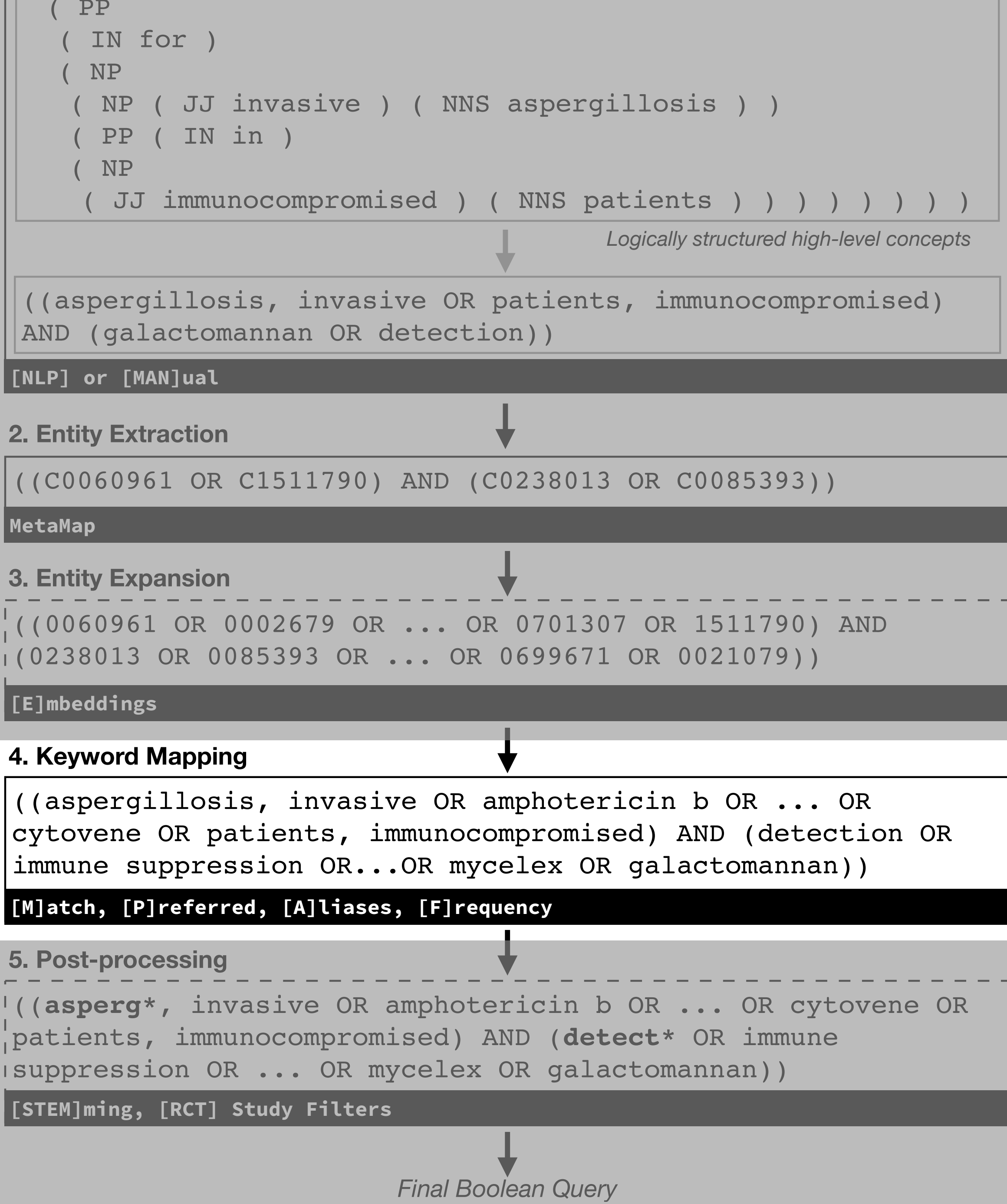
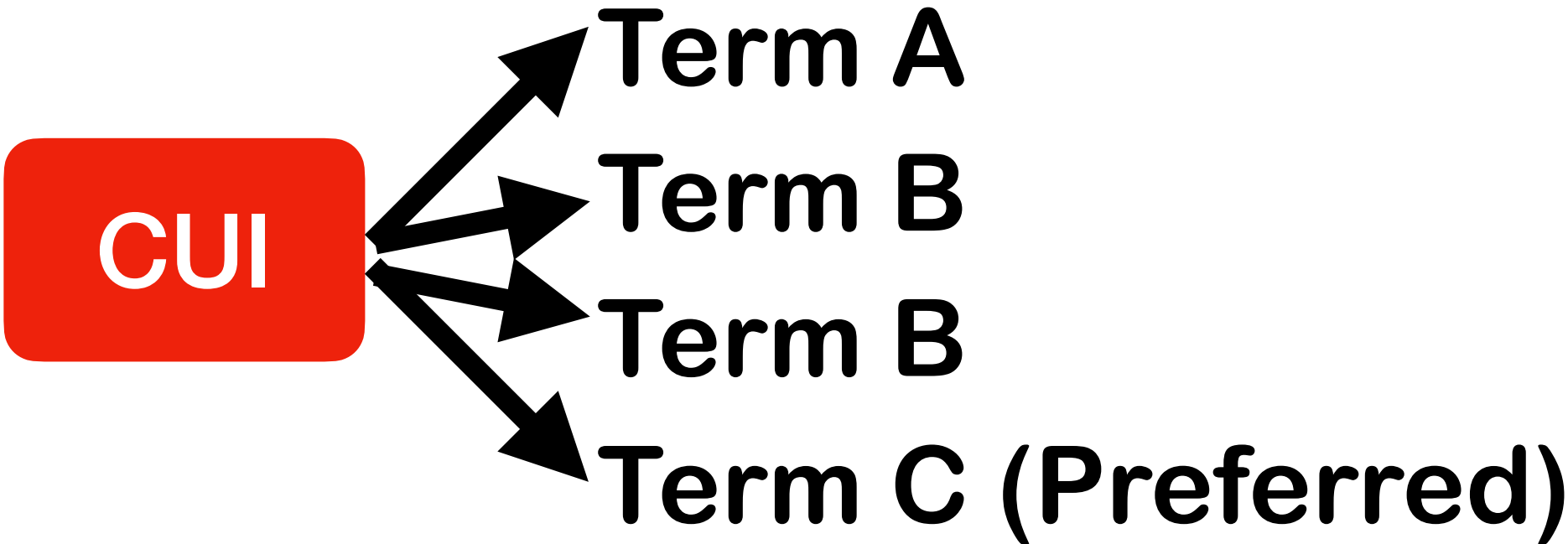
[M]atch, [P]referred, [A]liases, [F]requency

5. Post-processing

```
((asperg*, invasive OR amphotericin b OR ... OR cytovene OR
patients, immunocompromised) AND (detect* OR immune
```

