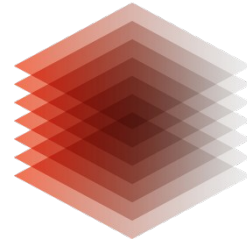

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

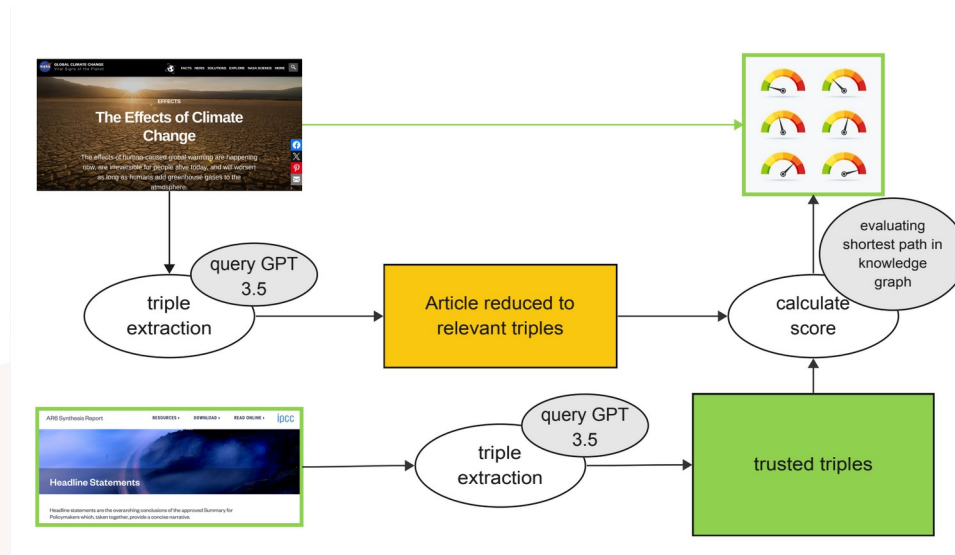


TIB

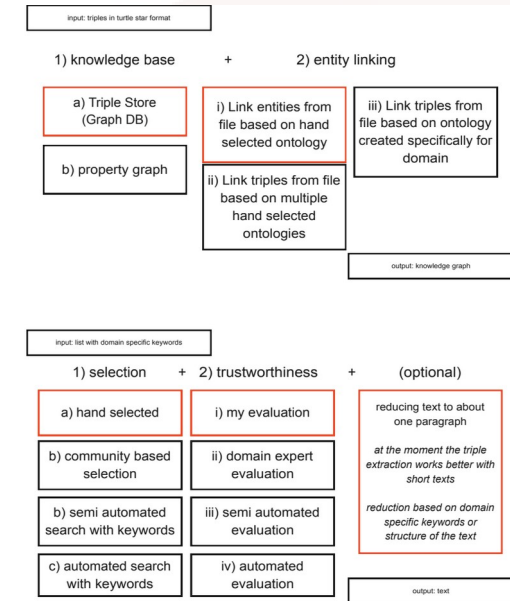
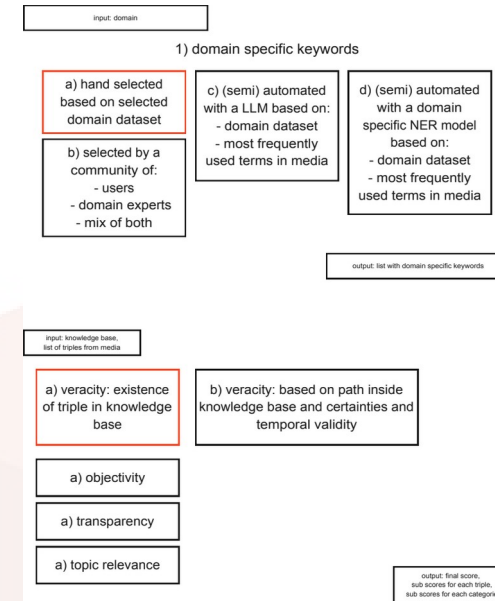
Master Thesis „Scientific Knowledge fit for society“ – Approach

