THE UNIVERSITY OF TOKYO TEAM

What information is provided in strategies about policy implementation, such as specific goals, timelines, budgetary commitments or policy actions and/or their governance/monitoring?

Chen Zhu

Qiuhan Zhao

Ryota Sawano

Sinndy Dayana Rico L.

東京大学
THE UNIVERSITY OF TOKYO

東京大学大学院
工学系研究科
SCHOOL OF ENGINEERING
THE UNIVERSITY OF TOKYO

Team Approach

Part 1: Concrete information related to STI policies

1. Keyword searching in "Strategies Database":
    - ○ Strategies
    - ○ Specific goals
    - ○ Budget commitment
    - ○ Policy actions
    - ○ Governance

Analysis of quantitative results

- ○ Keyword frequency aggregated at the country-level
- ○ Extracting paragraphs at document level (further development)

Part 2: Relationship between STI policies

1. Topic modeling.
2. Characteristic vector by country in accordance with the topics modeled.
3. Similarity between countries considering selected topics.

Analysis of results:

- ○ What topics are more related to Part 1 keywords.
- ○ Which countries are more focused on those topics in accordance with their characteristic vectors.

Part 3: MOIP and strategies

1. Classify data set (strategies) into MOIP related and NO-MOIP related groups:
    - ○ Using basic keywords search

2. Topic modeling in each group (MOIP and NO-MOIP) (NO-MOIP for further development of comparisons)

Analysis of results (further development)

東京大学
THE UNIVERSITY OF TOKYO

東京大学大学院
工学系研究科
SCHOOL OF ENGINEERING
THE UNIVERSITY OF TOKYO

Part 1

Concrete information related to STI policies

東京大学
THE UNIVERSITY OF TOKYO

東京大学大学院
工学系研究科
SCHOOL OF ENGINEERING
THE UNIVERSITY OF TOKYO

1. Keyword searching

- ➔ **Strategies:** strategy, plan, agenda, policy, program
- ➔ **Specific goals:** goal, directionality, aim, target, purpose, object(ive)
- ➔ **Timelines:** milestone, roadmap
- ➔ **Budgetary commitments:** budget assignment, budget allocation, grants
- ➔ **Policy actions:** implementation, execution, carrying out, action plan
- ➔ **Governance:** monitoring, foresight, impact assessment, policy intelligence, evaluation, policy coordination, feedback, lessons

1. Keyword searching: Six Indicators

The six indicators are also well related to the given aspects in question. The indicators were created by calculating the word count and word frequency of the selected sets.

strategies_set = ['strategy', 'plan', 'agenda', 'policy', 'program']

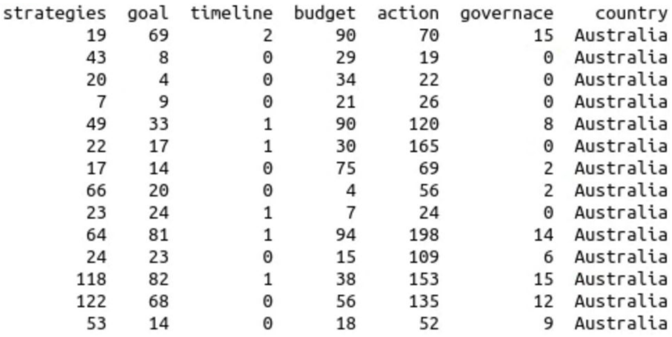
goal_set = ['goal', 'directionality', 'aim', 'target', 'purpose', 'objective', 'vision']

timeline_set = ['milestone', 'loadmap']

budget_set = ['budget', 'fund', 'grant', 'investment', 'budget allocation']

action_set = ['implement', 'execut', 'act']

governance_set = ['monitor', 'foresight', 'impact assessment', 'policy intelligence', 'evaluation', 'policy coordination', 'feedback', 'lessons']



```
strategies  goal  timeline  budget  action  governace    country
        19    69         2      90      70         15  Australia
        43     8         0      29      19          0  Australia
        20     4         0      34      22          0  Australia
         7     9         0      21      26          0  Australia
        49    33         1      90     120          8  Australia
        22    17         1      30     165          0  Australia
        17    14         0      75      69          2  Australia
        66    20         0       4      56          2  Australia
        23    24         1       7      24          0  Australia
        64    81         1      94     198         14  Australia
        24    23         0      15     109          6  Australia
       118    82         1      38     153         15  Australia
       122    68         0      56     135         12  Australia
        53    14         0      18      52          9  Australia
```

Example of results of word count for every policy document of Australia

Then for each policy document, we compute two vectors, respectively.

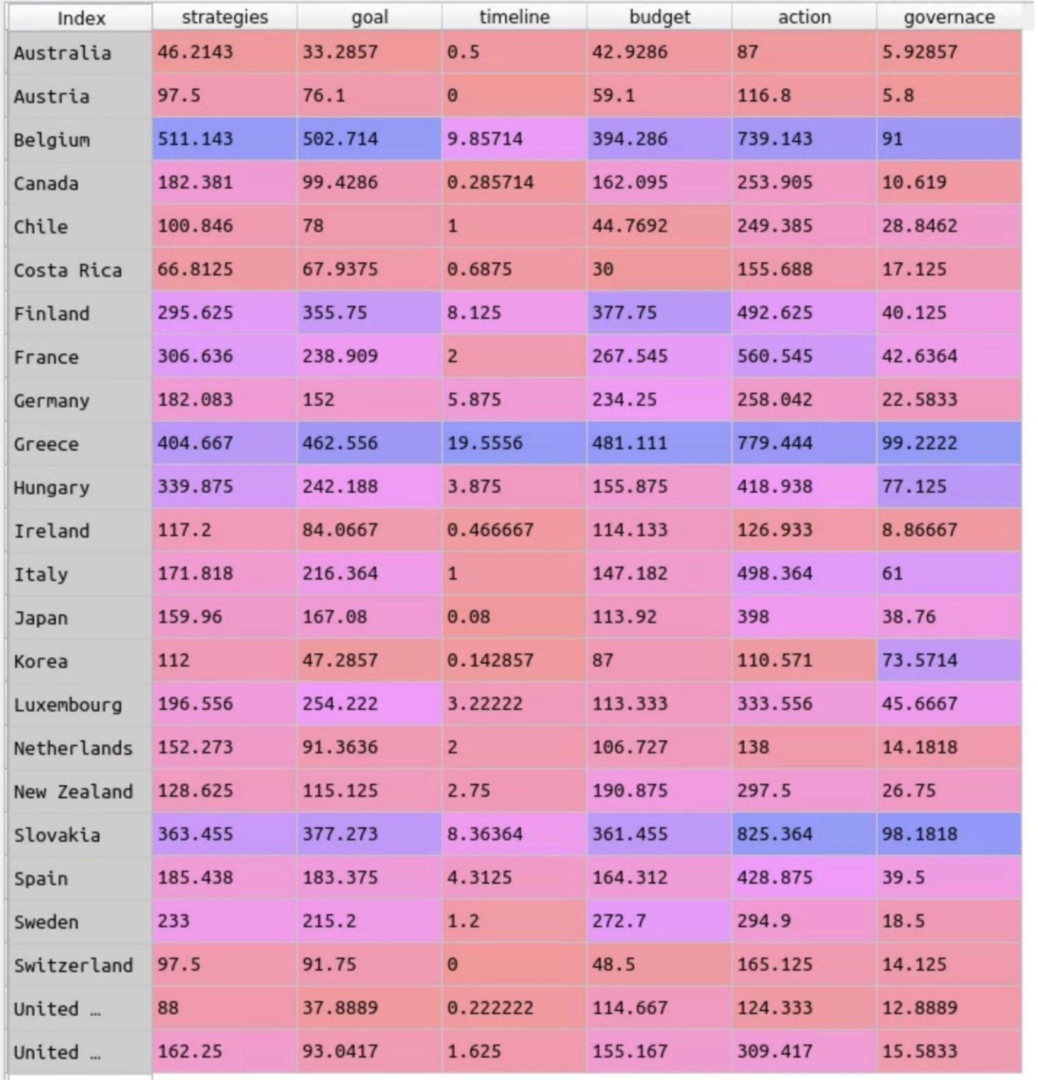
Vector of word count = [Count_strategy, Count_goal, Count_timeline, Count_budget, Count_action, Count_govern]