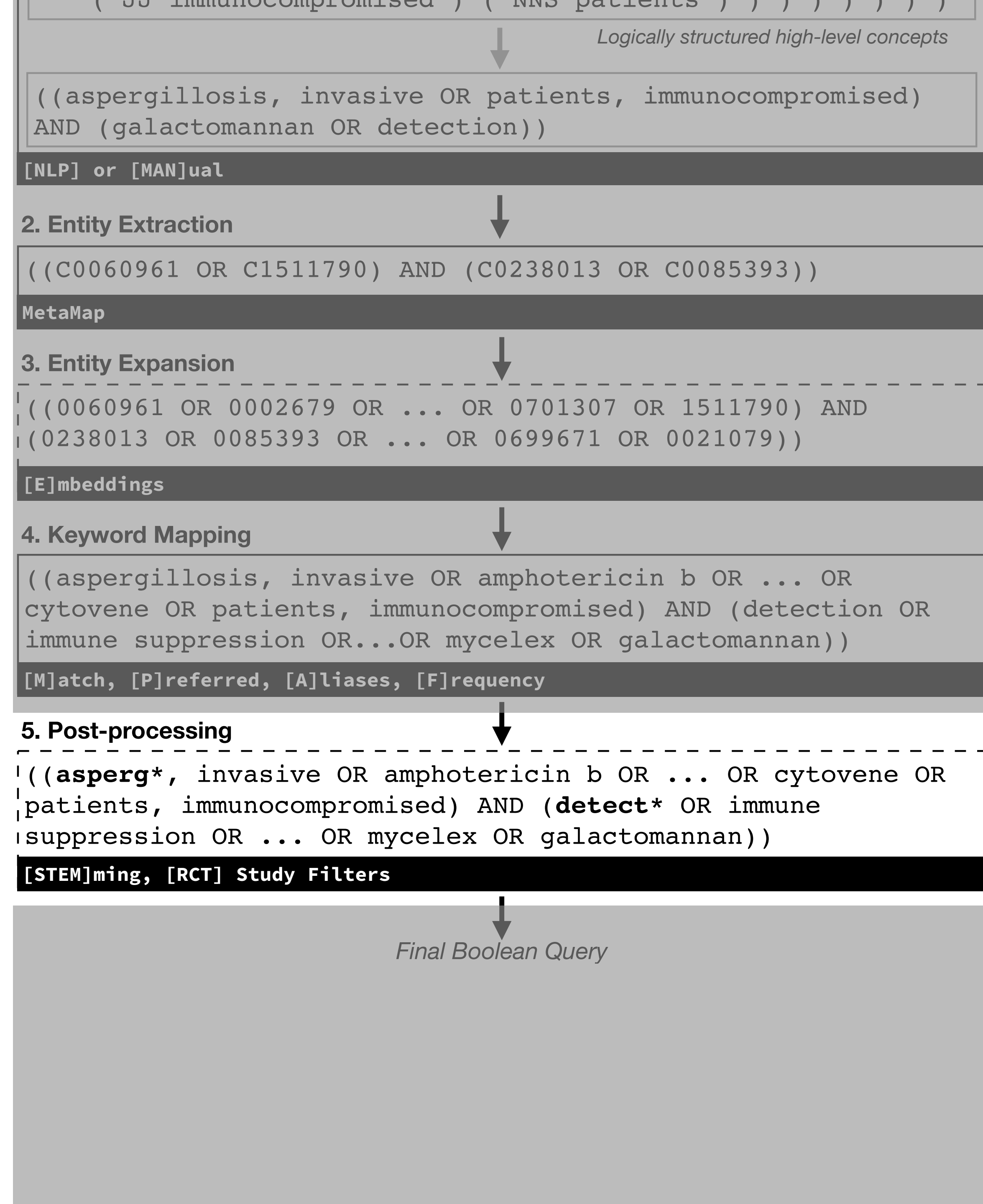

Keyword Mapping

- Map the entities into **keywords**
- **[M]atch**: original terms matched in query logic composition (cannot be used with entity expansion)
- **[P]referred**: UMLS `Preferred` term associated with each entity
- **[A]lias**: All of the aliases in UMLS for an entity
- **[F]requent**: Only most frequently used term for entity in UMLS is used



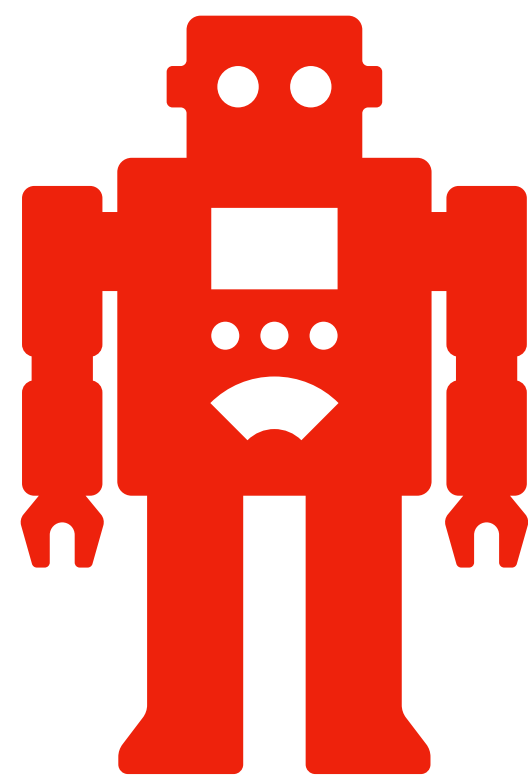
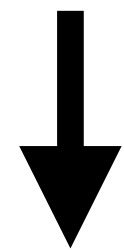
Post-processing

- Queries can and are typically processed further to broaden or narrow the scope
- **Stemming:** extract stems from a collection of existing systematic review queries and substitute terms in query for the longest match found in list of stems (if any)
- **Filters:** these are quasi-standard Boolean expressions designed to retrieve specific portions of databases. We add a common RCT filter to query



Automatic Objective Query Formulation

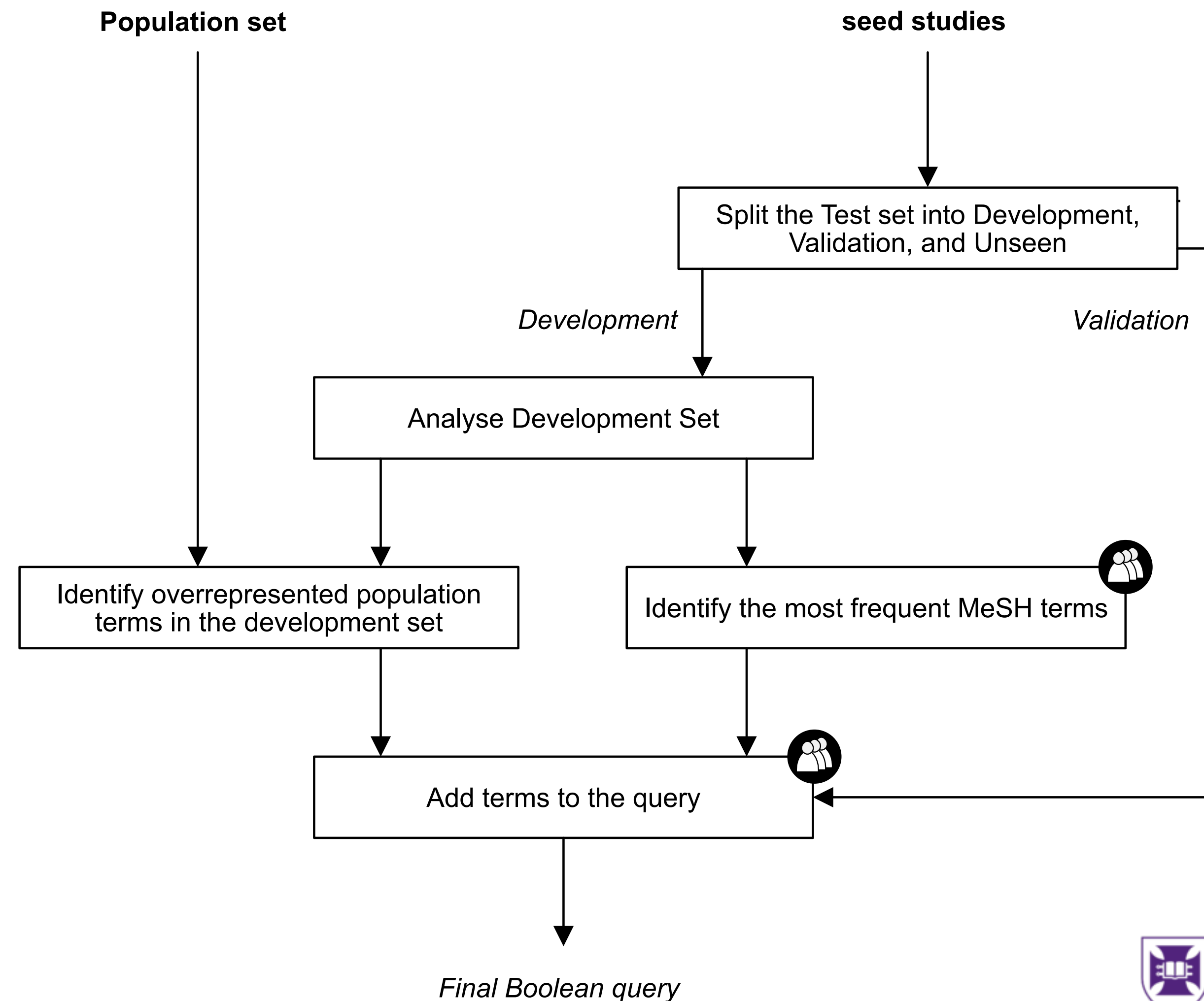
Scells, Zuccon, Koopman, Clark. A computational approach
for objectively derived systematic review search strategies.
ECIR 2020



1. (adrenergic* and antagonist*).tw.
2. (adrenergic* and block\$).tw.
- ...
18. 7 and 17