Constantin Tremel,
01. February 2023
Interim Presentation

Agenda



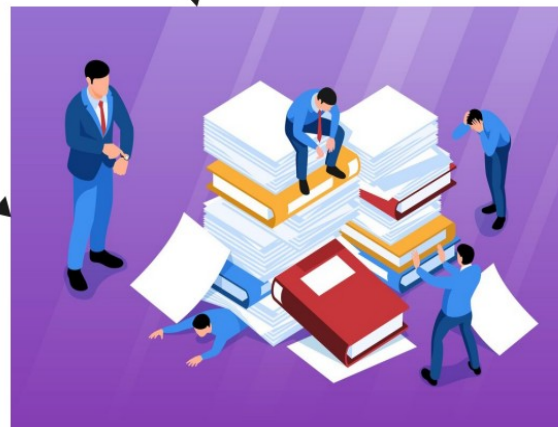
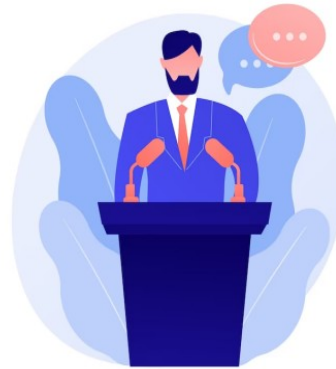
Approach – 5 min