Vector of word frequency = [Freq_strategy, Freq_goal, Freq_timeline, Freq_budget, Freq_action, Freq_govern]

1. Keyword searching:
Total keywords' count

Vector of keyword count aggregated at the country-level

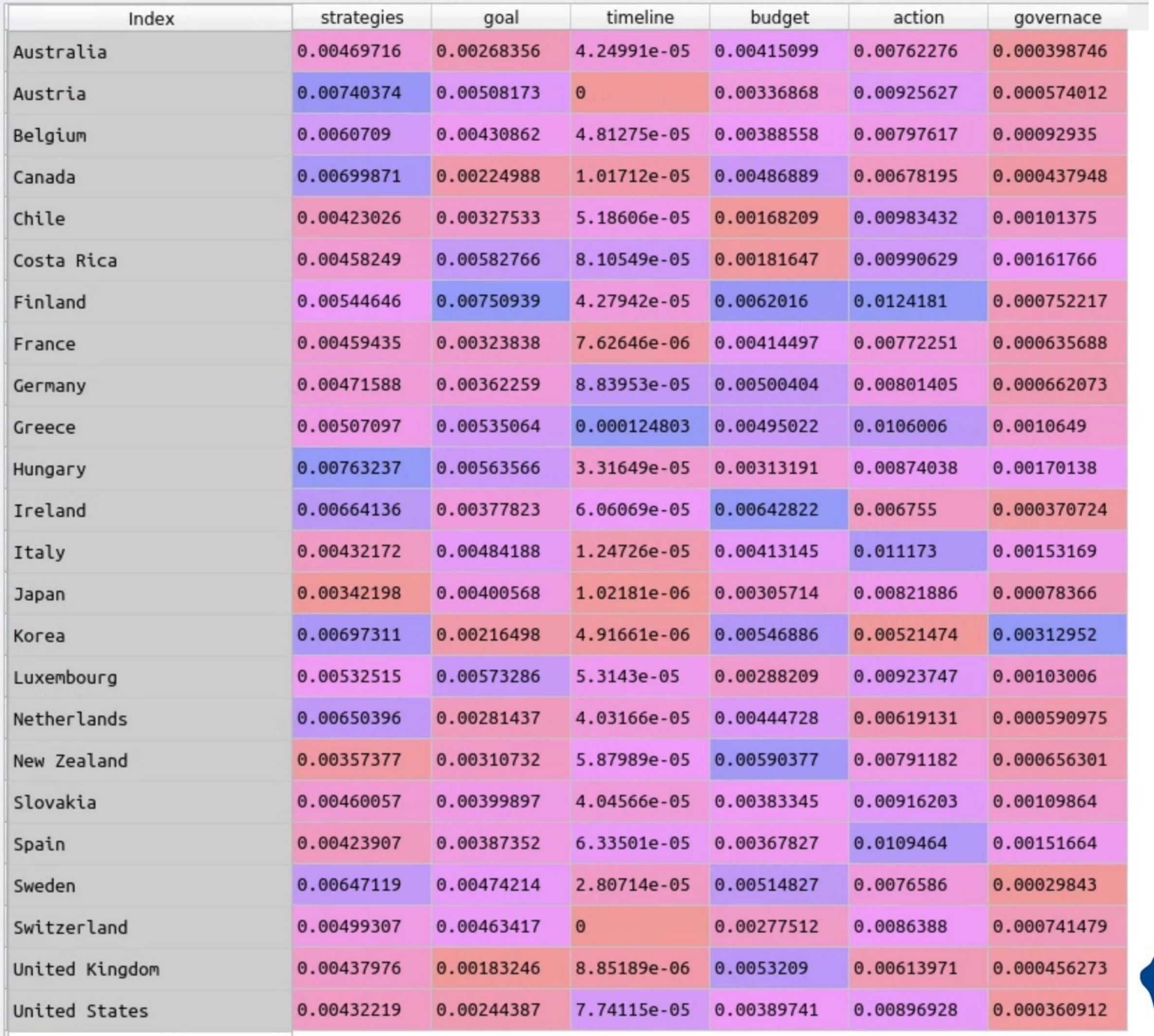
* Where aggregation means that the value in each cell corresponds to the mean of the keyword count of all policy documents by country.



Index	strategies	goal	timeline	budget	action	governace
Australia	46.2143	33.2857	0.5	42.9286	87	5.92857
Austria	97.5	76.1	0	59.1	116.8	5.8
Belgium	511.143	502.714	9.85714	394.286	739.143	91
Canada	182.381	99.4286	0.285714	162.095	253.905	10.619
Chile	100.846	78	1	44.7692	249.385	28.8462
Costa Rica	66.8125	67.9375	0.6875	30	155.688	17.125
Finland	295.625	355.75	8.125	377.75	492.625	40.125
France	306.636	238.909	2	267.545	560.545	42.6364
Germany	182.083	152	5.875	234.25	258.042	22.5833
Greece	404.667	462.556	19.5556	481.111	779.444	99.2222
Hungary	339.875	242.188	3.875	155.875	418.938	77.125
Ireland	117.2	84.0667	0.466667	114.133	126.933	8.86667
Italy	171.818	216.364	1	147.182	498.364	61
Japan	159.96	167.08	0.08	113.92	398	38.76
Korea	112	47.2857	0.142857	87	110.571	73.5714
Luxembourg	196.556	254.222	3.22222	113.333	333.556	45.6667
Netherlands	152.273	91.3636	2	106.727	138	14.1818
New Zealand	128.625	115.125	2.75	190.875	297.5	26.75
Slovakia	363.455	377.273	8.36364	361.455	825.364	98.1818
Spain	185.438	183.375	4.3125	164.312	428.875	39.5
Sweden	233	215.2	1.2	272.7	294.9	18.5
Switzerland	97.5	91.75	0	48.5	165.125	14.125
United …	88	37.8889	0.222222	114.667	124.333	12.8889
United …	162.25	93.0417	1.625	155.167	309.417	15.5833

