Introduction to automated screening techniques

Oct 1, 9:00 - 10:30

Cathleen PETIT et Devi VEYTIA
Postdoctoral researchers

FRB FONDATION POUR LA RECHERCHE SUR LA BIODIVERSITÉ | CESAB CENTRE DE SYNTHÈSE ET D'ANALYSE SUR LA BIODIVERSITÉ

The main stages of a review

- **Step 1: Defining the question**
  - 1.1 Formulating the question
  - 1.2 Conducting a pre-scoping exercise
  - 1.3 Building a test list of reference articles
  - 1.4 Identifying inclusion / exclusion criteria
  - 1.5 Refining and finalizing the question
- **Step 2 : Executing the literature search**
  - 2.1 Identifying relevant bibliographic sources
  - 2.2 Identifying search terms
  - 2.3 Refining search string against a Testlist
  - 2.4 Locating grey literature
  - 2.5 Choosing mangement software
  - 2.6 Managing duplicates
- **Step 3: Screening the literature**
  - 3.1 Title screening
  - 3.2 Abstract screening
  - 3.3 Searching for PDFs
  - 3.4 Full-text screening
- **Step 4 : Evaluating the retained articles**
  - *For systematic maps, this is not mandatory. May go to 5.1*
  - 4.1 Conducting a criticial analysis
- **Step 5: Synthesizing the meta-data**
  - 5.1 Extracting meta-data
  - 5.2 Extracting quantitative data for meta-anlayses

Source: Joseph Langridge, after Foo et al, 2021

Outline

1. What is the need?
2. Text mining and machine learning models
3. Applications of machine learning models for screening
   a. Helps in the process : Truncate screening ; relevance ranking
   b. Update existing reviews
   c. Fully automated screening
   d. Active learning
4. TD: Case study of a sample protocol workflow)
      i. Relevance ranking tools (AbstractR, Colandr, etc.)
      ii. Use of Abstrackr using example from Ouédraogo et al. (2021)

1. What is the need ?

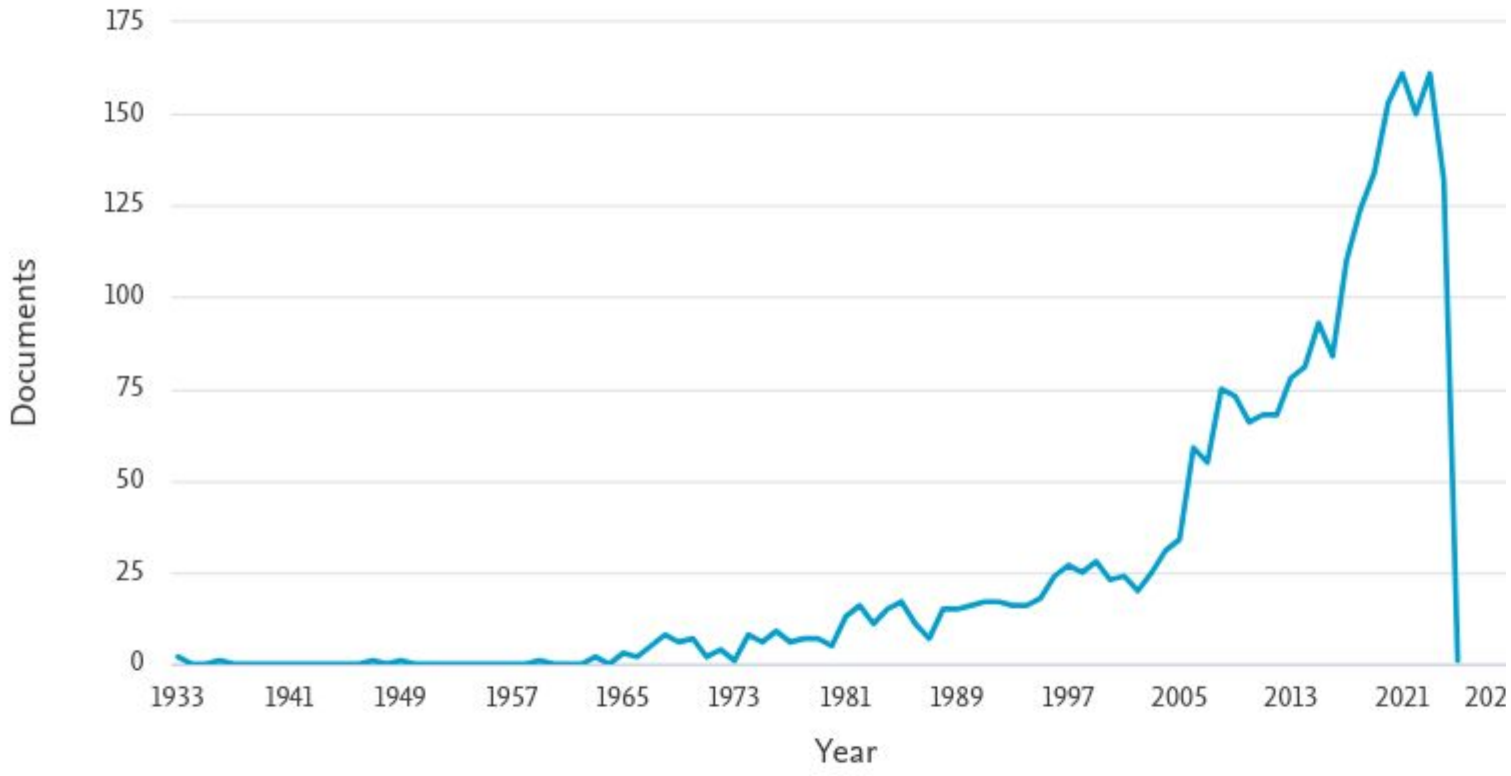
- More and more publication to deal with

  Reply to a narrow topic:

  How artificial light affects chiropterans ?

Scopus®

2501 documents found



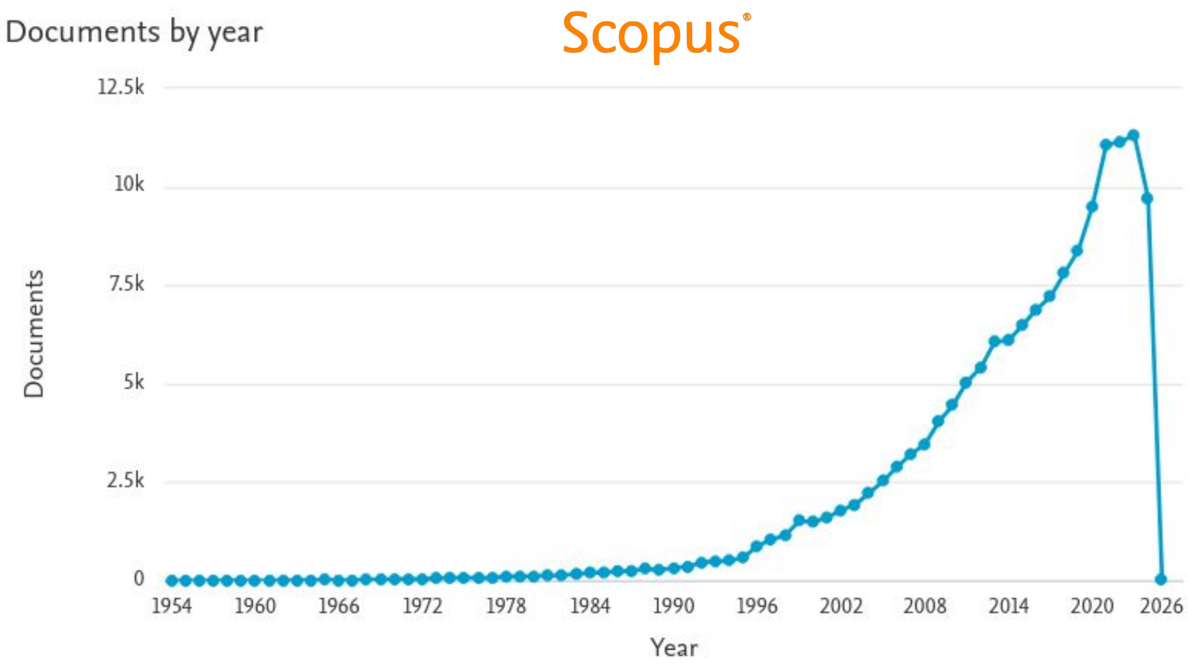
1. What is the need ?