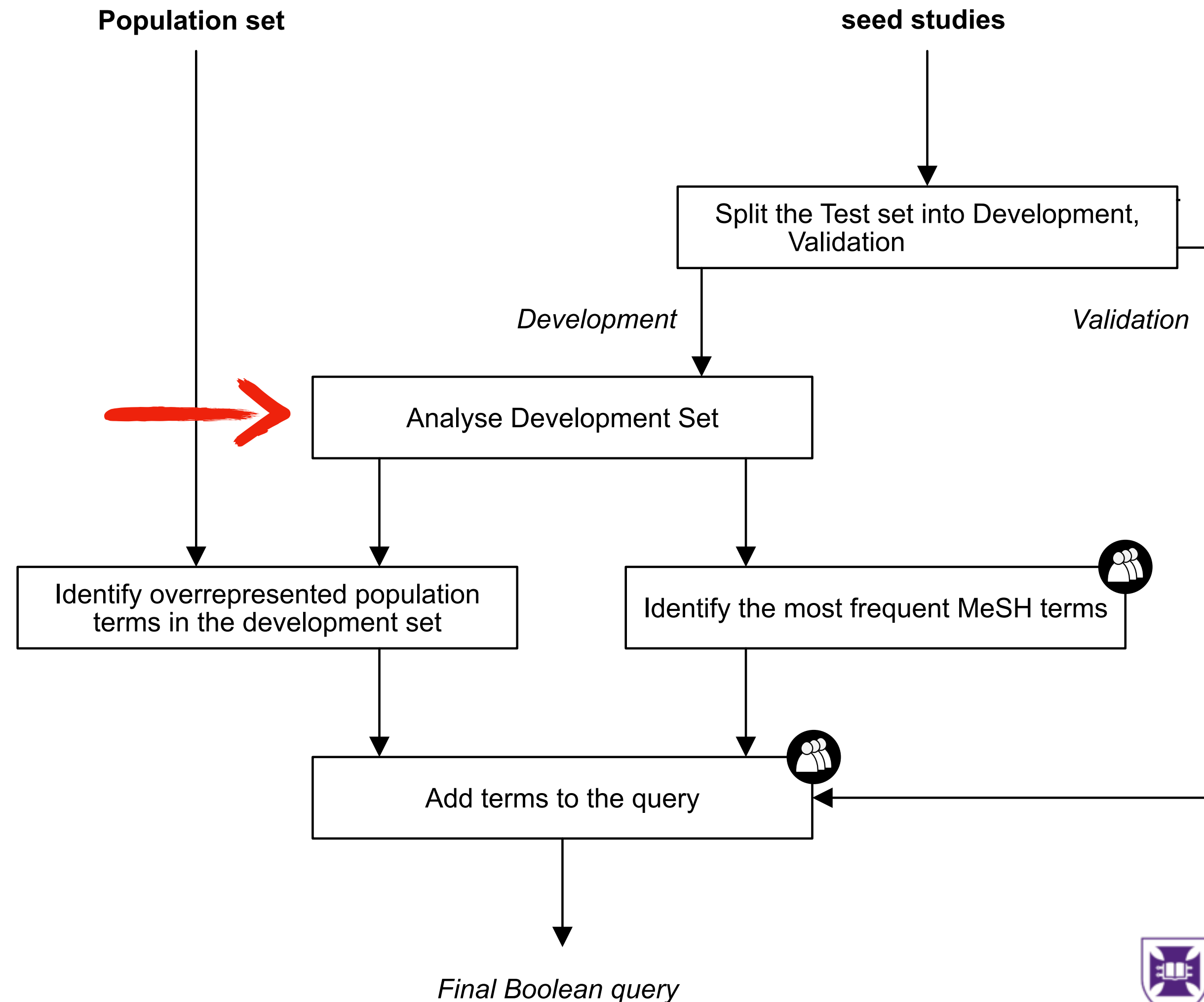
Automatic Objective Query Formulation

- Find prominent terms from docs
- Add these terms to query



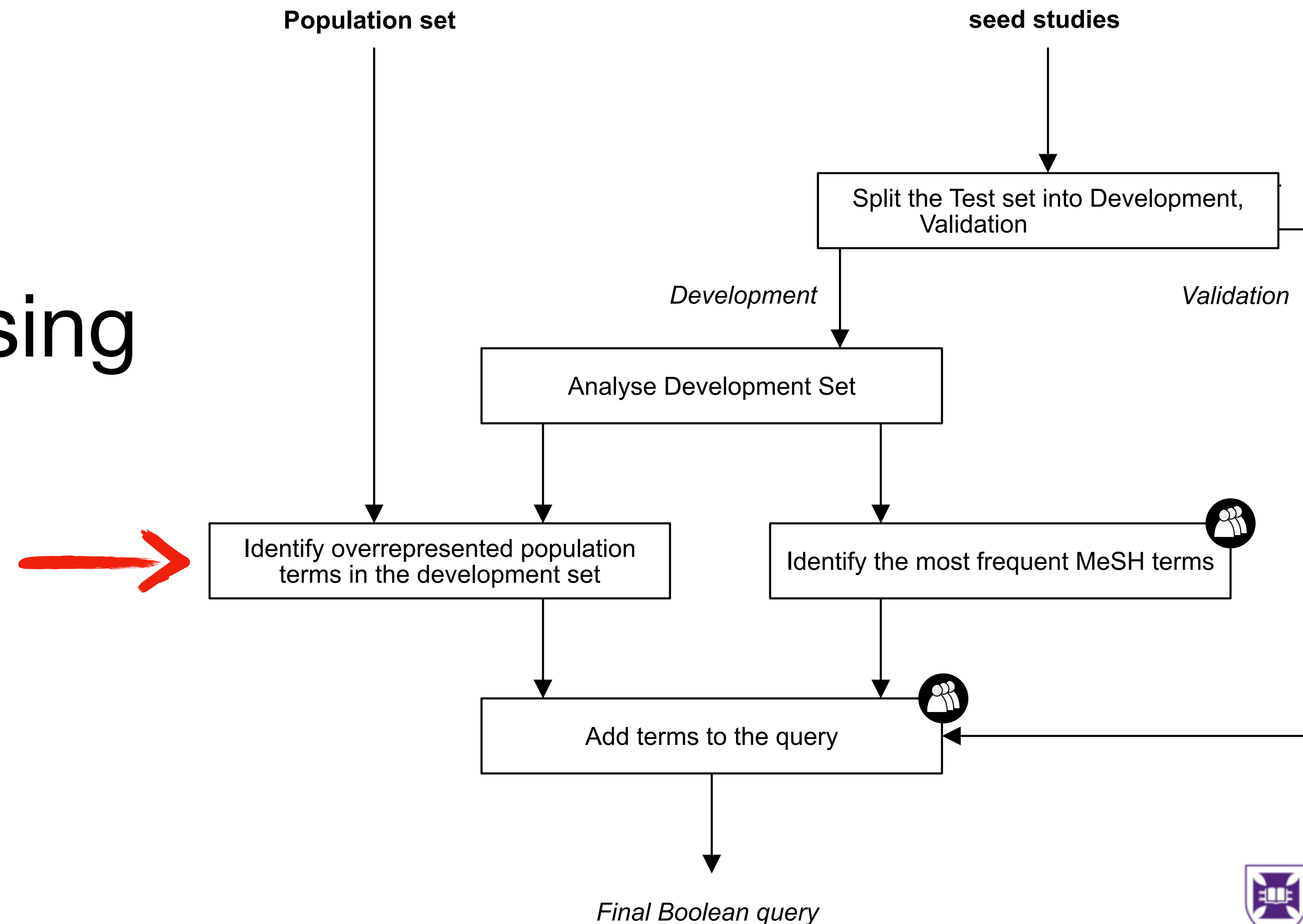
Automatic Objective Query Formulation

- Extract keywords



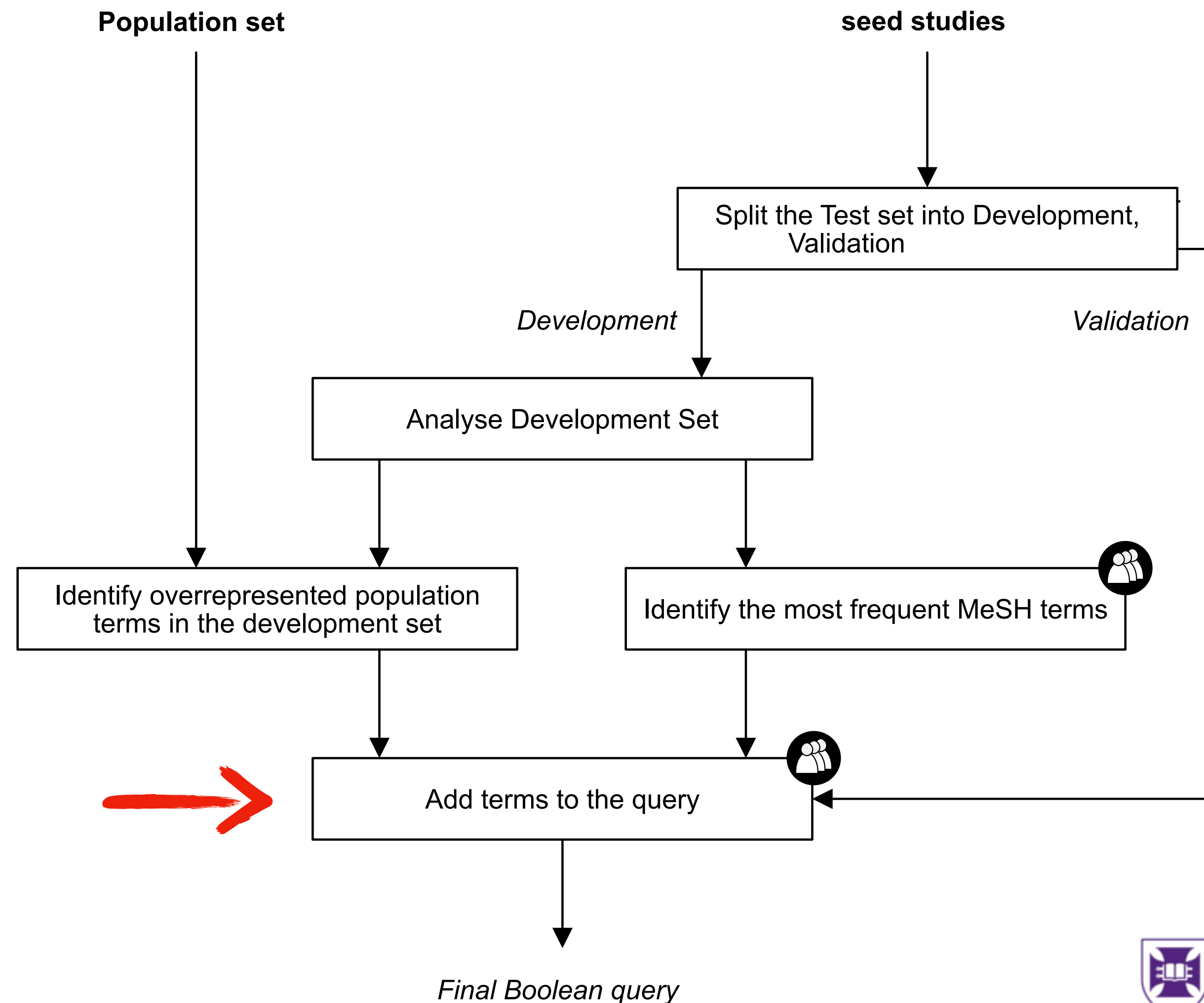
Automatic Objective Query Formulation

- Rank terms using document frequency



Automatic Objective Query Formulation

- Add terms to query



Formulating Boolean Queries with ChatGPT

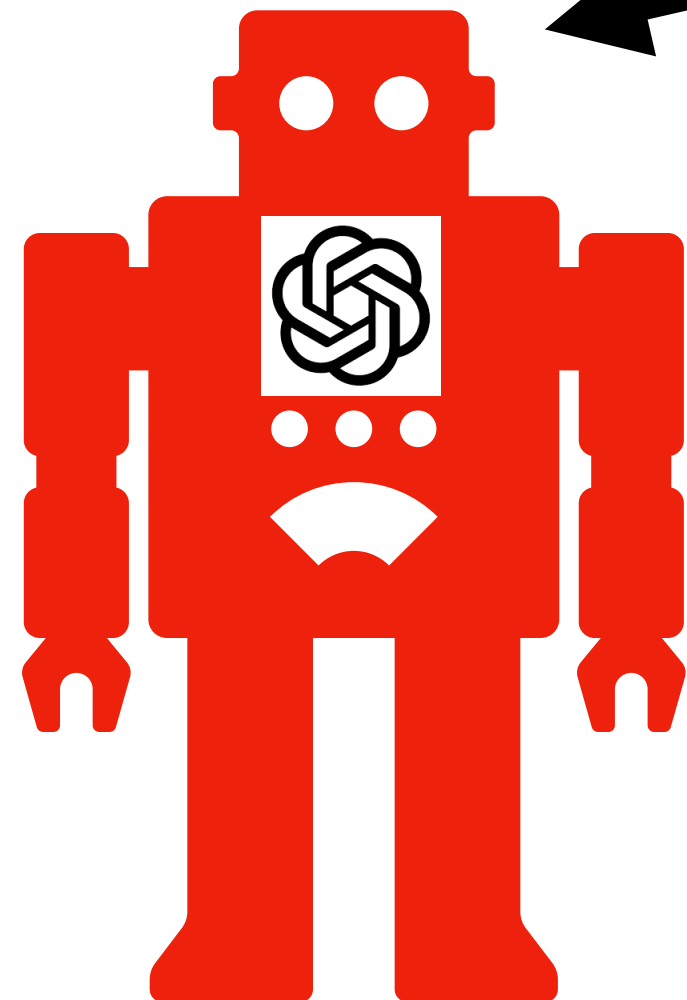
Wang, Scells, Koopman, Zuccon, 2023. Can ChatGPT write a good boolean query for systematic review literature search?
arXiv:2302.03495.

Prompting:

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?

Task-input:

{review_title}



1. (adrenergic* and antagonist*).tw.
2. (adrenergic* and block\$).tw.
- ...
18. 7 and 17

Single Prompts for Query Formulation

- Prompt instructs to formulate systematic review Boolean queries using the title of the review
- We explore 5 prompt variations
- **simple:** only uses one sentence to briefly state the task for ChatGPT

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?

Single Prompts for Query Formulation

- Prompt instructs to formulate systematic review Boolean queries using the title of the review
- We explore 5 prompt variations
 - **simple:** only uses one sentence to briefly state the task for ChatGPT
 - **detailed (x2):** background story is included, which justifies clearly what is required for ChatGPT to successfully complete the task

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.

Single Prompts for Query Formulation

- Prompt instructs to formulate systematic review Boolean queries using the title of the review
- We explore 5 prompt variations
 - **simple:** only uses one sentence to briefly state the task for ChatGPT
 - **detailed (x2):** background story is included, which justifies clearly what is required for ChatGPT to successfully complete the task

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.

Single Prompts for Query Formulation

- Prompt instructs to formulate systematic review Boolean queries using the title of the review
- We explore 5 prompt variations
 - **simple:** only uses one sentence to briefly state the task for ChatGPT