1. Keyword searching: Keywords’ frequency

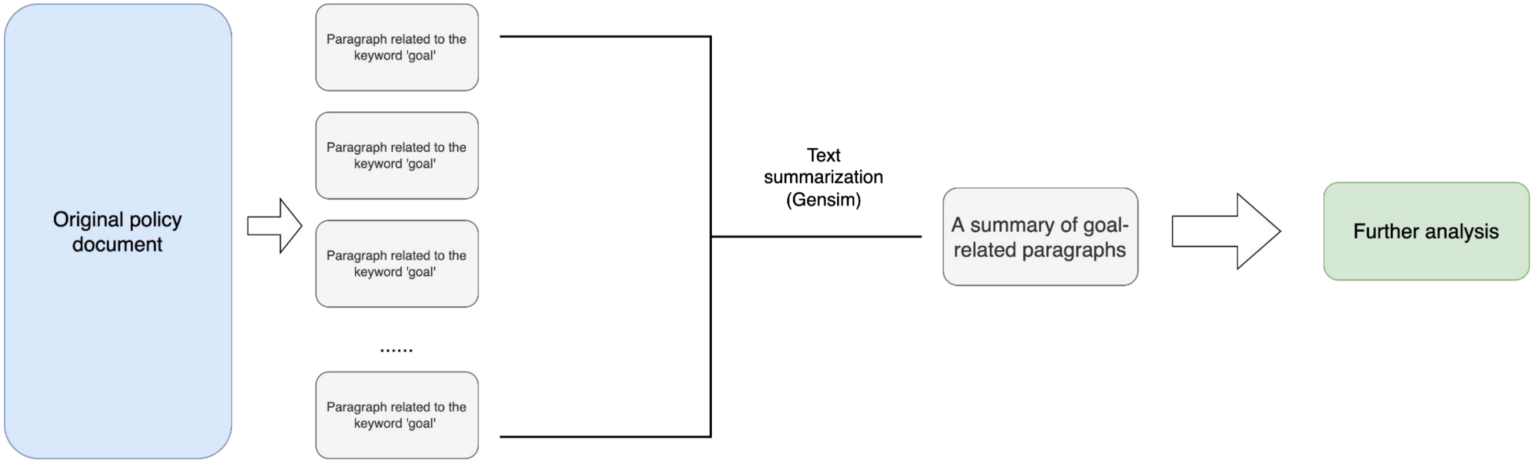
Vector of keyword frequency aggregated at the country-level



Index	strategies	goal	timeline	budget	action	governace
Australia	0.00469716	0.00268356	4.24991e-05	0.00415099	0.00762276	0.000398746
Austria	0.00740374	0.00508173	0	0.00336868	0.00925627	0.000574012
Belgium	0.0060709	0.00430862	4.81275e-05	0.00388558	0.00797617	0.00092935
Canada	0.00699871	0.00224988	1.01712e-05	0.00486889	0.00678195	0.000437948
Chile	0.00423026	0.00327533	5.18606e-05	0.00168209	0.00983432	0.00101375
Costa Rica	0.00458249	0.00582766	8.10549e-05	0.00181647	0.00990629	0.00161766
Finland	0.00544646	0.00750939	4.27942e-05	0.0062016	0.0124181	0.000752217
France	0.00459435	0.00323838	7.62646e-06	0.00414497	0.00772251	0.000635688
Germany	0.00471588	0.00362259	8.83953e-05	0.00500404	0.00801405	0.000662073
Greece	0.00507097	0.00535064	0.000124803	0.00495022	0.0106006	0.0010649
Hungary	0.00763237	0.00563566	3.31649e-05	0.00313191	0.00874038	0.00170138
Ireland	0.00664136	0.00377823	6.06069e-05	0.00642822	0.006755	0.000370724
Italy	0.00432172	0.00484188	1.24726e-05	0.00413145	0.011173	0.00153169
Japan	0.00342198	0.00400568	1.02181e-06	0.00305714	0.00821886	0.00078366
Korea	0.00697311	0.00216498	4.91661e-06	0.00546886	0.00521474	0.00312952
Luxembourg	0.00532515	0.00573286	5.3143e-05	0.00288209	0.00923747	0.00103006
Netherlands	0.00650396	0.00281437	4.03166e-05	0.00444728	0.00619131	0.000590975
New Zealand	0.00357377	0.00310732	5.87989e-05	0.00590377	0.00791182	0.000656301
Slovakia	0.00460057	0.00399897	4.04566e-05	0.00383345	0.00916203	0.00109864
Spain	0.00423907	0.00387352	6.33501e-05	0.00367827	0.0109464	0.00151664
Sweden	0.00647119	0.00474214	2.80714e-05	0.00514827	0.0076586	0.00029843
Switzerland	0.00499307	0.00463417	0	0.00277512	0.0086388	0.000741479
United Kingdom	0.00437976	0.00183246	8.85189e-06	0.0053209	0.00613971	0.000456273
United States	0.00432219	0.00244387	7.74115e-05	0.00389741	0.00896928	0.000360912

2. Analysis: Extracting paragraphs containing each keyword

A potential use of keywords search is using Gensim to summarize paragraphs at a policy document level as follows:


Original policy document
Paragraph related to the keyword 'goal'
Paragraph related to the keyword 'goal'
Paragraph related to the keyword 'goal'
......
Paragraph related to the keyword 'goal'
Text summarization (Gensim)
A summary of goal-related paragraphs
Further analysis

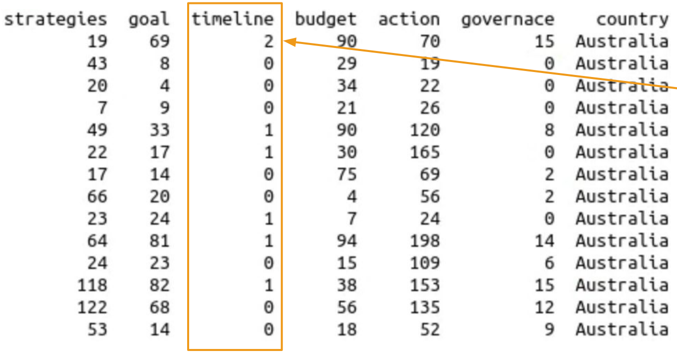
2. Analysis: Extracting paragraphs containing each keyword

We propose a keyword-based searching to extract **paragraphs** related to strategic goals, timelines, budgetary commitments, policy actions, and governance/monitoring.