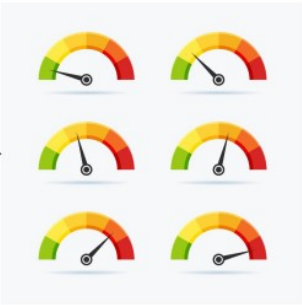
Additional material – 2 min

Questions & Comments

The problem



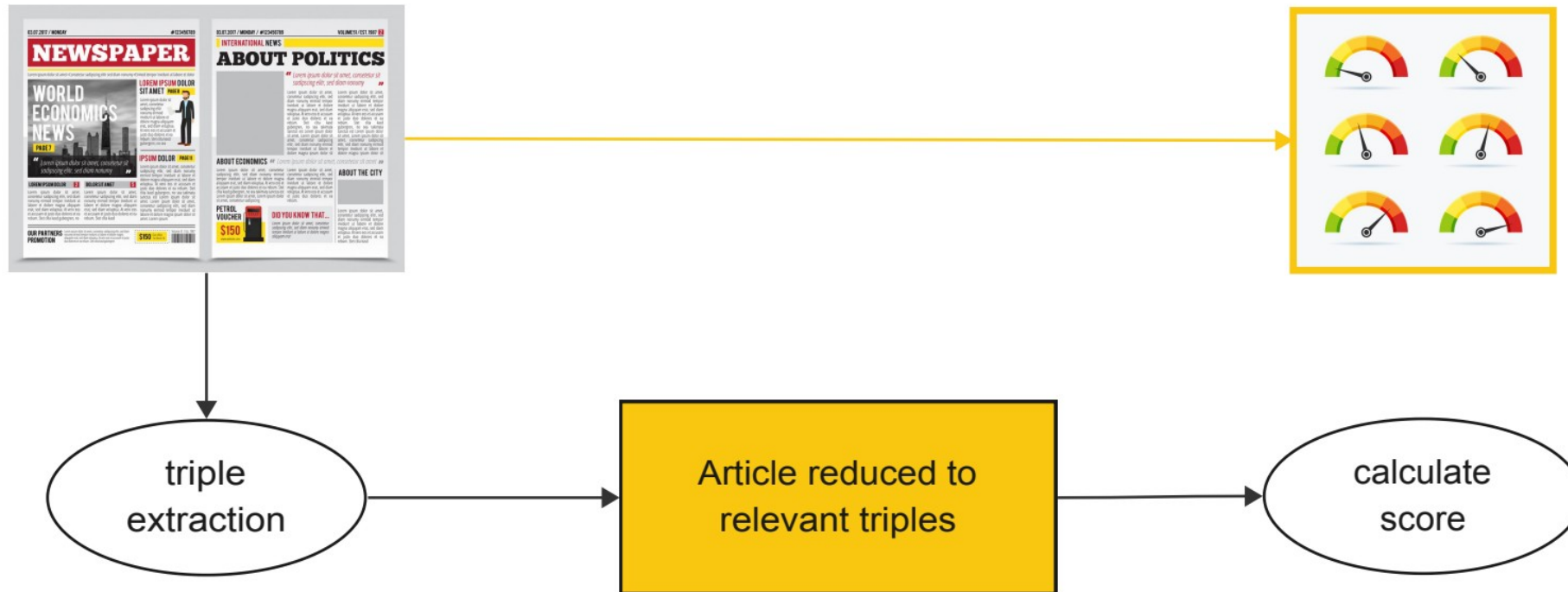
Approach

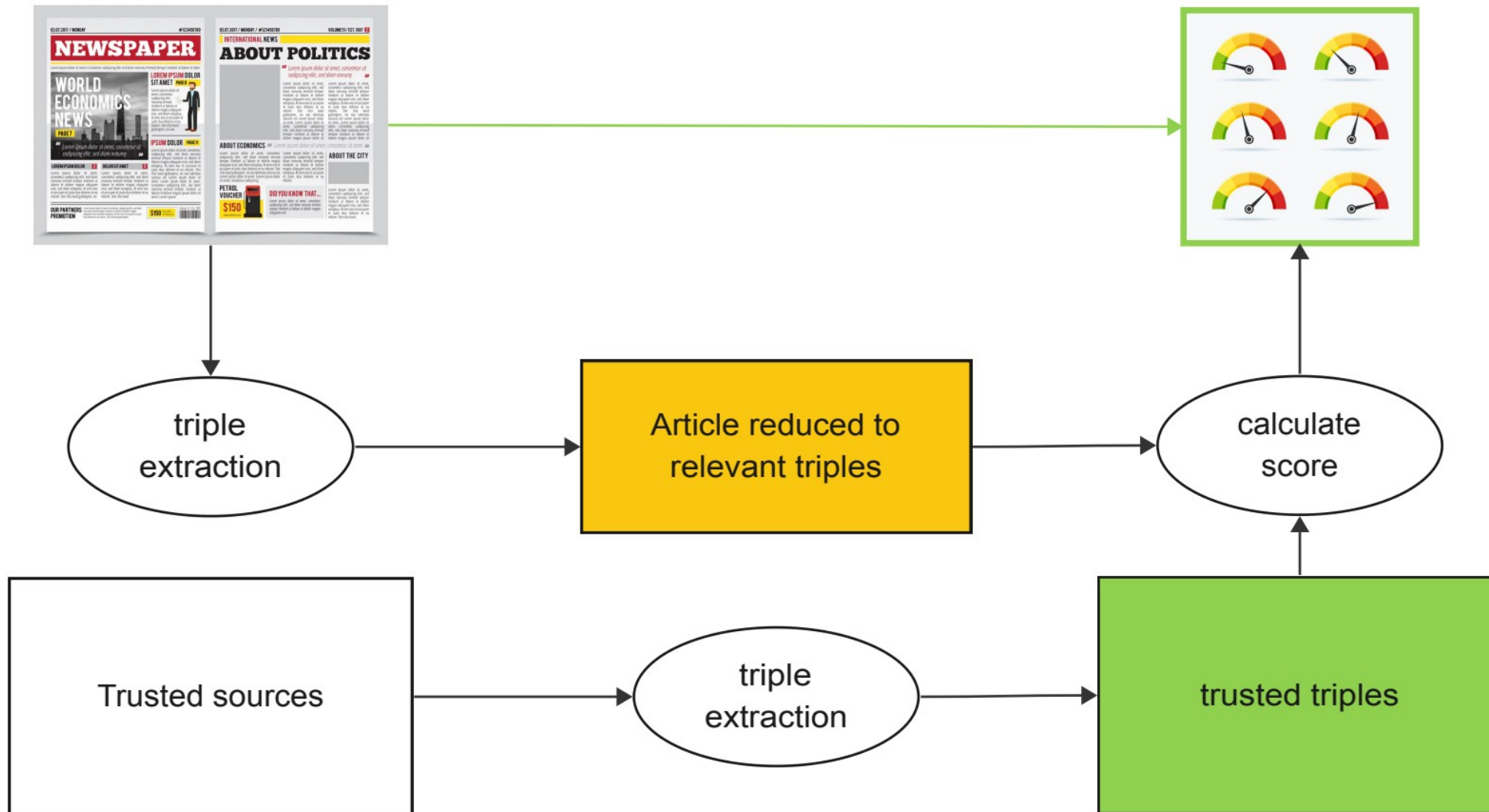


Approach

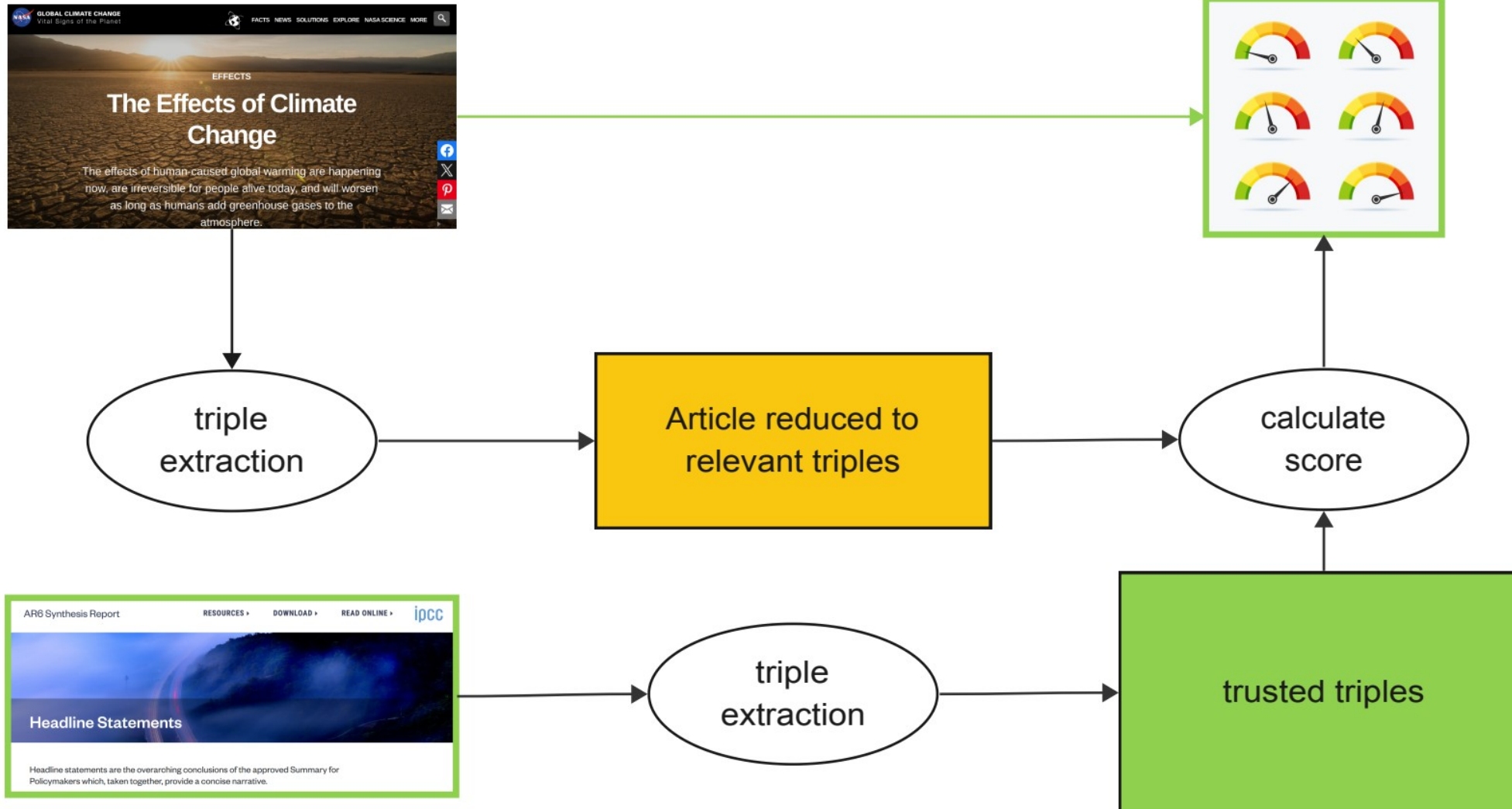


Approach

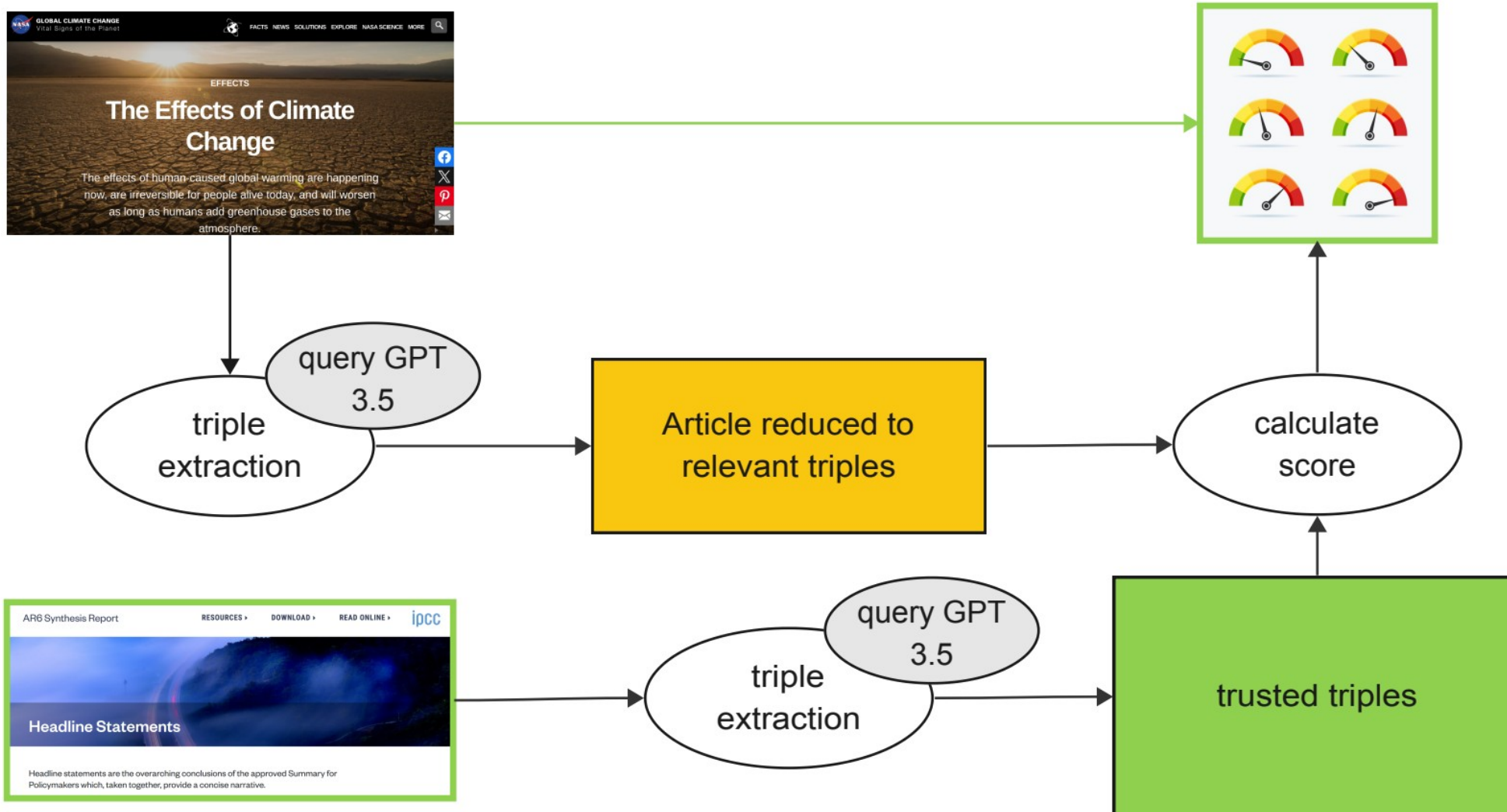




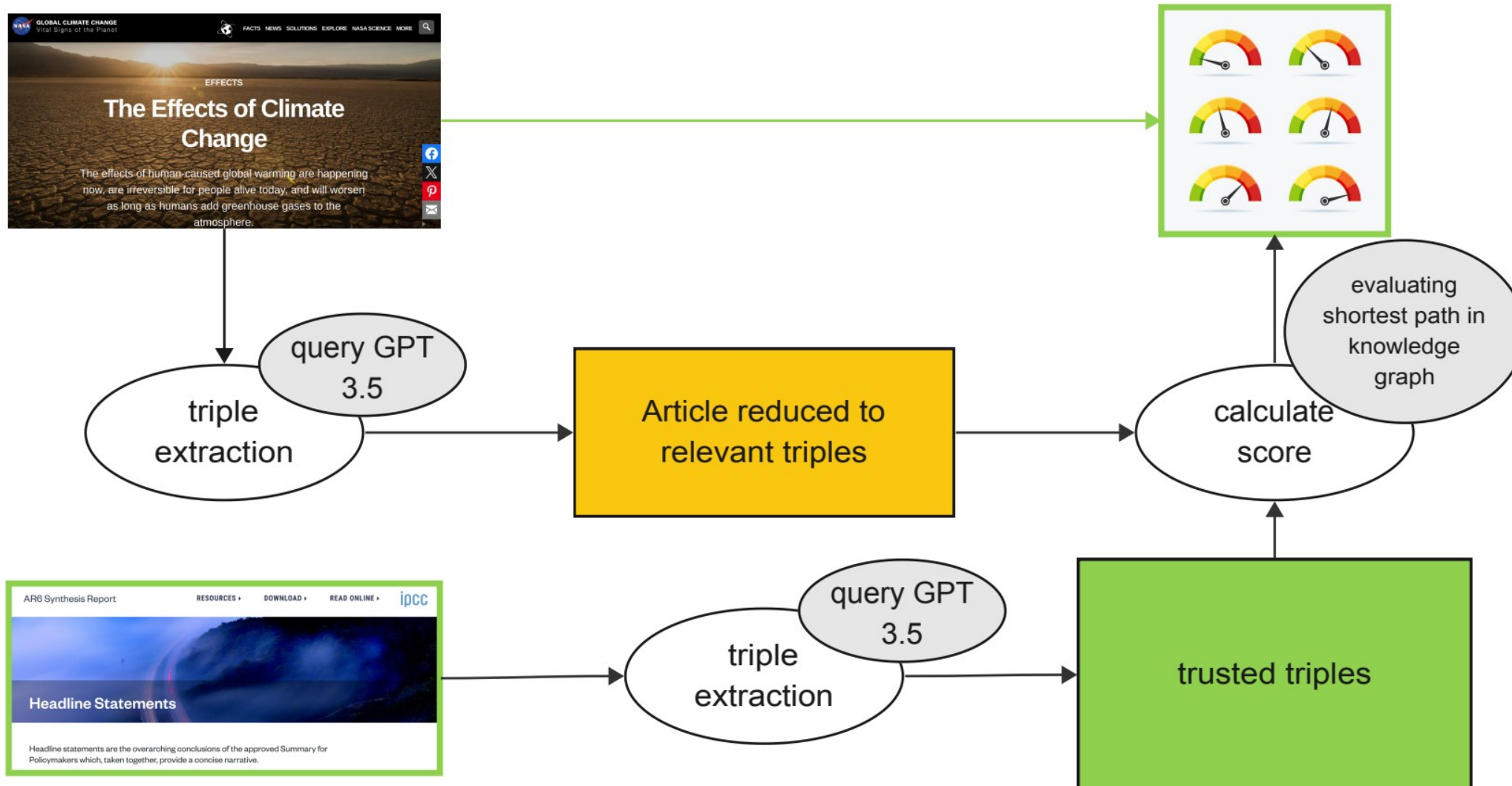
Approach



Approach



Approach



Additional material

input: text (length of roughly one paragraph)

a) Abstract Meaning
Representation (AMR)
Graphs

b) Named Entity
Recognition (NER)
with domain specific
model

c) LLM extracts triple +
simple error repair

output: triples in turtle star format

Additional material

input: domain