- More and more publication to deal with

    Reply to a broad topic :

    How is considered the human presence, activities and infrastructures into studies of human pressures on biodiversity?

152,178 documents found in field related to ecology



1. What is the need ?

- More and more publication to deal with
- Repetitive time consuming tasks

1. What is the need ?

- More and more publications to deal with
- Repetitive time consuming tasks
- New tools and machine learning models can be used

Summary FOR PDF :

Systematic evidence syntheses are time-intensive. Resource constraints are often a limiting factor in determining the scope of the research question.

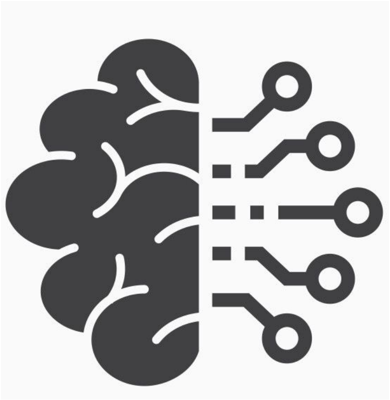
Even when search strategies are optimized for sensitivity, review teams are still faced with spending large amounts of time sifting through a large proportion of irrelevant studies.

This underlines the trade-off between:

the exponential increase in scientific publications, which places a limiting pressure driving synthesis questions to become narrower, and

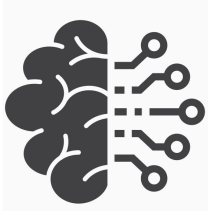
the needs/motivations of researchers and decision-makers who often require syntheses of broad-scale topical questions within narrow time frames.

Recent developments in text mining and machine learning techniques can help to expedite certain time-consuming steps of the evidence synthesis process, reducing the manual effort required and making more large-scale and timely evidence synthese possible



1. What is the need ?

- More and more publications to deal with
- Repetitive time consuming tasks
- New tools and machine learning models can be used



- **Step 1: Defining the question**
  - 1.1 Formulating the question
  - 1.2 Conducting a pre-scoping exercise
  - 1.3 Building a test list of reference articles
  - 1.4 Identifying inclusion / exclusion criteria
  - 1.5 Refining and finalizing the question
- **Step 2 : Executing the literature search**
  - 2.1 Identifying relevant bibliographic sources
  - 2.2 Identifying search terms
  - 2.3 Refining search string against a Testlist
  - 2.4 Locating grey literature
  - 2.5 Choosing mangement software
  - 2.6 Managing duplicates
- **Step 3: Screening the literature**
  - 3.1 Title screening
  - 3.2 Abstract screening
  - 3.3 Searching for PDFs
  - 3.4 Full-text screening
- **Step 4 : Evaluating the retained articles**
  - *For systematic maps, this is not mandatory. May go to 5.1*
  - 4.1 Conducting a criticial analysis
- **Step 5: Synthesizing the meta-data**
  - 5.1 Extracting meta-data
  - 5.2 Extracting quantitative data for meta-anlayses

1. What is the need ?

AI Research Question Prompt Generator

Unleash your academic potential with the click of a button! Our Research Question Prompt Generator is your secret weapon to effortlessly crafting thought-provoking questions that'll make your studies and papers stand out. Try it now and let inspiration strike!

taskade.com

- More and more publications to deal with
- Repetitive time consuming tasks
- New tools and machine learning models can be used

Elicit.com

Analyze research papers at superhuman speed

Automate time-consuming research tasks like summarizing papers, extracting data, and synthesizing your findings.

Step 1: Defining the question
- 1.1 Formulating the question
- 1.2 Conducting a pre-scoping exercise
- 1.3 Building a test list of reference articles
- 1.4 Identifying inclusion / exclusion criteria
- 1.5 Refining and finalizing the question

Step 2 : Executing the literature search
- 2.1 Identifying relevant bibliographic sources
- 2.2 Identifying search terms
- 2.3 Refining search string against a Testlist
- 2.4 Locating grey literature
- 2.5 Choosing mangement software
- 2.6 Managing duplicates

Step 3: Screening the literature
- 3.1 Title screening
- 3.2 Abstract screening
- 3.3 Searching for PDFs
- 3.4 Full-text screening

Step 4 : Evaluating the retained articles
- *For systematic maps, this is not mandatory. May go to 5.1*
- 4.1 Conducting a criticial analysis

Step 5: Synthesizing the meta-data
- 5.1 Extracting meta-data
- 5.2 Extracting quantitative data for meta-anlayses

AI Search Engine for Research

Find & understand the best science, faster.

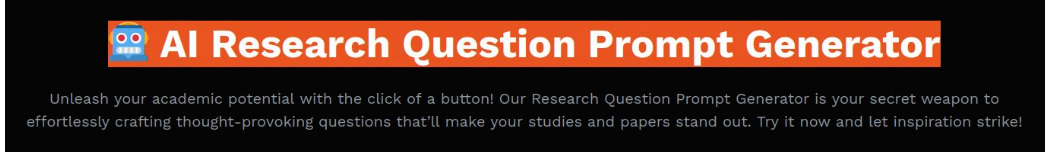
Ask the research...

Does exercise improve cognition?

Can cash transfers reduce poverty?

https://consensus.app/

1. What is the need ?



taskade.com

- More and more publications to deal with
- Repetitive time consuming tasks
- New tools and machine learning models can be used

Step 1: Defining the question
- 1.1 Formulating the question
- 1.2 Conducting a pre-scoping exercise
- 1.3 Building a test list of reference articles
- 1.4 Identifying inclusion / exclusion criteria
- 1.5 Refining and finalizing the question

Step 2 : Executing the literature search
- 2.1 Identifying relevant bibliographic sources
- 2.2 Identifying search terms
- 2.3 Refining search string against a Testlist
- 2.4 Locating grey literature
- 2.5 Choosing mangement software
- 2.6 Managing duplicates

Step 3: Screening the literature
- 3.1 Title screening
- 3.2 Abstract screening
- 3.3 Searching for PDFs
- 3.4 Full-text screening

Step 4 : Evaluating the retained articles
- *For systematic maps, this is not mandatory. May go to 5.1*
- 4.1 Conducting a criticial analysis

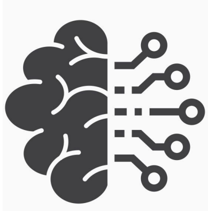
Step 5: Synthesizing the meta-data
- 5.1 Extracting meta-data
- 5.2 Extracting quantitative data for meta-anlayses

colandr

https://consensus.app/

1. What is the need ?

- More and more publications to deal with
- Repetitive time consuming tasks
- New tools and machine learning models can be used



Step 1: Defining the question
- 1.1 Formulating the question
- 1.2 Conducting a pre-scoping exercise
- 1.3 Building a test list of reference articles
- 1.4 Identifying inclusion / exclusion criteria
- 1.5 Refining and finalizing the question

Step 2 : Executing the literature search
- 2.1 Identifying relevant bibliographic sources
- 2.2 Identifying search terms
- 2.3 Refining search string against a Testlist
- 2.4 Locating grey literature
- 2.5 Choosing mangement software
- 2.6 Managing duplicates

Step 3: Screening the literature
- 3.1 Title screening
- 3.2 Abstract screening
- 3.3 Searching for PDFs
- 3.4 Full-text screening

Step 4 : Evaluating the retained articles
- *For systematic maps, this is not mandatory. May go to 5.1*
- 4.1 Conducting a criticial analysis

Step 5: Synthesizing the meta-data
- 5.1 Extracting meta-data
- 5.2 Extracting quantitative data for meta-anlayses

2. Text mining and machine learning models

Introduction & key terms

Text mining retrieves and distils information from unstructured text by:

1. Determining relevant text	Screening
2. Describing the characteristics of included text	Metadata coding/mapping
3. Data mining: identifying patterns within relevant texts	Synthesis/analysis

These steps map well to the systematic protocol, and can be used to assist human reviewers in these steps.