Paragraph/Targeted sentence (-n/+m): n sentences + targeted sentence + m sentences.

* Where a targeted sentence is the sentence that contains the keyword.

** Further analysis*



```
strategies  goal  timeline  budget  action  governace    country
        19    69         2      90      70          15  Australia
        43     8         0      29      19           0  Australia
        20     4         0      34      22           0  Australia
         7     9         0      21      26           0  Australia
        49    33         1      90     120           8  Australia
        22    17         1      30     165           0  Australia
        17    14         0      75      69           2  Australia
        66    20         0       4      56           2  Australia
        23    24         1       7      24           0  Australia
        64    81         1      94     198          14  Australia
        24    23         0      15     109           6  Australia
       118    82         1      38     153          15  Australia
       122    68         0      56     135          12  Australia
        53    14         0      18      52           9  Australia
```

Example of results of word count for every policy document of Australia

Example:

Because the words related to “timeline_set” appears 2 times (word count) in the first policy document of Australia, there will be 2 extracted paragraphs for “timeline” is such a document. They can be summarized in 1 paragraph by using Gensim.

Therefore, in total, there will be 7 summarized paragraphs related to “timeline” for all policy documents of Australia.

The same process would be done for other keywords and countries.

2. Analysis: Extracting paragraphs

Example: Targeted sentence (-/+5) of keyword ‘goal’
Related paragraph

Departments and agencies should prioritize evidence-based standards and research to rapidly establish microorganism, plant, and animal safety and efficacy for products developed using gene editing, to better accelerate biotechnology product adoption and socially responsible use. Additionally, departments and agencies should focus on R&D that enables biotechnology, omics, scientific collections, biosecurity, and data analytics to drive economic growth across multiple sectors including healthcare, pharmaceuticals, manufacturing, and agriculture. 5. American Space Exploration and Commercialization R&D investments should continue to leverage efforts underway at American universities and in the private sector and focus on ensuring American leadership in space by supporting the Trump Administration's call for a return of Americans to the Moon's surface by 2024 and utilizing the Moon as a proving-ground for a future human mission to Mars. Departments and agencies should prioritize in-situ resource utilization on the Moon and Mars, cryogenic fuel storage and management, in-space manufacturing and assembly, and advanced space-related power and propulsion capabilities. **Departments and agencies should also prioritize activities that ensure an industrial base for commercial activity in space and that will broadly speed private-sector progress in meeting stated Government goals and furthering the space economy.** Finally, departments and agencies should seek opportunities to work with advanced materials, additive manufacturing, and machine learning capabilities that have broad potential applications in space and on Earth. PRIORITY CROSSCUTTING ACTIONS 1. Build and Leverage a Diverse, Highly Skilled American Workforce The Trump Administration's 2018 report, Charting a Course for Success: America's Strategy for STEM Education (STEM Strategy), articulates a vision that "all Americans will have lifelong access to high quality STEM education and the United States will be the global leader in STEM literacy, innovation, and employment". 11 Achieving this vision depends on a multisector seamless STEM education and training ecosystem that can meet the needs of all Americans from all backgrounds and ZIP codes and can adapt to the changing, and often growing, demands for ST:8M knowledge and skills in both the workplace and in everyday life. STC.

Paragraph related to the keyword 'goal'

2. Analysis: Extracting paragraphs

Example: Targeted sentence (-/+5) of keyword ‘goal’.
A summary of goal-related paragraphs

Advanced Manufacturing: Department and agency R&D investments should support the goals in the National Science and Technology Council (NSTC) report, Strategies for American Leadership in Advanced Manufacturing. American Energy and Environmental Leadership Advancing energy technologies, understanding our unexplored ocean and expanding use of ocean data, and improving our Earth system prediction capabilities are Administration priorities that will enhance the nation's economic vitality, national security, and environmental quality. American Space Exploration and Commercialization R&D investments should continue to leverage efforts underway at American universities and in the private sector and focus on ensuring American leadership in space by supporting the Trump Administration's call for a return of Americans to the Moon's surface by 2024 and utilizing the Moon as a proving-ground for a future human mission to Mars. Departments and agencies should prioritize in-situ resource utilization on the Moon and Mars, cryogenic fuel storage and management, in-space manufacturing and assembly, and advanced space-related power and propulsion capabilities.

A summary of goal-related paragraphs

Part 2

Relationship between STI policies in each country

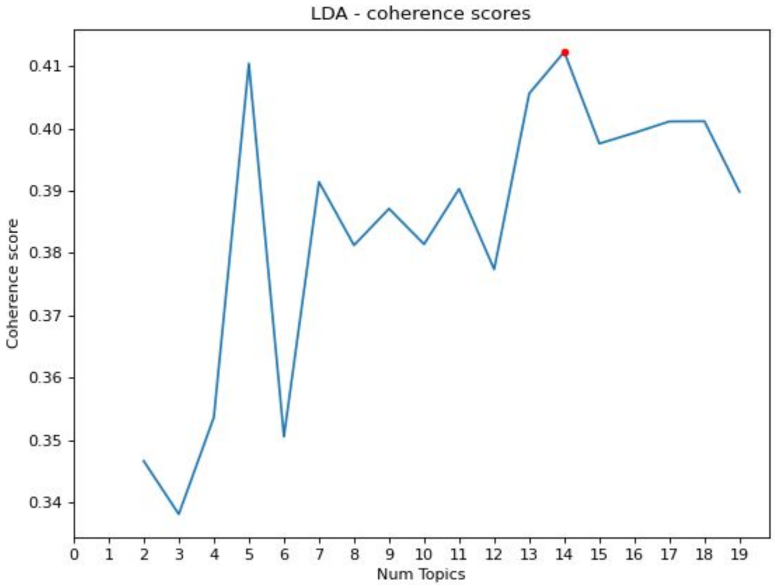
1. Topic Modeling results

We use topic modeling to visualize and gain an overall feeling of these strategic documents.

One key issue of topic modeling is to select the number of topics. Here, the coherence score is used. According to the figure below, we select 14 topics for the following analysis.

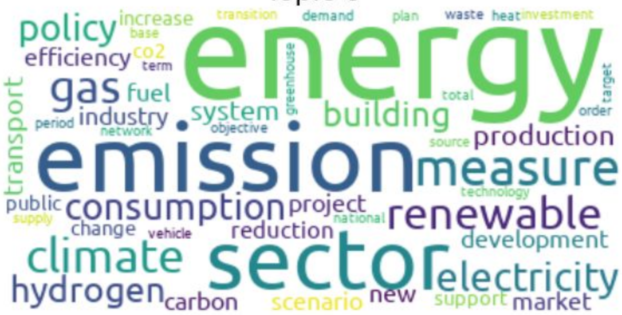
The coherence score for a single topic measures the semantic similarity (represented by co-occurrence of words) between words within the topic => we hope words appearing in the same topic should be related to each other.