 - **detailed (x2):** background story is included, which justifies clearly what is required for ChatGPT to successfully complete the task
 - **with examples (x2):** also includes an expected query formulation example, so that ChatGPT knows what is expected for it to generate a high-quality answer

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.

Single Prompts for Q

- Prompt instructs to formulate systematic review Boolean queries using the title of the review
- We explore 5 prompt variations
 - **simple:** only uses one sentence to briefly state the task for ChatGPT
 - **detailed (x2):** background story is included, which justifies clearly what is required for ChatGPT to successfully complete the task
 - **with examples (x2):** also includes an expected query formulation example, so that ChatGPT knows what is expected for it to generate a high-quality answer

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.

Single Prompts for Query Refinement

- Prompts that provide ChatGPT with a title of a systematic review and a corresponding Boolean query for that review
- Instruct the model to produce a modification of the query that leads to higher search effectiveness.
- We explore 2 prompt variations
 - **simple:** only uses one sentence to briefly state the task for ChatGPT

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.

Single Prompts for Query Refinement

- Prompts that provide ChatGPT with a title of a systematic review and a corresponding Boolean query for that review
- Instruct the model to produce a modification of the query that leads to higher search effectiveness.
- We explore 2 prompt variations
 - **simple:** only uses one sentence to briefly state the task for ChatGPT
 - **with examples:** also includes an expected query formulation example, so that ChatGPT knows what is expected for it to generate a high-quality answer

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:

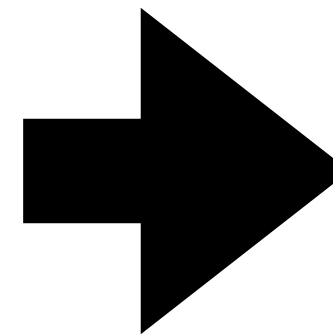
“**{example_review_refined_query}**”. Now your task is to correct a systematic review Boolean query: “**{initial_query}**” for information need “**{review_title}**”, so it can retrieve fewer irrelevant documents and more relevant documents.