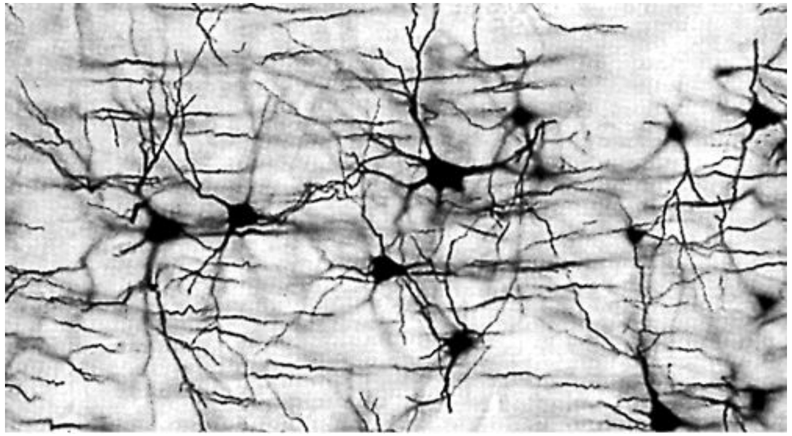
Various methods exist (e.g. automatic term recognition, topic clustering, etc.), but we will focus on how machine learning models can be used to predict document relevance and expedite literature screening

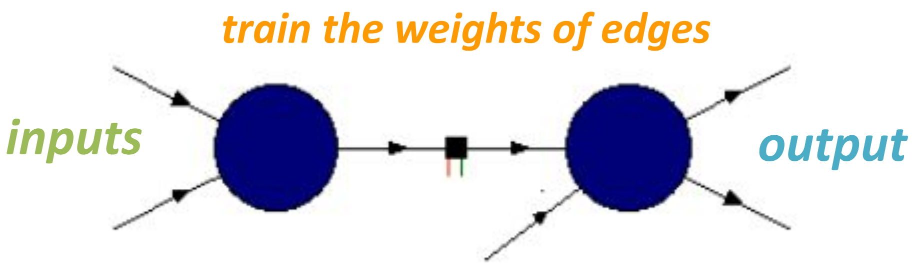
2. Text mining and machine learning models:
Development and recent advancements

WWII – Alan Turing's B-type unorganised machine

- The human brain is a network composed of connected neurons. We learn from different experiences (***inputs***) which strengthen or weaken connections between different neurons

- Turing hypothesized this could be simulated by a machine. The strength (***weights***) of the connections (***edges***) between artificial neurons can be updated from ***inputs*** ⟶ The machine can be ***trained*** to perform specific tasks (***output***)





But, artificial neural network models didn't perform well at predictive tasks. Until…

Source: AlanTuring.net

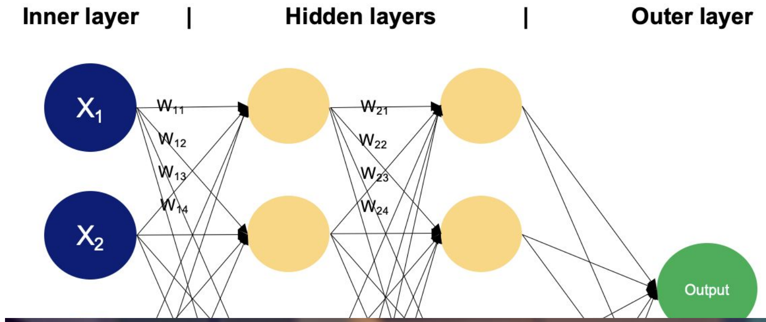
2. Text mining and machine learning models:
Development and recent advancements

<u>3 Key advancements → The AI revolution</u>

1. Big sources of data online → needed to train models

2. Theoretical advancements in neural network models architecture → stacking multiple layers of models allows information to be passed through the different layers improving performance

3. Advancements in Graphical Processing Units (developed to get increasingly high-res FPS video games) allowed for the parallel processing of these multi-layered models → allows for multiple model calculations to be done simultaneously needed for multi-level neural networks

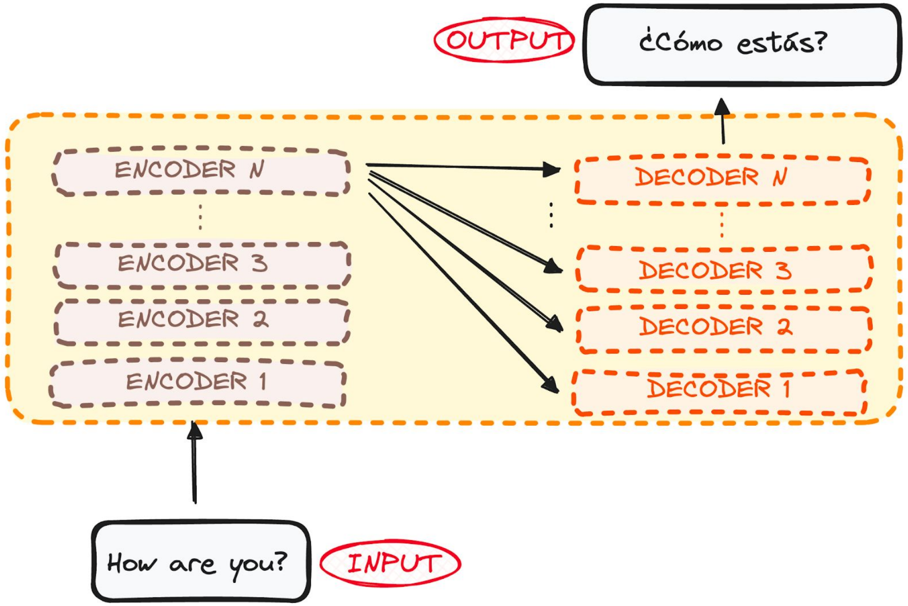


Source: [AI in medicine, Podcast interview with Isacc Kohane, M.D., PhD]

2. Text mining and machine learning models:
Development and recent advancements

Transformer models

- *Attention Is All You Need* (published by Google in 2017) presented the transformer model.
- Doesn't just look at the frequencies of which words that co-occur together, it also accounts for the *context* → the position of words in a sentence/ordering dramatically improved model performance
- E.g. Google's BERT large language model (trained from text data in wikipedia) that is used in their search engine



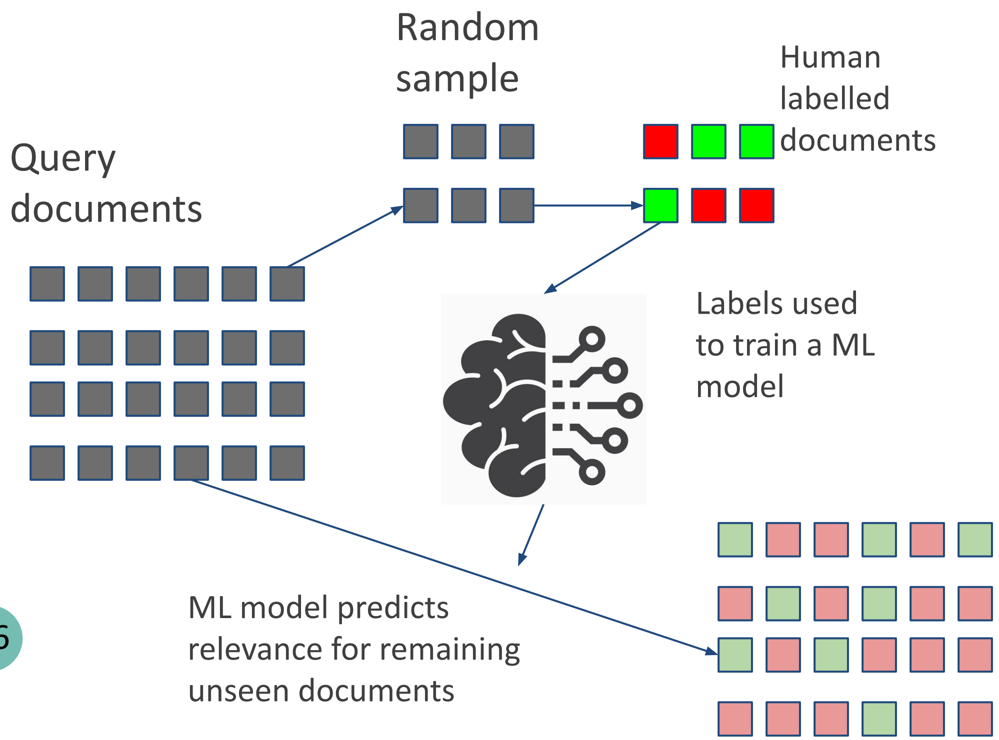
For more information:

https://www.datacamp.com/tutorial/how-transformers-work

2. Text mining and machine learning models

Using machine learning models to predict document relevance

Transformer models (pre-trained on large amounts of text data) can then be fine-tuned to perform specific natural language processing tasks. For reviews, we can use them to predict whether a document is relevant (inclusion) or not (exclusion)