And, we select the number of topics where the overall topic coherence can be maximized.



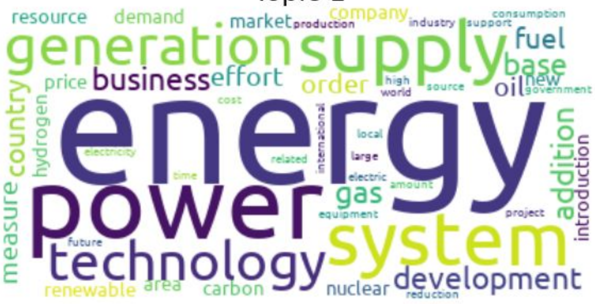
1. Topic Modeling results

Topic 0



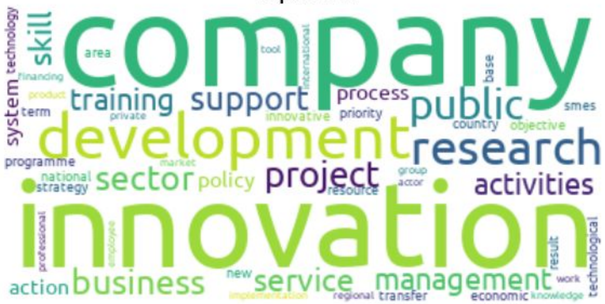
Green technology

Topic 1



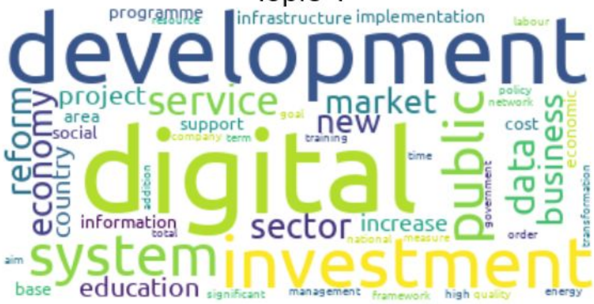
Energy

Topic 11



Management

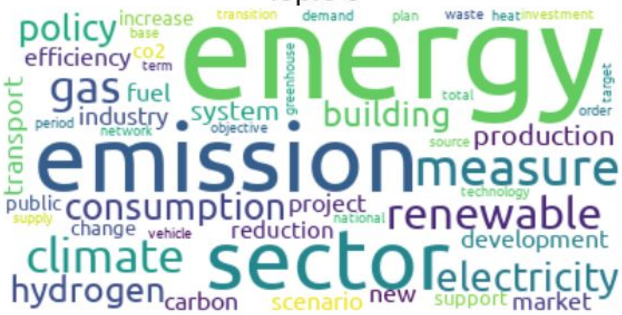
Topic 4



Digital transformation

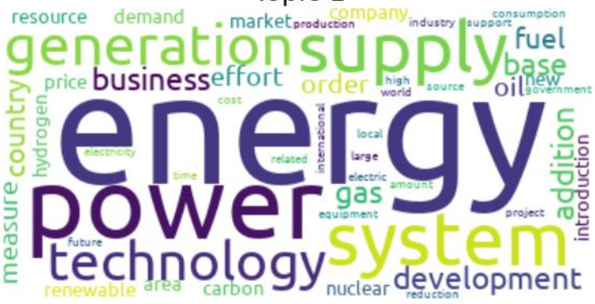
1. Topic Modeling results

Topic 0



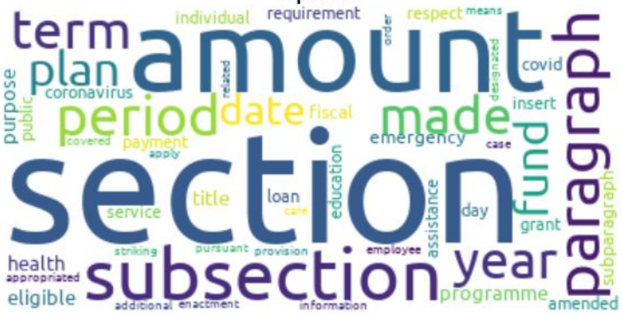
Green technology

Topic 1



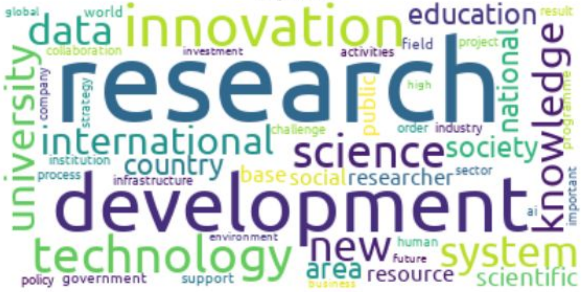
Energy

Topic 2



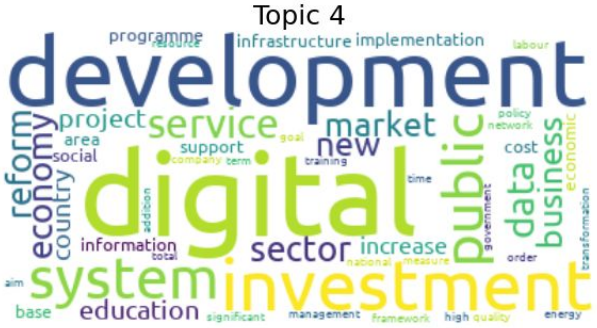
Public finance

Topic 3

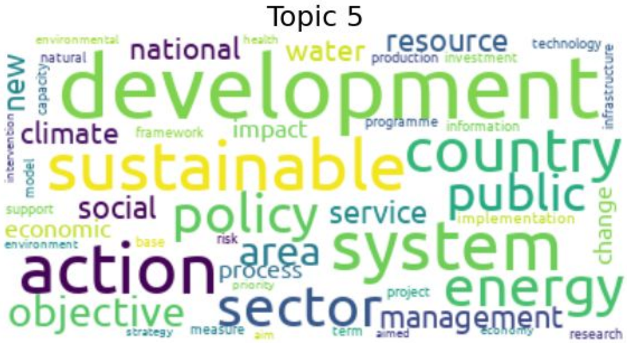


R&D

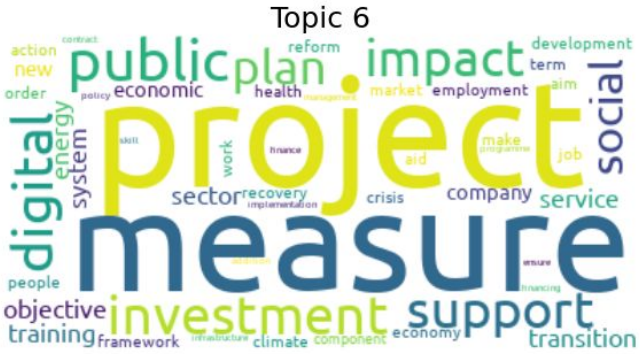
1. Topic Modeling results



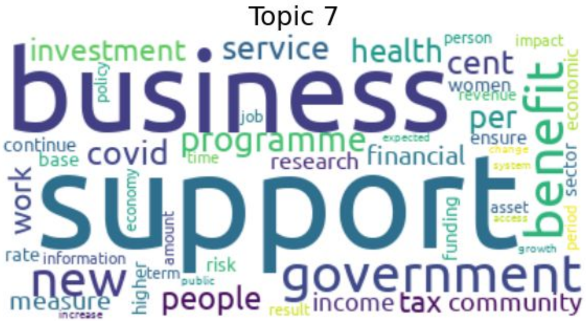
Digital transformation



Sustainability

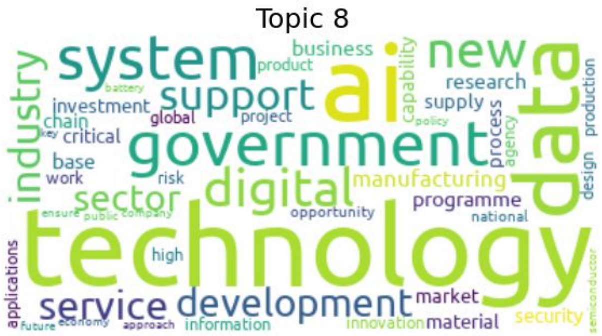


Assessment

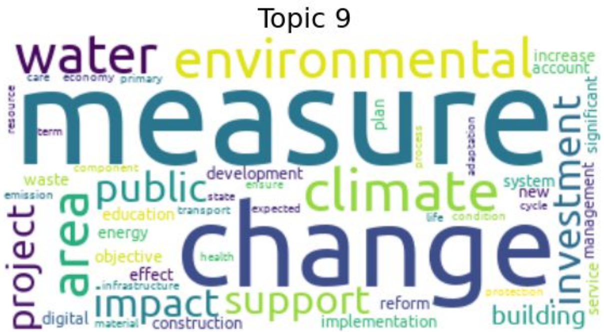


Public Support

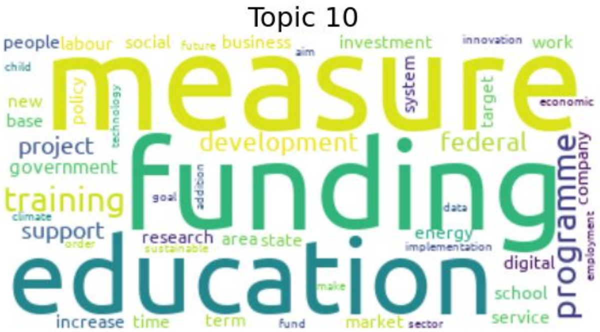
1. Topic Modeling results



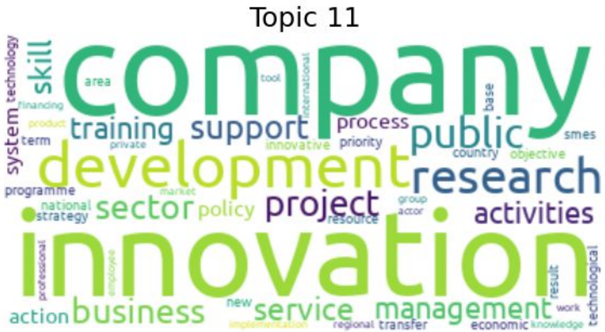
AI



Climate change

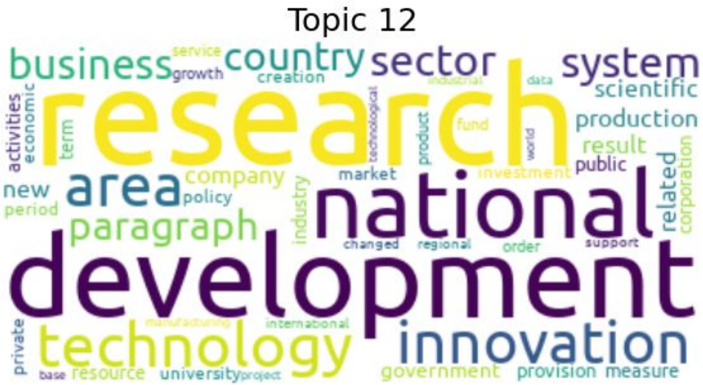