1) domain specific keywords

a) hand selected
based on selected
domain dataset

b) selected by a
community of:
- users
- domain experts
- mix of both

c) (semi) automated
with a LLM based on:
- domain dataset
- most frequently
used terms in media

d) (semi) automated
with a domain
specific NER model
based on:
- domain dataset
- most frequently
used terms in media

output: list with domain specific keywords

input: list with domain specific keywords

1) selection + 2) trustworthiness + (optional)

a) hand selected

i) my evaluation

b) community based
selection

ii) domain expert
evaluation

b) semi automated
search with keywords

iii) semi automated
evaluation

c) automated search
with keywords

iv) automated
evaluation

reducing text to about
one paragraph

*at the moment the triple
extraction works better with
short texts*

*reduction based on domain
specific keywords or
structure of the text*

output: text

input: knowledge base,
list of triples from media

a) veracity: existence
of triple in knowledge
base

b) veracity: based on path inside
knowledge base and certainties and
temporal validity

a) objectivity

a) transparency

a) topic relevance

output: final score,
sub scores for each triple,
sub scores for each categorie

input: triples in turtle star format

1) knowledge base + 2) entity linking

a) Triple Store
(Graph DB)

b) property graph

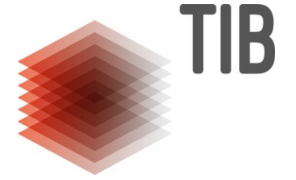
i) Link entities from
file based on hand
selected ontology

iii) Link triples from
file based on ontology
created specifically for
domain

ii) Link triples from file
based on multiple
hand selected
ontologies

output: knowledge graph

Questions



1. Is this approach up to date?
2. Is this connection of methods efficient?
3. Would you use this tool?

Final Questions

4. Triple extraction:

What method would you use if you have:

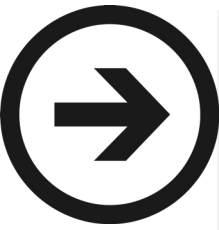
- a) a deadline soon?
- b) a lot of time?

5. Score calculation:

Is there a better way to check whether triples exist in a knowledge graph?