Guided Prompts for Query Formulation

- Multi-step (conversational) prompt that follows the same logic from one of objective automated query formulation method
- Four step pipeline:
 1. Boolean query terms are identified using one seed study

Prompt

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.



ChatGPT Output

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.

Guided Prompts for Query Formulation

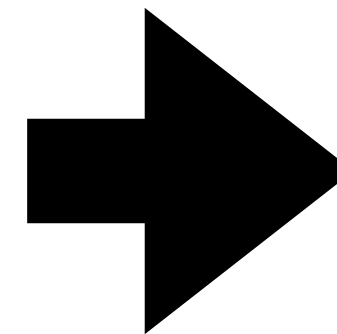
- Multi-step (conversational) prompt that follows the same logic from one of objective automated query formulation method
- Four step pipeline:
 1. Boolean query terms are identified using one seed study
 2. terms are classified into four categories: terms relating to health conditions (A), terms relating to a treatment (B), terms relating to types of study design (C), or others (N/A)

Prompt

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).



ChatGPT Output

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.

Guided Prompts for Query Formulation

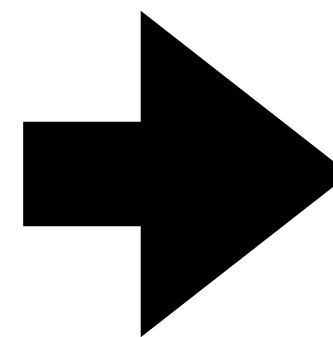
- Multi-step (conversational) prompt that follows the same logic from one of objective automated query formulation method
- Four step pipeline:
 1. Boolean query terms are identified using one seed study
 2. terms are classified into four categories: terms relating to health conditions (A), terms relating to a treatment (B), terms relating to types of study design (C), or others (N/A)
 3. terms in the same category are combined by 'OR', and final queries are combined by 'AND'

Prompt

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 (itemB1[Title/Abstract] itemB3[Title/Abstract]) OR AND itemA2[Title/Abstract]) itemB2[Title/Abstract] (itemC1[Title/Abstract] AND OR itemC2[Title/Abstract] OR itemC3[Title/Abstract]))



ChatGPT Output

((differentiated[Title/Abstract] OR thyroid[Title/Abstract] OR carcinoma[Title/Abstract] OR papillary[Title/Abstract] OR microcarcinoma[Title/Abstract]) OR incidence[Title/Abstract] AND (cancer[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 fibrosis[Title/Abstract] OR infiltrate[Title/Abstract] OR development[Title/Abstract] frequency[Title/Abstract])) AND OR (autopsy[Title/Abstract] OR surgical[Title/Abstract] OR material[Title/Abstract] OR series[Title/Abstract] OR specimens[Title/Abstract] OR cases[Title/Abstract]))

Guided Prompts for Query Formulation

- Multi-step (conversational) prompt that follows the same logic from one of objective automated query formulation method
- Four step pipeline:
 1. Boolean query terms are identified using one seed study
 2. terms are classified into four categories: terms relating to health conditions (A), terms relating to a treatment (B), terms relating to types of study design (C), or others (N/A)
 3. terms in the same category are combined by 'OR', and final queries are combined by 'AND'
 4. ChatGPT is asked to refine the query by adding more terms such as MeSHTerms

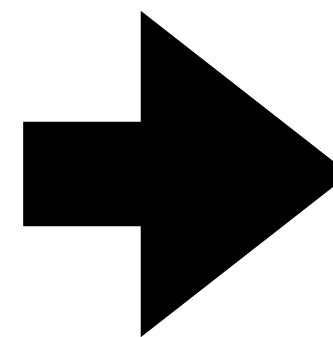
Prompt

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.

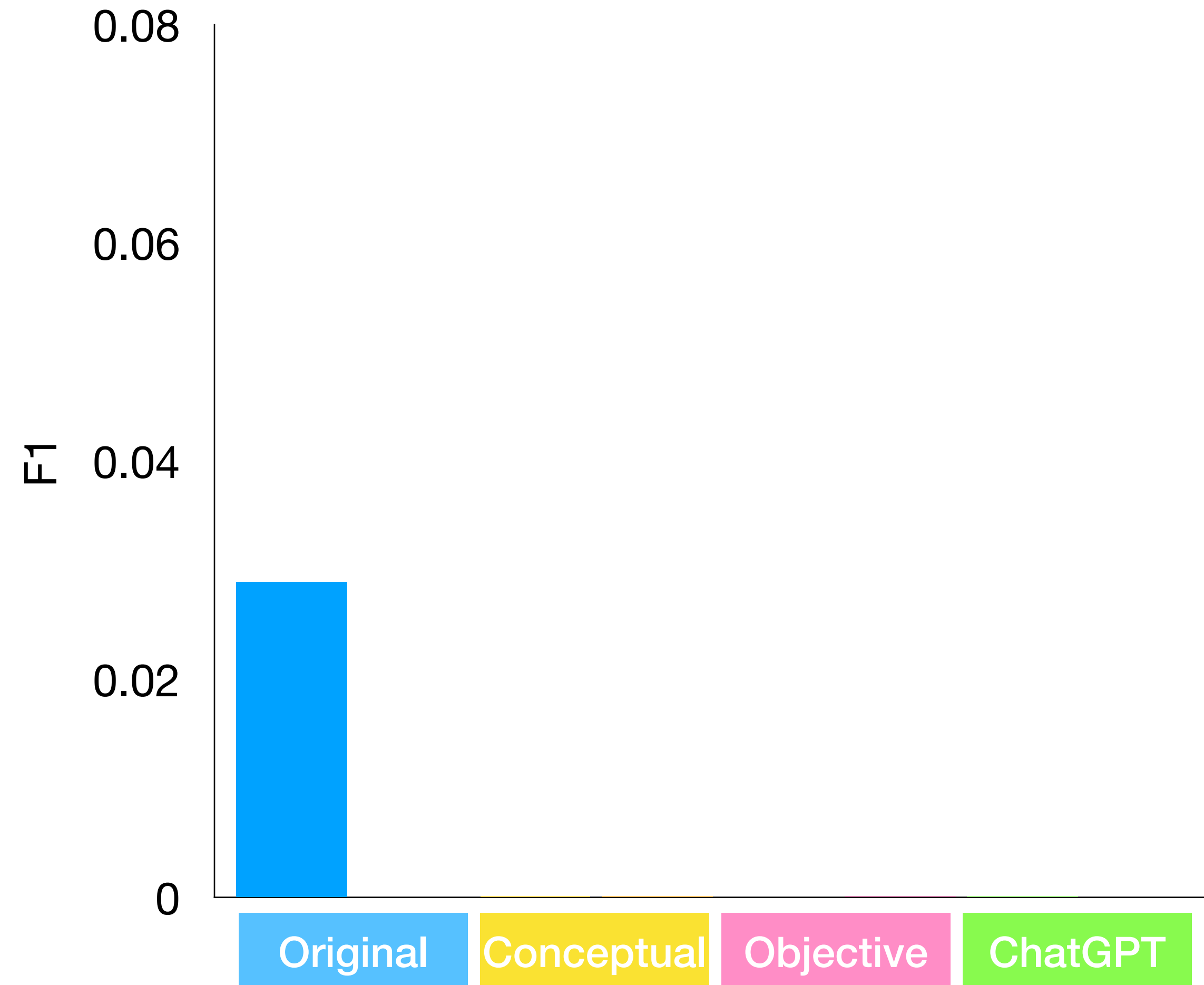
The final query still needs to be executable on PubMed, so it should be a valid query.