Education & Training

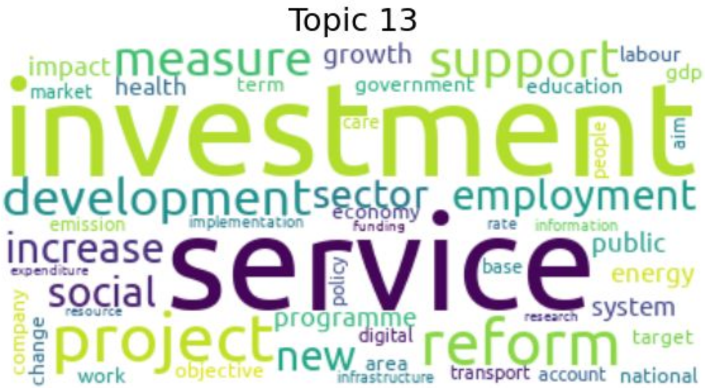


Management

1. Topic Modeling results


Topic 12

National Innovation System


Topic 13

Investment

1. Topic Modeling results

Considering the elements obtained for each of the 14 topics, we selected which are more related to policy implementation in accordance with keywords of Part 1.

0. Green tech.
1. Energy
2. Public finance
3. R&D
4. Digital transformation
5. Sustainability
6. Assessment
7. Public Support
8. AI
9. Climate change
10. Educ & training
11. Management
12. Nat. Innov. Sys.
13. Investment

1. Public finance
2. Assessment
3. Public Support
4. Educ & training
5. Management
6. Nat. Innov. Sys.
7. Investment

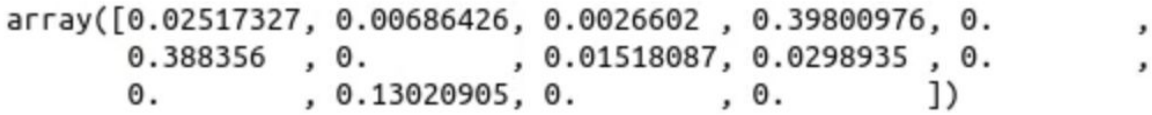
2. Characteristic vector by country

For each country, we created a characteristic vector based on the topic modeling results. The formula is given as follows:

A country characteristic vector = Mean(policy document vectors belonging to the country)

A policy document vector is a 14-dimensional vector where each element represents its relatedness to the corresponding topic.