6. Do you have any other comments?

input: text (length of roughly one paragraph)



a) Abstract Meaning
Representation (AMR)
Graphs

b) Named Entity
Recognition (NER)
with domain specific
model

c) LLM extracts triple +
simple error repair

output: triples in turtle star format

**You**

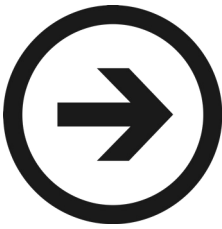
You will perform the open information extraction task. You will identify the named entities in the content and then extract the relations between them. Based on the provided testimony, you will return triples, which are formatted as [named entity A, relation, named entity B]. START of the testimony:

Human activities, principally through emissions of greenhouse gases, have unequivocally caused global warming, with global surface temperature reaching 1.1°C above 1850–1900 in 2011–2020. Global greenhouse gas emissions have continued to increase, with unequal historical and ongoing contributions arising from unsustainable energy use, land use and land-use change, lifestyles and patterns of consumption and production across regions, between and within countries, and among individuals (high confidence).

END of the testimony. The extracted triples formatted as .ttl are:

```
1 @prefix ex: <http://example.org/> .
2
3 ex:HumanActivities
4   ex:cause ex:GlobalWarming .
5
6 ex:GreenhouseGasEmissions
7   ex:cause ex:GlobalWarming .
8
9 ex:GlobalSurfaceTemperature
10  ex:reach "1.1°C above 1850-1900 in 2011-2020"^^ex:Temperature .
11
12 ex:GlobalGreenhouseGasEmissions
13   ex:increaseContinuously true .
14
15 ex:UnsustainableEnergyUse
16   ex:contributeTo ex:GreenhouseGasEmissions .
17
18 ex:LandUseAndLandUseChange
19   ex:contributeTo ex:GreenhouseGasEmissions .
20
21 ex:LifestylesAndPatterns
22   ex:contributeTo ex:GreenhouseGasEmissions .
23
24 ex:Contributions
25   ex:ariseFrom ex:UnsustainableEnergyUse, ex:LandUseAndLandUseChange, ex:LifestylesAndPatterns ;
26   ex:varyAcross ex:Regions ;
27   ex:varyBetweenAndWithin ex:Countries ;
28   ex:varyAmong ex:Individuals .
29
30 ex:Confidence
31   ex:hasLevel "high confidence"^^ex:ConfidenceLevel .
32
```

input: triples in turtle star format



1) knowledge base

+

2) entity linking

a) Triple Store
(Graph DB)

b) property graph

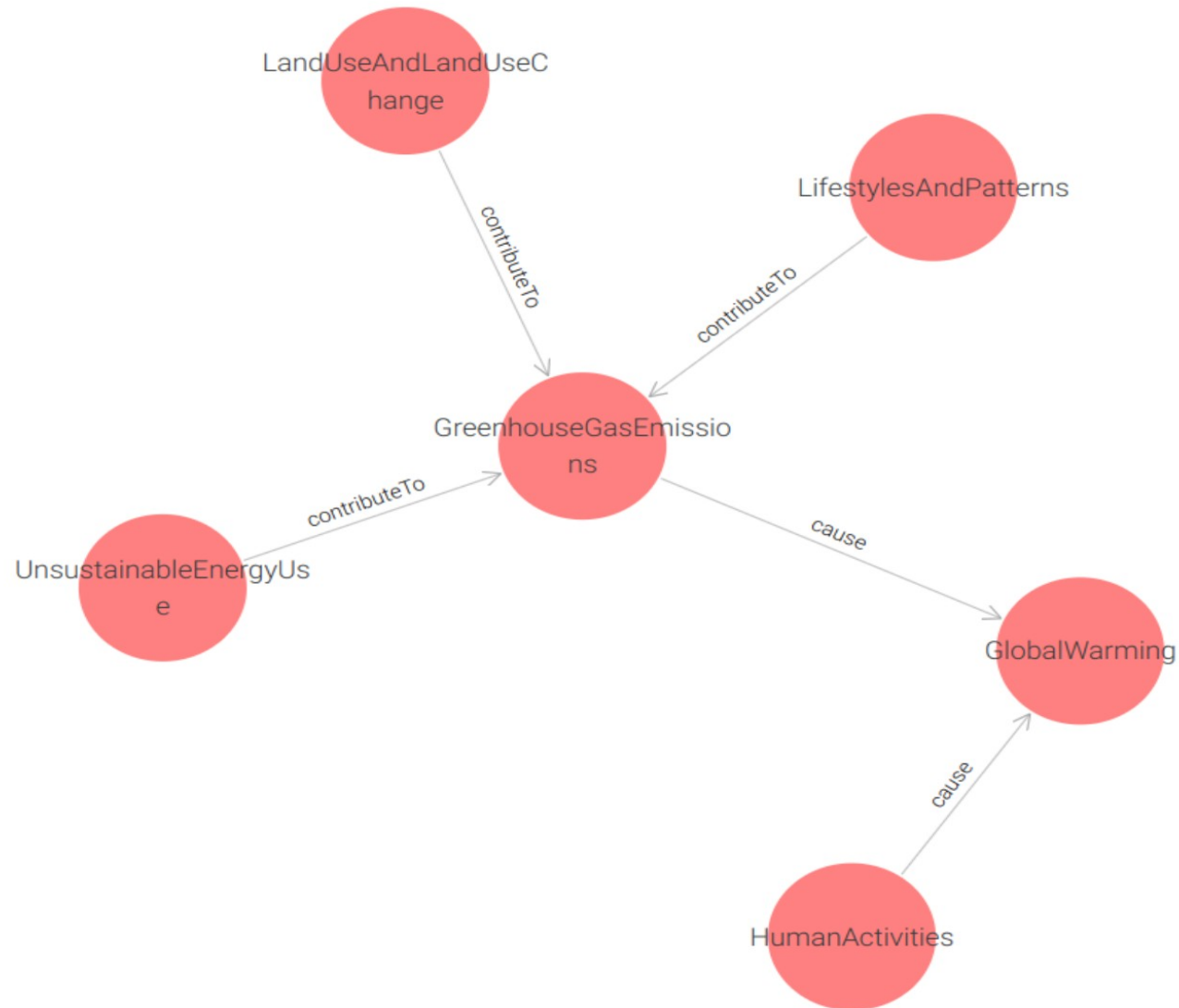
i) Link entities from
file based on hand
selected ontology