ChatGPT Output

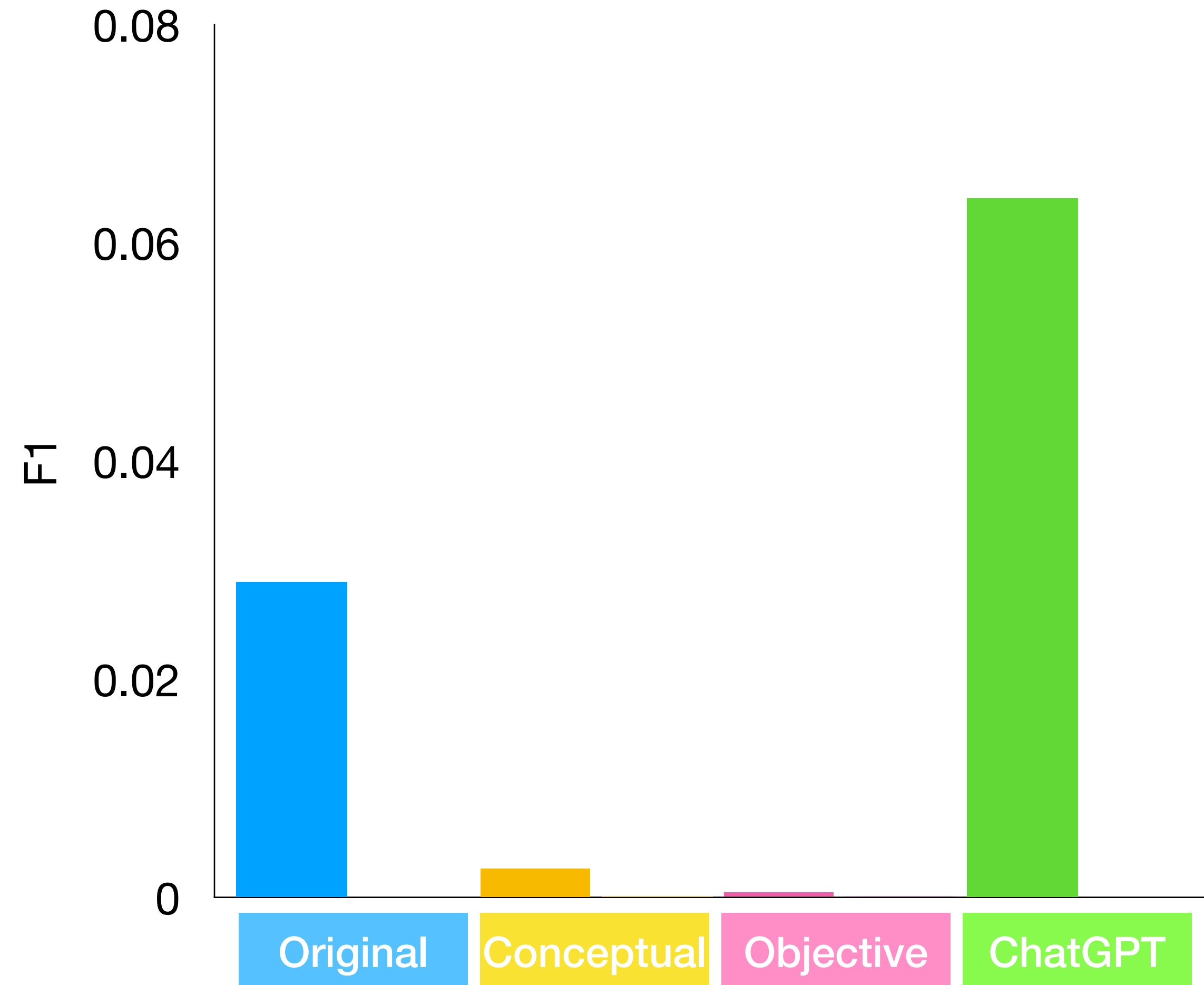
((differentiated thyroid cancer[MeSH] OR "differentiated thyroid"[All Fields] OR "thyroid carcinoma"[All Fields] OR "papillary microcarcinoma"[All Fields]) AND (prevalence[All Fields] OR incidence[MeSH] OR "etiology of"[All Fields] OR "risk factors"[All Fields] OR gender[All Fields] OR hormonal[All Fields] 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]))

How Well does Automatic Query Formulation Perform?



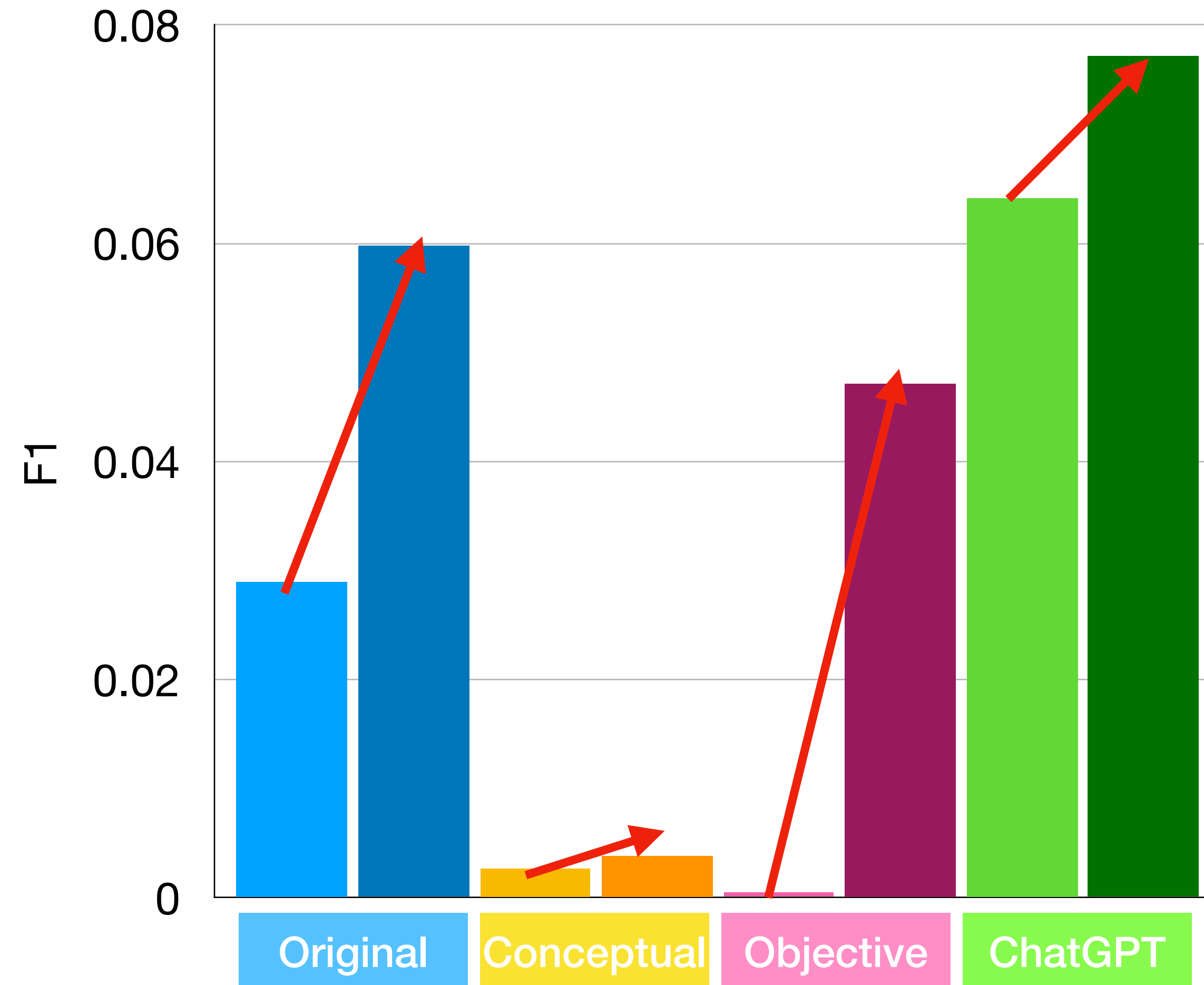
- Results obtained on CLEF TAR 2017 + 2018 (Diagnostic Test Accuracy)
- Once queries are obtained, they are executed on PubMed
- All methods show trade off prec-rec w.r.t. original query
- Results on Seed Collection are similar

How Well does Automatic Query Formulation Perform?