There is an example of how a vector looks.



```
array([0.02517327, 0.00686426, 0.0026602 , 0.39800976, 0.        ,
       0.388356  , 0.        , 0.01518087, 0.0298935 , 0.        ,
       0.        , 0.13020905, 0.        , 0.        ])
```

A resulting subvector is created for each policy document with the 7 selected topics directly related to policy implementation.

2. Characteristic vector by country

Topic0	Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8	Topic9	Topic10	Topic11	Topic12	Topic13	Country
Green Tech	Energy	Public Finance	R&D	DX	Sustainability	Assessment	Public Support	AI	Environmental	Education	Biz Dev	Nat. Inno. Sys.	Investment	
0.01	0.00	0.00	0.31	0.00	0.00	0.01	0.45	0.17	0.00	0.00	0.04	0.00	0.00	United Kingdom
0.03	0.01	0.00	0.40	0.00	0.39	0.00	0.02	0.03	0.00	0.00	0.13	0.00	0.00	Chile
0.02	0.04	0.10	0.19	0.00	0.01	0.00	0.05	0.58	0.00	0.00	0.00	0.00	0.00	United States
0.01	0.00	0.00	0.17	0.02	0.02	0.00	0.62	0.10	0.00	0.00	0.04	0.00	0.00	New Zealand
0.00	0.00	0.01	0.03	0.57	0.17	0.03	0.04	0.01	0.00	0.01	0.08	0.00	0.04	Greece
0.03	0.01	0.01	0.19	0.04	0.39	0.00	0.00	0.04	0.00	0.00	0.26	0.04	0.00	Costa Rica
0.00	0.00	0.00	0.15	0.00	0.11	0.00	0.02	0.00	0.00	0.20	0.06	0.00	0.46	Finland
0.00	0.00	0.02	0.19	0.04	0.22	0.01	0.11	0.01	0.01	0.09	0.31	0.00	0.00	Switzerland
0.09	0.02	0.00	0.19	0.01	0.02	0.00	0.22	0.45	0.00	0.00	0.00	0.00	0.00	Australia
0.04	0.00	0.00	0.20	0.02	0.11	0.23	0.01	0.13	0.01	0.10	0.05	0.00	0.07	Luxembourg
0.29	0.00	0.00	0.01	0.01	0.08	0.29	0.01	0.01	0.06	0.17	0.01	0.00	0.05	Belgium
0.04	0.01	0.00	0.31	0.00	0.03	0.00	0.00	0.14	0.02	0.38	0.05	0.00	0.00	Germany
0.01	0.34	0.01	0.36	0.07	0.00	0.01	0.03	0.05	0.00	0.01	0.02	0.09	0.00	Japan
0.10	0.12	0.00	0.34	0.09	0.00	0.01	0.00	0.32	0.00	0.02	0.00	0.00	0.00	Korea
0.17	0.01	0.01	0.24	0.02	0.01	0.00	0.07	0.08	0.10	0.17	0.03	0.00	0.08	Sweden
0.01	0.00	0.05	0.33	0.02	0.02	0.03	0.19	0.19	0.00	0.07	0.03	0.00	0.06	Ireland
0.01	0.00	0.00	0.16	0.00	0.01	0.29	0.02	0.02	0.00	0.01	0.41	0.00	0.06	France
0.09	0.01	0.01	0.11	0.01	0.17	0.19	0.00	0.04	0.00	0.01	0.34	0.00	0.01	Spain
0.04	0.00	0.01	0.21	0.25	0.06	0.00	0.02	0.03	0.07	0.08	0.21	0.00	0.03	Hungary
0.12	0.00	0.00	0.54	0.00	0.00	0.00	0.02	0.11	0.00	0.11	0.07	0.00	0.02	Netherlands
0.19	0.01	0.00	0.08	0.11	0.03	0.00	0.03	0.02	0.29	0.04	0.07	0.00	0.14	Slovakia
0.00	0.00	0.01	0.11	0.04	0.22	0.03	0.03	0.05	0.00	0.00	0.42	0.06	0.02	Italy
0.06	0.01	0.00	0.20	0.00	0.00	0.00	0.42	0.25	0.00	0.00	0.03	0.00	0.00	Canada
0.11	0.03	0.00	0.27	0.02	0.01	0.04	0.01	0.05	0.05	0.33	0.06	0.00	0.01	Austria

Top1 of each Topic   Top2-3 of each Topic

2. Characteristic vector by country. Analysis.

Several insights can be obtained from the characteristic vector matrix. Main conclusions from the 7 prioritized topics are:

➔ **Public Finance:** relatedness to the topic ranges from 0.00 to 0.10, where only one country (US) has 0.10 and fourteen countries obtained 0.00. It implies that there is not a big focus on financing from public investments in most of the policy documents in most of the countries. Only US, Ireland, and Switzerland have values greater than 0.20.

➔ **Assessment:** 12 countries presented a relatedness different to zero. France (0.29), Belgium (0.29), and Luxembourg (0.23) have policy documents that treat the assessment/evaluation aspect of regulations. This is a vital point when implementing STIP within a medium-long time perspective.

➔ **Public support:** Most of countries' documents deal with public support by some way (83%), where the highest relatedness to the topic corresponds to New Zealand (0.62), United Kingdom (0.45), and Canada (0.42). However, this contrast with the low values presented in most of the other implementation-related documents.

➔ **Education & training:** The topic appears in 16 countries but only 6 of them showed values greater than 0.10. Germany, Austria, and Finland are on the top. This also matches with the characteristics of their education systems.