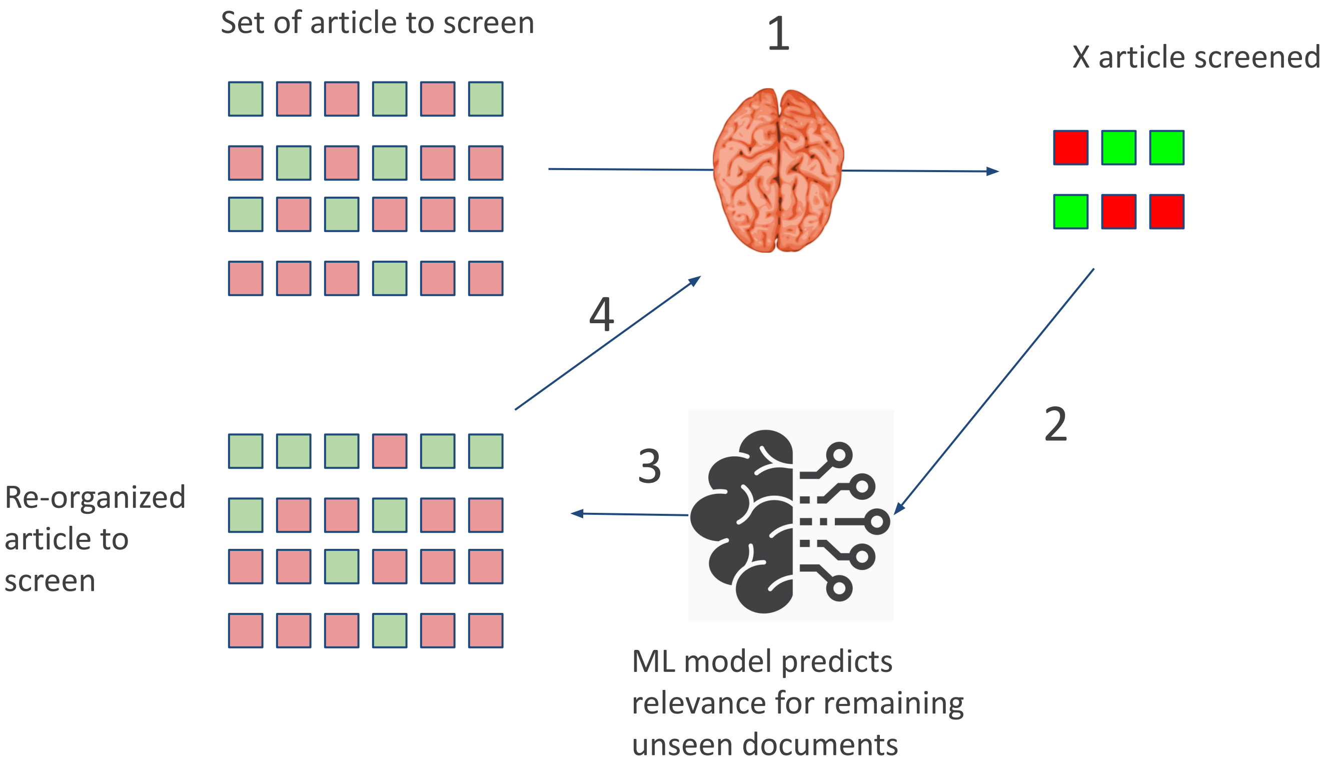


This involves providing a sample of publication text with reviewer screening decisions, and using this to train the model to predict relevance across unseen publications.

Adapted from ED Fig. 1, Callaghan et al. (2021)

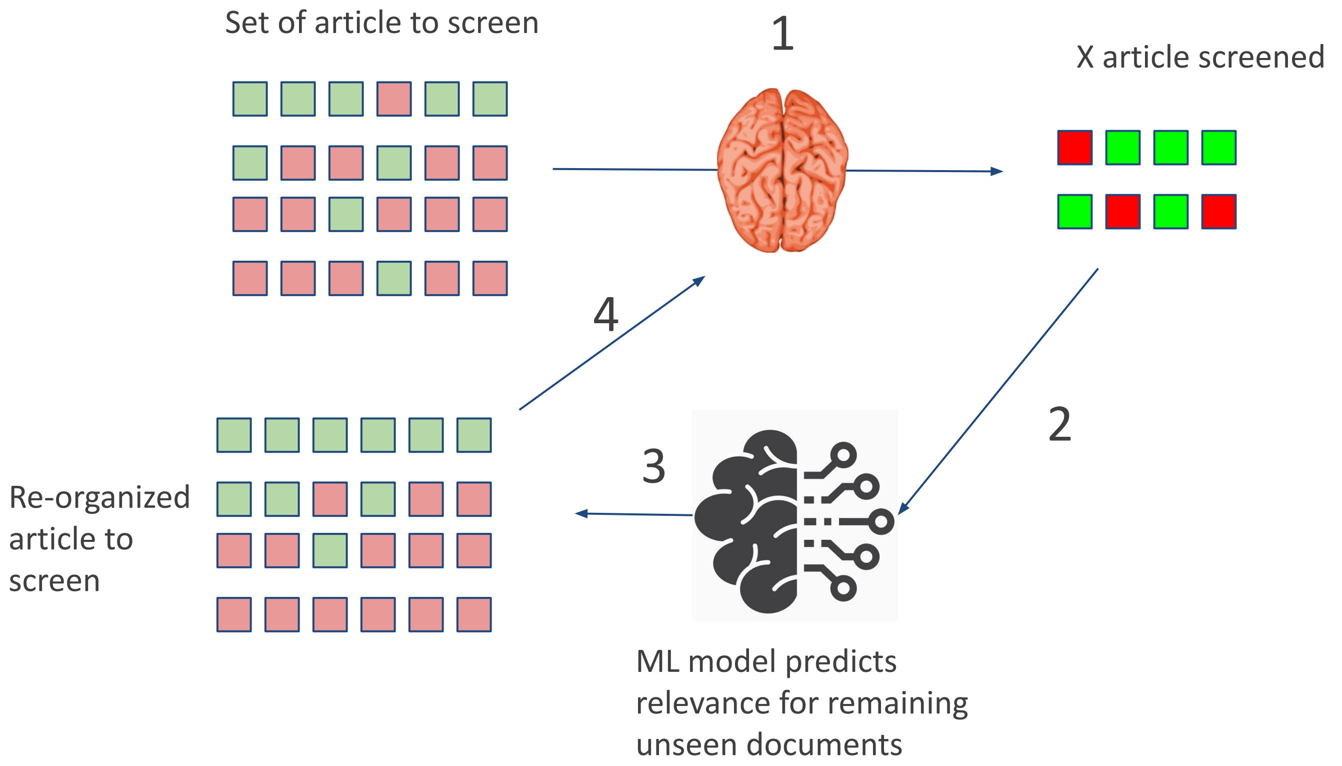
3. Applications
a) Relevance ranking

- Model that predicts relevance ranking
  (e.g. method Cohen et al 2009 ; e.g. use. Apriyani et al., 2024)