ii) Link triples from file
based on multiple
hand selected
ontologies

iii) Link triples from
file based on ontology
created specifically for
domain

output: knowledge graph

Visual graph



input: list with domain specific keywords

1) selection

+

2) trustworthiness

+

(optional)

a) hand selected

i) my evaluation

b) community based
selection

ii) domain expert
evaluation

b) semi automated
search with keywords

iii) semi automated
evaluation

c) automated search
with keywords

iv) automated
evaluation

reducing text to about
one paragraph

*at the moment the triple
extraction works better with
short texts*

*reduction based on domain
specific keywords or
structure of the text*

output: text

Headline statements are the overarching conclusions of the approved Summary for Policymakers which, taken together, provide a concise narrative.

A Current Status and Trends

Observed Warming and its Causes

A.1 Human activities, principally through emissions of greenhouse gases, have unequivocally caused global warming, with global surface temperature reaching 1.1°C above 1850–1900 in 2011–2020. Global greenhouse gas emissions have continued to increase, with unequal historical and ongoing contributions arising from unsustainable energy use, land use and land-use change, lifestyles and patterns of consumption and production across regions, between and within countries, and among individuals (*high confidence*). {2.1, Figure 2.1, Figure 2.2}.

Observed Changes and Impacts

A.2 Widespread and rapid changes in the atmosphere, ocean, cryosphere and biosphere have occurred. Human-caused climate change is already affecting many weather and climate extremes in every region across the globe. This has led to widespread adverse impacts and related losses and damages to nature and people (*high confidence*). Vulnerable communities who have historically contributed the least to current climate change are disproportionately affected (*high confidence*). {2.1, Table 2.1, Figure 2.2 and 2.3} (Figure SPM.1)

input: knowledge base,
list of triples from media

a) veracity: existence
of triple in knowledge
base

b) veracity: based on path inside
knowledge base and certainties and
temporal validity

a) objectivity

a) transparency

a) topic relevance

output: final score,
sub scores for each triple,
sub scores for each categorie

input: domain

1) domain specific keywords

a) hand selected
based on selected
domain dataset

b) selected by a
community of:
- users
- domain experts
- mix of both

c) (semi) automated
with a LLM based on:
- domain dataset
- most frequently
used terms in media

d) (semi) automated
with a domain
specific NER model
based on:
- domain dataset
- most frequently
used terms in media

output: list with domain specific keywords

Open ▾



vocabIPCC.txt

~/Documents/Masterarbeit/Scripts/Tests/LLM

Save



1 human activity
2
3 emission
4 greenhouse gas
5 CO2
6
7 climate change
8 global warming
9
10 weather
11
12 adaptation
13 implementation
14 action
15
16 mitigation
17
18 policy
19 law
20 climate goal
21
22 risk
23
24 impact
25 loss
26 damage
27
28

Future
expansions:



Media type

Language

Score Criteria

- Text
 - Political Bills
 - Coalition agreements
 - Draft laws
 - Blog post
 - Social media post
 - Audio
 - Podcast
 - Radio shows
 - Video
 - Short viral videos (e.g. TikTok, Instagram)
 - Long video essays (e.g. YouTube)
- Temporal relevance (“Barack Obama is president” true in the past)
 - Clearness
 - Domain relevance
 - Transparency