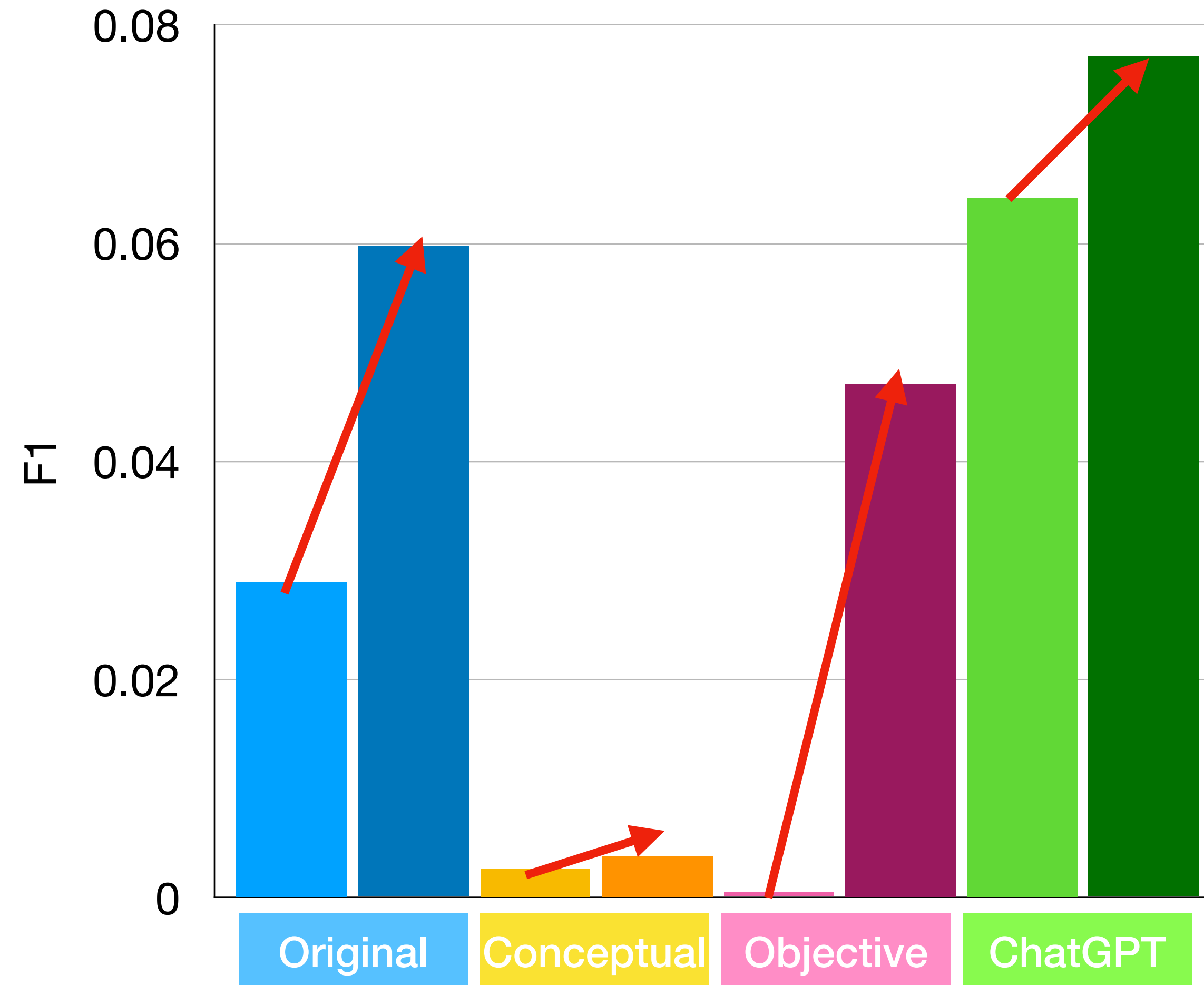
- Conceptual and Objective formulation perform poorly in terms of F1
 - But note that objective has highest recall with large trade-offs for precision
- ChatGPT (single prompt with example) largely outperforms the original query
 - Gains are in precision, with recall taking a hit
 - Other single prompt methods do not perform as well

How Well does Automatic Query Formulation Perform?



- Query refinement performed with ChatGPT (single prompt)
 - Refinement improves all queries
 - Gains are in precision

How Well does Automatic Query Formulation Perform?

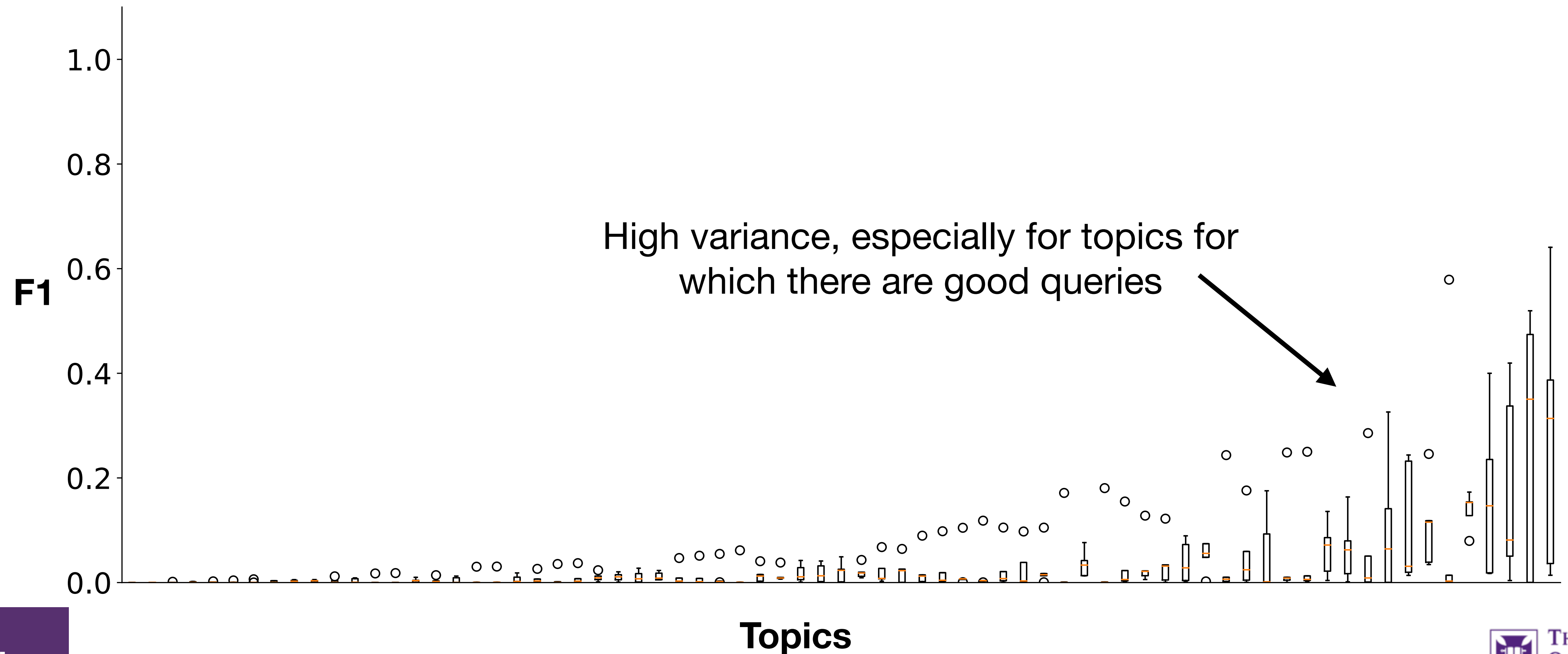


- Query refinement performed with ChatGPT (single prompt)
 - Refinement improves all queries
 - Gains are in precision
- **Guided prompting** (not shown) provides even higher effectiveness, delivering both precision and recall improvements

What makes an automatically formulated query successful?

- Analysis performed on CHatGPT generated queries
- Poorly performing ChatGPT generated queries tend to **retrieve a large number of results** — mostly due to a large number of OR clauses
- Some ChatGPT queries contain many **incorrect MeSH Terms**
 - Generated queries: avg 3 MeSH Terms generated, 55% not in MeSH vocabulary
 - Refined queries: avg 1.5 MeSH Terms generated, 66% not in MeSH vocabulary
- **Relevance judgements** were done using the original query so may be biased towards this
 - number of unjudged documents for “failing queries” more than 94.8%

Careful, LLMs are stochastic



Other Works on Query Automation

- Alharbi, Stevenson, 2020. *Refining Boolean queries to identify relevant studies for systematic review updates*. JAMIA
 - Improve effectiveness in identifying relevant studies for **updates to systematic reviews** using automated transformations of Boolean queries.
- MacFarlane, Russell-Rose, Shokrane, 2022. *Search strategy formulation for systematic reviews: Issues, challenges and opportunities*. Intelligent Systems with Applications
 - set of **design principles** and a conceptual framework for **search strategy formulation**
- Zang, Russell-Rose, 2023. *A Prototype “Debugger” for Search Strategies*. CHIIR