3. Applications
a) Relevance ranking

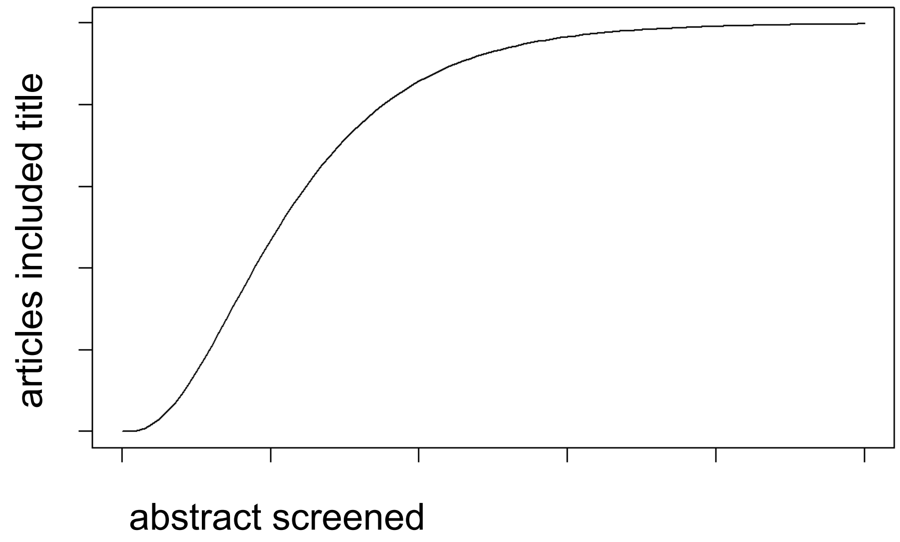
- Model that predicts relevance ranking
  (e.g. method [Cohen et al 2009]() ; e.g. use. [Apriyani et al., 2024]())



3. Applications
a) Relevance ranking

- As a reviewer applies screening decisions to article text, these data are used to update a model which predicts relevance for the remaining unseen documents. (e.g. method Cohen et al 2009 ; e.g. use. Apriyani et al., 2024)

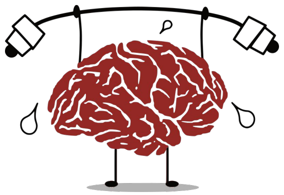
- Truncate screening : stop the screening effort after a certain inclusion/exclusion ratio.
  - threshold e.g. 5% in Cheng et al., 2023
  - Asymptote e.g. Rubenstein et al ., 2023



3. Applications
a) Relevance ranking

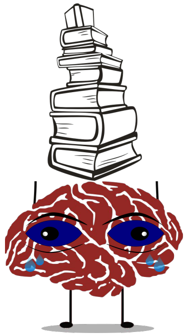
- As a reviewer applies screening decisions to article text, these data are used to update a model which predicts relevance for the remaining unseen documents. (e.g. Apriyani et al., 2024)

- Truncate screening : stop the screening effort after a certain inclusion/exclusion ratio.
  - threshold e.g. 5% in Cheng et al., 2023
  - Asymptote e.g. Rubenstein et al ., 2023



Strengths:

- Can significantly reduce review effort



Limitations:

- The effectiveness depends on the representativeness of all the articles that have been screened
- Size limitation on the scope of the review
- All relevant articles need to be manually screened.

3. Applications
a) Relevance ranking

Summary for PDF :

As a reviewer applies screening decisions to article text, these data are used to update a model which predicts relevance for the remaining unseen documents. Thus, rather than viewing the documents in no particular order, those most similar to the studies already included are moved to the top of the list. This increases the probability that the next document viewed will be included in the review. A reviewer need only screen until their inclusion rate drops below a certain threshold (e.g. 5% in Cheng et al., 2023) rather than screening all the documents.

Strengths: Can significantly reduce review effort

Limitations: Effectiveness is dependent on the studies already identified being representative of those remaining, there is still a size limitation on the scope of the review, as all relevant articles still need to be manually screened.

3. Applications

b) Updates of existing reviews

- The articles included and excluded from existing reviews can be used to train a model, which can then be applied to screen new search results (e.g. Cohen et al 2005 , Cohen et al 2009).

- Mainly use in medicine reviews

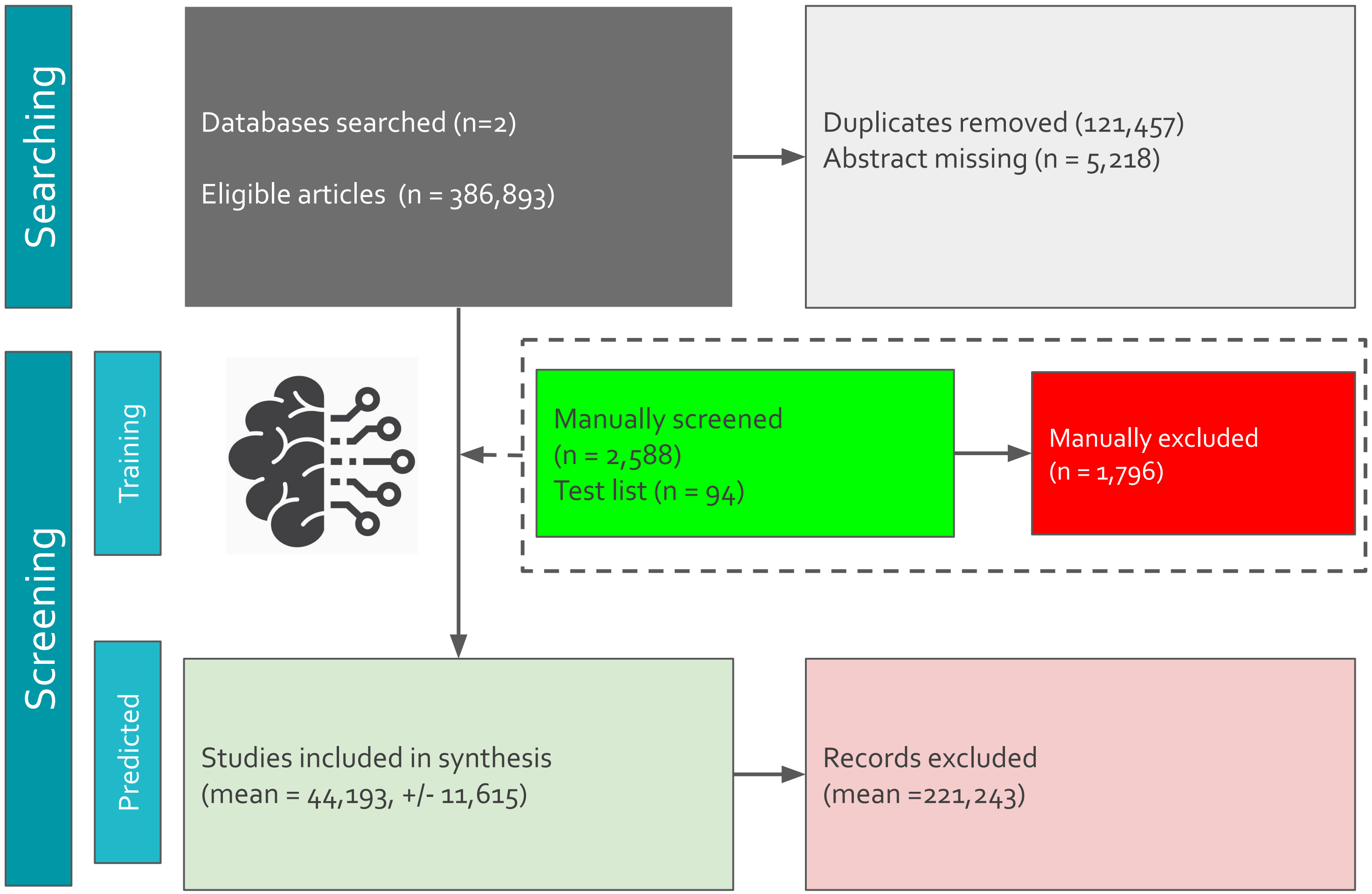
Fusidic acid in dermatology: an **updated review**
H Schöfer, L Simonsen - European Journal of Dermatology, 2010 - jle.com
Studies on the clinical efficacy of fusidic acid in skin and soft-tissue infections (SSTIs), notably those due to Staphylococcus aureus, are reviewed. Oral fusidic acid (tablets dosed at 250 …
☆ Enregistrer 99 Citer Cité 109 fois Autres articles Les 7 versions

[HTML] The hallucinogenic world of tryptamines: an **updated review**
AM Araújo, F Carvalho, ML Bastos… - Archives of …, 2015 - Springer
… This **review** provides a comprehensive update on tryptamine hallucinogens, concerning their historical background, prevalence, patterns of use and legal status, chemistry, toxicokinetics…
☆ Enregistrer 99 Citer Cité 294 fois Autres articles Les 11 versions

[HTML] Keratoconus: An **updated review**
J Santodomingo-Rubido, G Carracedo, A Suzaki… - Contact Lens and …, 2022 - Elsevier
… This article provides an **updated review** on the definition, epidemiology, histopathology, aetiology and pathogenesis, clinical features, detection, classification, and management and …
☆ Enregistrer 99 Citer Cité 403 fois Autres articles Les 15 versions

Dexmedetomidine: an **updated review**
AT Gerlach, JF Dasta - Annals of Pharmacotherapy, 2007 - journals.sagepub.com
… To **review** recent literature on the safety and efficacy of … Sedation for adult ICU patients: A narrative **review** including … **Review** Article: Dexmedetomidine: Does it Have Potential in …
☆ Enregistrer 99 Citer Cité 559 fois Autres articles Les 8 versions

3. Applications
c) Fully automated screening

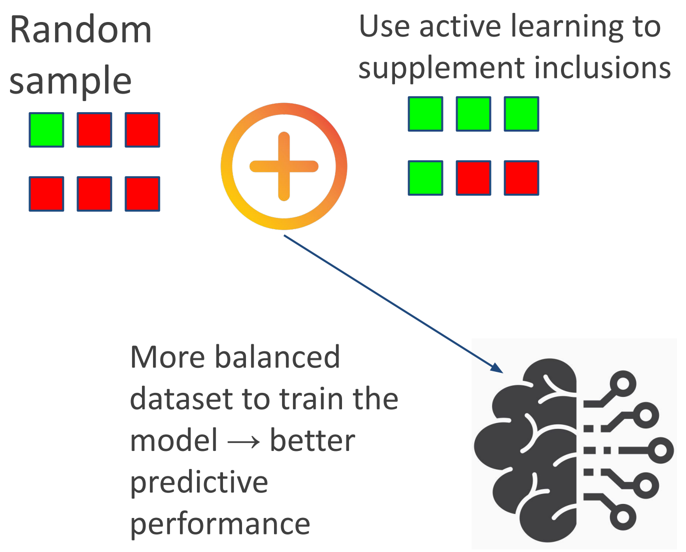


Example from Veytia et al. (2024) *In Prep*

3. Applications

d) Active learning →fully automated screening

Problem: In the example of the previous application, search results are often unbalanced with high numbers of irrelevant documents as opposed to relevant. Thus a random sample to train the machine learning model may have very few examples of inclusions. This is a problem as the model needs to train on data that provides information on *what to include* as well as *what to exclude*, and therefore needs a balanced sample



Active learning helps address this by using relevance ranking to increase the likelihood of seeing relevant documents, to provide supplementary inclusions to the random sample

3. Applications

- abstrackR is a web-application that makes citation-screening process of systematic reviews "easier" (http://abstrackr.cebm.brown.edu/)

- Colandr is an open access machine-learning assisted online platform for conducting reviews and syntheses of text-based evidence (e.g. articles, documents, etc...). You must register via this link : (https://www.colandrapp.com/signin) & (https://scheng.shinyapps.io/colandr_stats/)

- Non-open access : https://www.rayyan.ai/ OR https://www.laser.ai/

TD: Using relevance ranking to screen articles

(http://abstrackr.cebm.brown.edu/)

From: Ouédraogo et al. (2021, doi: 10.1186/s13750-021-00237-9)

Table 1 Eligibility criteria

Include	Exclude
Population:	
- All tropical reef-building coral species (hermatypic scleractinian species, *Millepora* species, *Heliopora* species and *Tubipora* species) living in the shallow and the mesophotic zones	- Cold-water or deep-water corals - Ahermatypic corals - Free-living dinoflagellates (not as symbionts in corals) - Studies conducted in coral reefs but not about corals (e.g. about coral reef fishes)
Exposure:	
- All natural (e.g. nitrate), geogenic (e.g. nickel) and synthetic (e.g. diuron) chemicals coming from human activities - Studies assessing the impact of human activities (e.g. river discharge, distance to a dump or to an industrial effluent source, tourism) on corals without reference to a chemical	- Studies assessing the impact of chemicals coming from natural sources (e.g. nutrients from guano) - Studies assessing the impact of organic carbon - Studies assessing the impact of sedimentation per se or impact of physical disturbances on coral - Marine debris, macro-plastics
Comparator:	
- Studies comparing population exposed to chemicals and population unexposed to chemicals - Studies comparing population exposed to chemicals and population prior to exposure to chemicals - Studies comparing population exposed to a range of concentrations/levels of chemicals	
Outcome:	
- All outcomes related to tropical reef-building corals, from molecular to community level - Studies reporting evidence of ingestion, concentration or accumulation/uptake of chemicals in the population studied without reporting health consequences - Studies assessing impacts on coral microbiome/symbionts	
Language:	
All articles written in English or French (in case a title or an abstract could not be found in English or French, it was directly screened on full-text)	
Type of document:	
Journal article, book chapter, report, conference proceeding, PhD or MSc thesis	Presentation, editorial material, letter or news item, conference or meeting abstract, poster
Type of content:	
In-situ or ex-situ studies	Reviews and meta-analyses, modelling studies without experimental data

Include	Exclude
Population:	
- All tropical reef-building coral species (hermatypic scleractinian species, Millepora species, Heliopora species and Tubipora species) living in the shallow and the mesophotic zones	- Cold-water or deep-water corals - Ahermatypic corals - Free-living dinoflagellates (not as symbionts in corals) - Studies conducted in coral reefs but not about corals (e.g. about coral reef fishes)
Exposure	
- All natural (e.g. nitrate), geogenic (e.g. nickel) and synthetic (e.g. diuron) chemicals coming from human activities - Studies assessing the impact of human activities (e.g. river discharge, distance to a dump or to an industrial effluent source, tourism) on corals without reference to a chemical	- Studies assessing the impact of chemicals coming from natural sources (e.g. nutrients from guano) - Studies assessing the impact of organic carbon - Studies assessing the impact of sedimentation per se or impact of physical disturbances on coral - Marine debris, macro-plastics
Comparator	
- Studies comparing population exposed to chemicals and population unexposed to chemicals - Studies comparing population exposed to chemicals and population due to exposure to chemicals - Studies comparing population exposed to a range of concentrations/levels of chemicals	
Outcome	
- All outcomes related to tropical reef-building corals, from molecular to community level - Studies reporting evidence of ingestion, concentration or accumulation/uptake of chemicals in the population studied without reporting health consequences - Studies assessing impacts on coral microbiome/symbionts	
Language	
All articles written in English or French (in case a title or an abstract could not be found in English or French, it was directly screened on full-text)	
Type of document	
Journal article, book chapter, report, conference proceeding, PhD or MSc thesis	Presentation, editorial material, letter or news item, conference or meeting abstract, poster
Type of content	
In-situ or ex-situ studies	Reviews and meta-analyses, modelling studies without experimental data