Tools

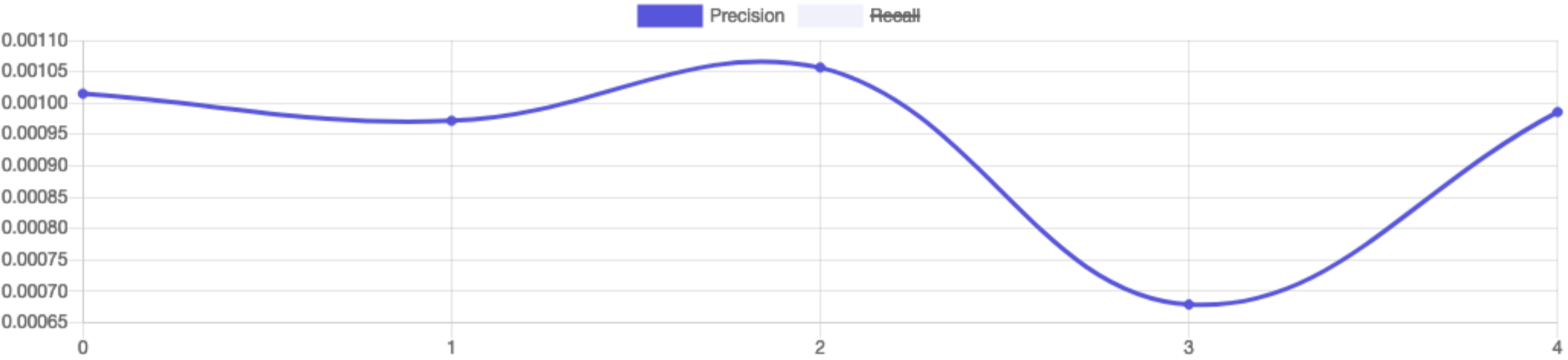
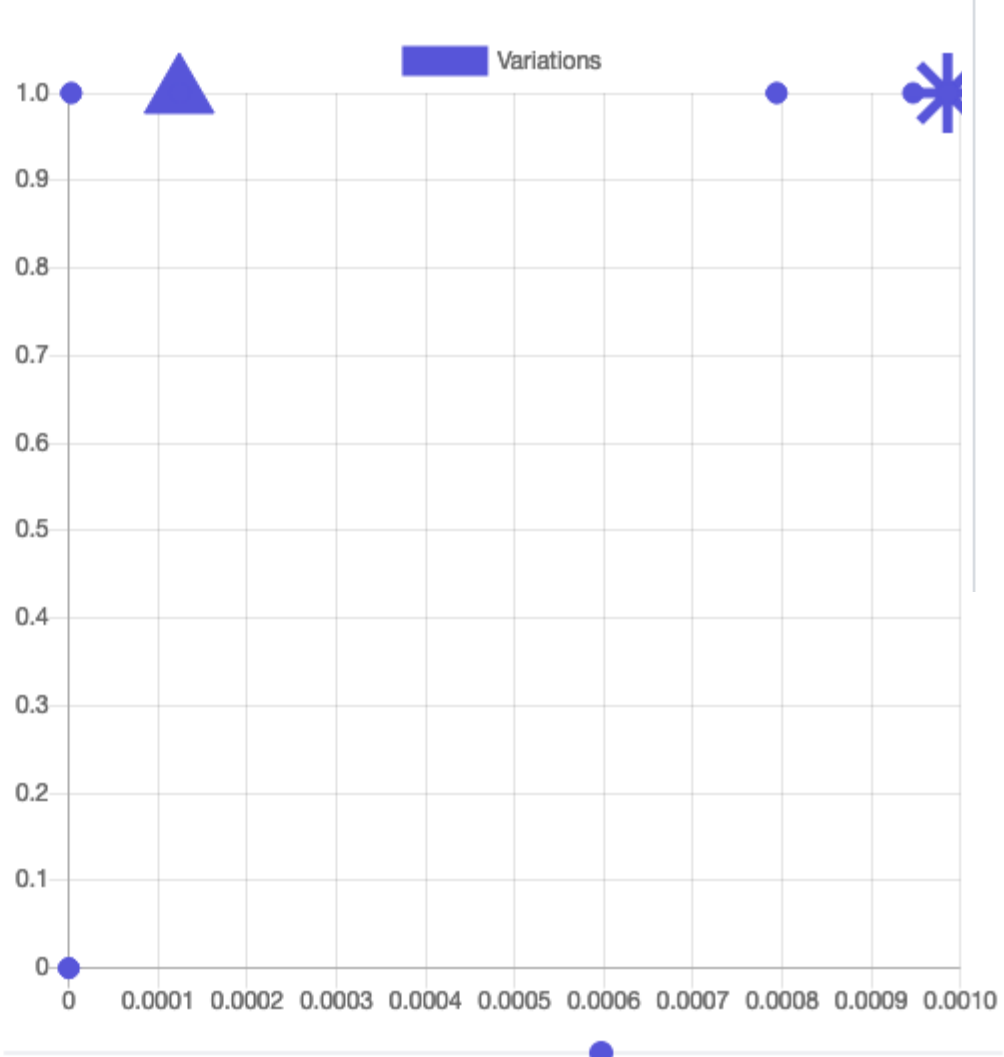
corous layer[Title/Abstract] OR
OR imaging procedure[All Fields]]

Continue Exploring

QueryVis >

Results >

- Number of variations: 14
- Type of transformation of this variation: Original
- F1: 0.0019696117051209907
- Precision: 0.0009857766511758908
- Recall: 1
- # Retrieved: 7101



Query Formulation

SET OF RELEVANT STUDY PMIDS

Relevant Set of Study PMIDs * : (Separate By Comma, Minimum 30)

18374892,18801517,15066293,10080594,19670452,18374503,15590163,14665356,18207665,19570619,15737905,20378156,11104883,10924979,15541117,18325680,14697417,18963536,9749478,11490252,19455340,10792092,20427255,18538495,20399462,19376564,12084197,19945997,15476513,17570425,11597800,20303100,20231039,10758314,14581420,17293235,19571899

Select Query Format * :

Pubmed

Clear All

→ Formulate

Query With MeSH

((obstruction[Title/Abstract] OR radioactive[Title/Abstract] OR ss[Title/Abstract] OR stool[Title/Abstract] OR aware[Title/Abstract] OR invasion[Title/Abstract] OR voiding[Title/Abstract] OR fecal[Title/Abstract] OR lrp[Title/Abstract] OR conference[Title/Abstract] OR undetectable[Title/Abstract] OR seminal[Title/Abstract] OR rose[Title/Abstract] OR clinicians[Title/Abstract] OR initiation[Title/Abstract] OR detectable[Title/Abstract] OR transient[Title/Abstract] OR anticipated[Title/Abstract] OR minor[Title/Abstract] OR fox[Title/Abstract] OR cohorts[Title/Abstract] OR prostatic[Title/Abstract] OR

Query Without MeSH

((obstruction[Title/Abstract] OR radioactive[Title/Abstract] OR ss[Title/Abstract] OR stool[Title/Abstract] OR aware[Title/Abstract] OR invasion[Title/Abstract] OR voiding[Title/Abstract] OR fecal[Title/Abstract] OR lrp[Title/Abstract] OR conference[Title/Abstract] OR undetectable[Title/Abstract] OR seminal[Title/Abstract] OR rose[Title/Abstract] OR clinicians[Title/Abstract] OR initiation[Title/Abstract] OR detectable[Title/Abstract] OR transient[Title/Abstract] OR anticipated[Title/Abstract] OR minor[Title/Abstract] OR fox[Title/Abstract] OR cohorts[Title/Abstract] OR prostatic[Title/Abstract] OR

Auto Doc

GENERAL INFORMATION

Years Covered By Search * : ?

1990 - 2000

Query Author (Optional) : ?

Harrisen Scells

Search Strategy Summary * : ?

Use the query to search for relevant studies.

Language Restrictions * : ?

English only.

DATABASE INFORMATION

Date Search Was Run * : ?



Take Aways (1)

- **Query formulation** is key to systematic review success — and there are big **opportunities for automation** and **improvements** that will have a downstream effect on systematic review effectiveness

Take Aways (2)

- Automatic formulation and refinement can be very successful, but yet **many open challenges**
 - Effective integration of MeSH terms (see also Wang, Scells, Koopman, Zuccon, Int.Sys. Appl. 2022)
 - LLMs very promising, but:
 - Variance among prompts — **formulating good prompts** is essential: but what makes a good prompt?
 - **Stochastic nature of LLMs** can largely affect effectiveness: if multiple generations are performed with ChatGPT for the same prompt, different queries are generated
 - Possibility to **specialise LLMs** for the domain
 - LLMs appear largely effective for “simple” tasks, more challenges for complex tasks: need for **specific training/instructions**

Take Aways (3)

- Information specialists have **limited or no insight into the effectiveness of generated queries** (especially if these “look” ok)
- Unable to identify high-yield queries (even among those for same prompt)
- Opportunity for computational methods for predicting query effectiveness, especially b/w variations