FRB
FONDATION
POUR LA RECHERCHE
SUR LA BIODIVERSITÉ

CESAB
CENTRE DE SYNTHÈSE ET D'ANALYSE
SUR LA BIODIVERSITÉ

Machine Learning applications for metadata extraction et categorization

Oct 1, 14:00 - 15:00

Devi VEYTIA
Postdoctoral researcher

The main stages of a review

- **Step 1: Defining the question**
  - 1.1 Formulating the question
  - 1.2 Conducting a pre-scoping exercise
  - 1.3 Building a test list of reference articles
  - 1.4 Identifying inclusion / exclusion criteria
  - 1.5 Refining and finalizing the question
- **Step 2 : Executing the literature search**
  - 2.1 Identifying relevant bibliographic sources
  - 2.2 Identifying search terms
  - 2.3 Refining search string against a Testlist
  - 2.4 Locating grey literature
  - 2.5 Choosing mangement software
  - 2.6 Managing duplicates
- **Step 3: Screening the literature**
  - 3.1 Title screening
  - 3.2 Abstract screening
  - 3.3 Searching for PDFs
  - 3.4 Full-text screening
- **Step 4 : Evaluating the retained articles**
  - *For systematic maps, this is not mandatory. May go to 5.1*
  - 4.1 Conducting a criticial analysis
- **Step 5: Synthesizing the meta-data**
  - 5.1 Extracting meta-data
  - 5.2 Extracting quantitative data for meta-anlayses

Source: Joseph Langridge, after Foo et al, 2021

Introduction

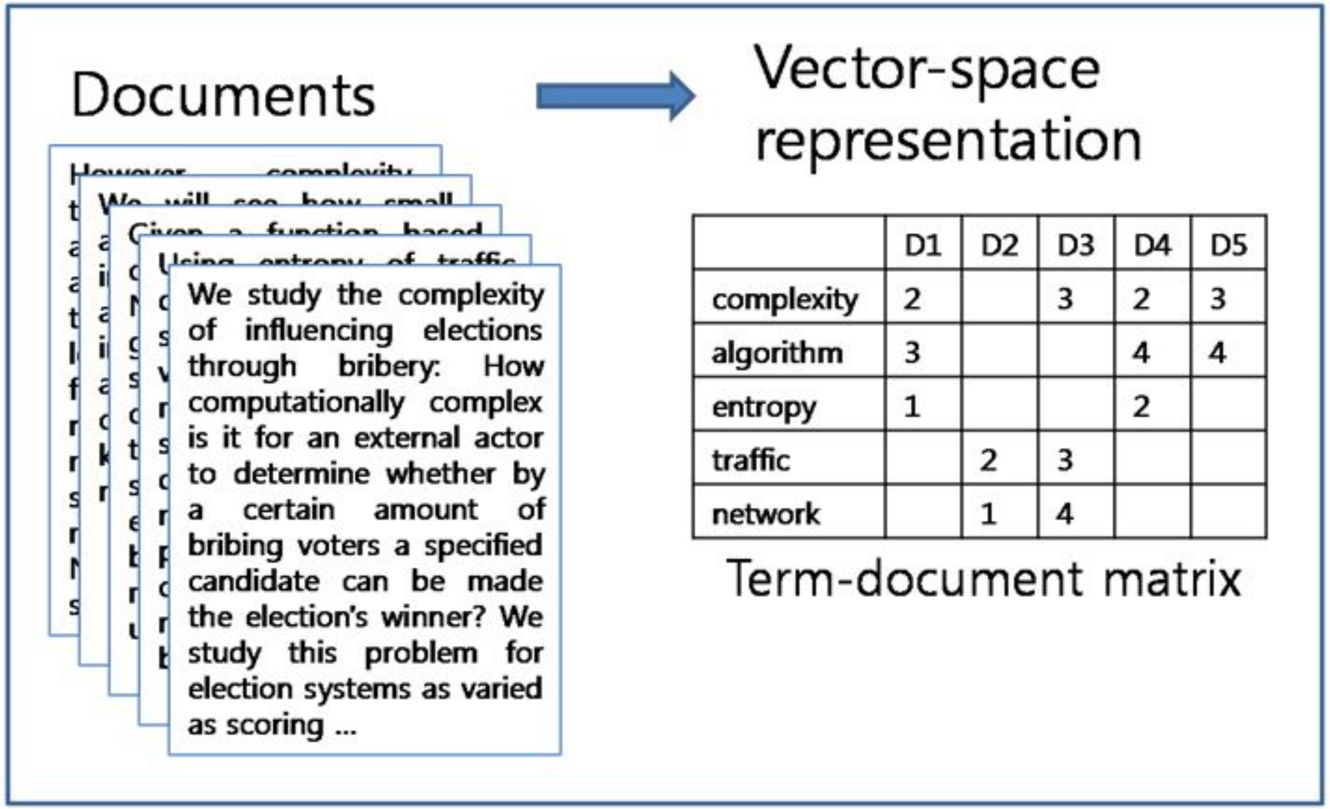
Text mining also has applications for describing the distribution and extent of literature with respect to metadata variables. Examples include:

1. Topic clustering (e.g. Stansfield et al. 2010, Callaghan et al. 2020)
2. Predictive labelling (e.g. Colandr)
3. Automatic classification (e.g. Callaghan et al. 2021, Veytia et al. 2023)

Topic clustering

Groups documents based on topic by:

1. Deconstructing text into a feature matrix of words and their frequencies of occurrence

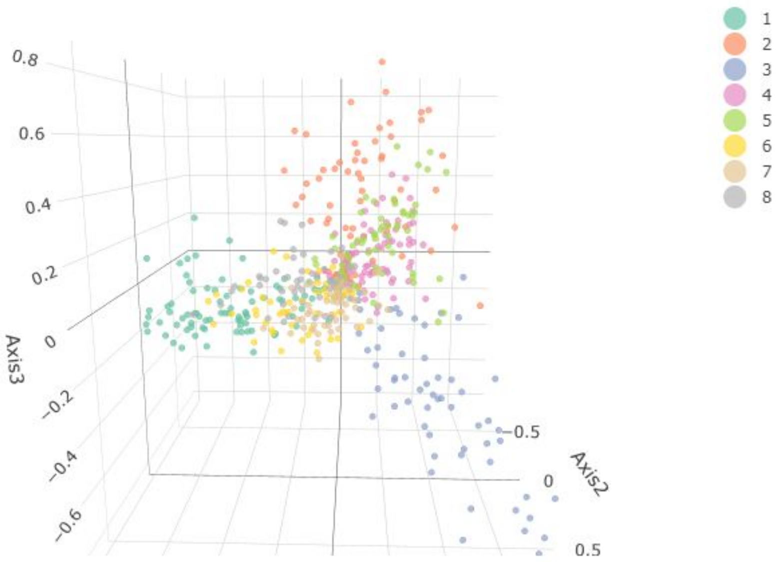


	D1	D2	D3	D4	D5
complexity	2		3	2	3
algorithm	3			4	4
entropy	1			2	
traffic		2	3		
network		1	4		

Topic clustering

Groups documents based on topic by:

1. Deconstructing text into a feature matrix of words and their frequencies of occurrence
2. Use these data to cluster documents with similar features



Each group is defined by several keywords that frequently occurred in the shared documents. These keywords can be used to infer the topic. E.g:

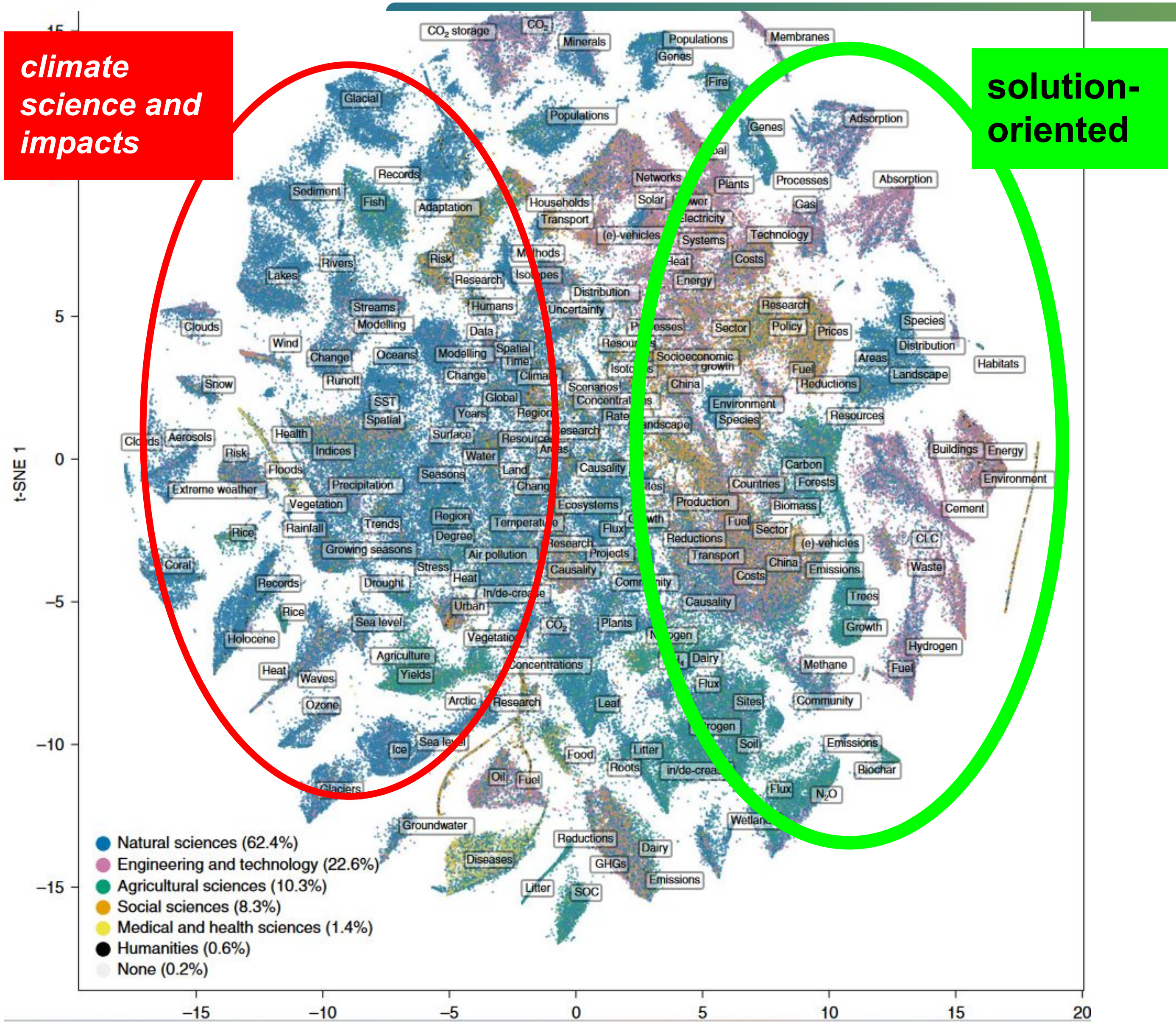
- Topic 1: manage, ecosystem, resource
- Topic 2: coral, reef, restore
- Topic 3: population, genetic, distributed

But these groupings were *not* very helpful to identifying relevant articles for our main research question, which was focused on separating articles according to specific intervention types

Advantages: Fast and easy – Useful for scoping the topics of studies before manual coding. Topics informed *by data* rather than reviewer bias

Disadvantages: The clusters generated automatically do not necessarily reflect the conceptual framework of the review

Topic clustering: Application



Callaghan et al. (2020, *NCC*)

- Dominance of the natural sciences in climate change research
- over-representation of social science and under-representation of technical solutions in literature cited in the IPCC compared with publication trends in WOS

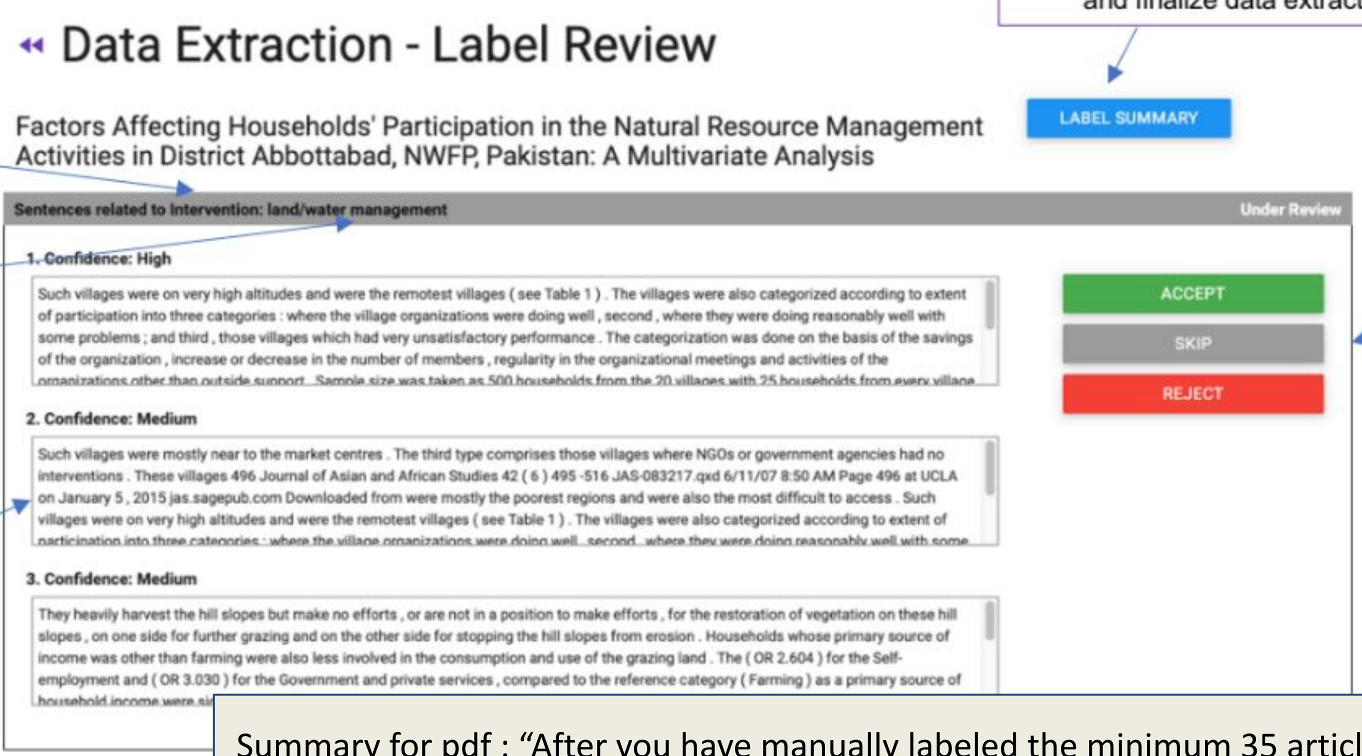
Predictive labelling

Provide suggestions to expedite data extraction (e.g. [Colandr])

Click to see PDF and manual labelling screen to enter other non-“select one” or “select many” fields and finalize data extraction

Data field

Suggested label



After reading sentences, user chooses to accept, skip, or reject the label

For each suggested label for a data field, colandr will display a set of sentences that it has high, medium, and low confidence are related to the suggested label.

Summary for pdf : “After you have manually labeled the minimum 35 articles to inform the machine learning and natural language processing model, colandr will begin to suggest labels for any data fields that are set to “select one” or “select many.” These labels will appear when you click for an article.”

Automatic classification

Overcomes the limitations of topic clustering by predicting relevance for user-defined labels

TD: Using DistilBERT for multi-label text classification
→ Work through google collab notebook of how a multilabel model can be trained and evaluated

<u>Learning outcomes:</u>

- You do not need to understand every line of code → focus on understanding the pipeline/steps that are involved in training and validating a machine learning model
- Apply this understanding to evaluate when machine learning models are fit for purpose – what are their strengths and limitations.
- This will help inform whether you want to use this approach in your own work, and critically evaluate other studies that use machine learning models.

Reflection

Do you think the model performed well?

What steps could we take to improve the model fit?

- Modelling and meta-analysis labels performed poorly compared to study and review. Could improve by adding more samples to poorly performing labels, and if necessary removing the poorly performing labels all together.

What limitations do you see in applying this approach?

- Will not be useful if the information needed to code the label is not included in the title/abstract text
- It is not straightforward to tell a priori how the model will perform
- Suitable for mapping but not suited for reviews – human oversight is still needed for quality assessment