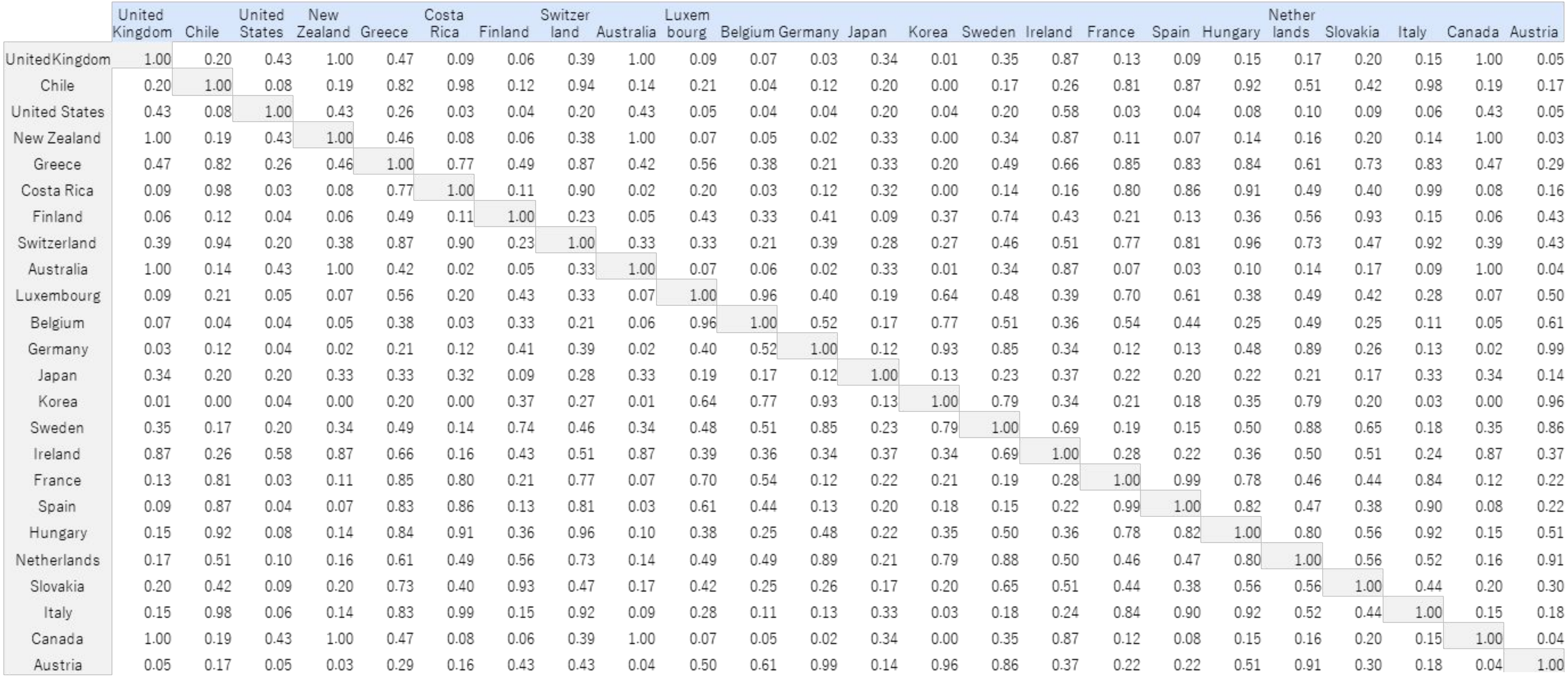
2. Characteristic vector by country. Analysis.

Main conclusions from the 7 selected topics are (continuation):

- ➔ **Management:** Only policy documents from 3 countries are not related to this topic (Australia, US, and Korea). Italy, France, Spain, and Switzerland have the greater values. An interesting point here is that 2 Latin American countries have values greater than 0.10 (Costa Rica and Chile).
- ➔ **National Innovation System:** This is the topic with less values across all countries. Only Japan, Italy, and Costa Rica received positive relatedness, 0.09. 0.06, and 0.04 respectively. One can conclude that there is no clear mentions to a innovation systems in the longtime perspective and then, it may be considered as an issue when implementing the STI policies.
- ➔ **Investment:** This topic is related to the Public Finance topic but this is more opened to any kind of investment in STI, including different types of resources (services, expenditures, infrastructure, labour, etc.). 13 countries have relatedness, where Finland is in the top with a large advantage (0.46 in comparison to the second ranked country, Slovakia, which obtained 0.14).

3. Similarity between countries

To obtain the similarity between countries considering the 7 selected topics, we used the cosine similarity matrix derived from their vector representations. It resulted in a 24x24 matrix as follows.



	United Kingdom	Chile	United States	New Zealand	Greece	Costa Rica	Finland	Switzerland	Australia	Luxembourg	Belgium	Germany	Japan	Korea	Sweden	Ireland	France	Spain	Hungary	Netherlands	Slovakia	Italy	Canada	Austria
UnitedKingdom	1.00	0.20	0.43	1.00	0.47	0.09	0.06	0.39	1.00	0.09	0.07	0.03	0.34	0.01	0.35	0.87	0.13	0.09	0.15	0.17	0.20	0.15	1.00	0.05
Chile	0.20	1.00	0.08	0.19	0.82	0.98	0.12	0.94	0.14	0.21	0.04	0.12	0.20	0.00	0.17	0.26	0.81	0.87	0.92	0.51	0.42	0.98	0.19	0.17
United States	0.43	0.08	1.00	0.43	0.26	0.03	0.04	0.20	0.43	0.05	0.04	0.04	0.20	0.04	0.20	0.58	0.03	0.04	0.08	0.10	0.09	0.06	0.43	0.05
New Zealand	1.00	0.19	0.43	1.00	0.46	0.08	0.06	0.38	1.00	0.07	0.05	0.02	0.33	0.00	0.34	0.87	0.11	0.07	0.14	0.16	0.20	0.14	1.00	0.03
Greece	0.47	0.82	0.26	0.46	1.00	0.77	0.49	0.87	0.42	0.56	0.38	0.21	0.33	0.20	0.49	0.66	0.85	0.83	0.84	0.61	0.73	0.83	0.47	0.29
Costa Rica	0.09	0.98	0.03	0.08	0.77	1.00	0.11	0.90	0.02	0.20	0.03	0.12	0.32	0.00	0.14	0.16	0.80	0.86	0.91	0.49	0.40	0.99	0.08	0.16
Finland	0.06	0.12	0.04	0.06	0.49	0.11	1.00	0.23	0.05	0.43	0.33	0.41	0.09	0.37	0.74	0.43	0.21	0.13	0.36	0.56	0.93	0.15	0.06	0.43
Switzerland	0.39	0.94	0.20	0.38	0.87	0.90	0.23	1.00	0.33	0.33	0.21	0.39	0.28	0.27	0.46	0.51	0.77	0.81	0.96	0.73	0.47	0.92	0.39	0.43
Australia	1.00	0.14	0.43	1.00	0.42	0.02	0.05	0.33	1.00	0.07	0.06	0.02	0.33	0.01	0.34	0.87	0.07	0.03	0.10	0.14	0.17	0.09	1.00	0.04
Luxembourg	0.09	0.21	0.05	0.07	0.56	0.20	0.43	0.33	0.07	1.00	0.96	0.40	0.19	0.64	0.48	0.39	0.70	0.61	0.38	0.49	0.42	0.28	0.07	0.50
Belgium	0.07	0.04	0.04	0.05	0.38	0.03	0.33	0.21	0.06	0.96	1.00	0.52	0.17	0.77	0.51	0.36	0.54	0.44	0.25	0.49	0.25	0.11	0.05	0.61
Germany	0.03	0.12	0.04	0.02	0.21	0.12	0.41	0.39	0.02	0.40	0.52	1.00	0.12	0.93	0.85	0.34	0.12	0.13	0.48	0.89	0.26	0.13	0.02	0.99
Japan	0.34	0.20	0.20	0.33	0.33	0.32	0.09	0.28	0.33	0.19	0.17	0.12	1.00	0.13	0.23	0.37	0.22	0.20	0.22	0.21	0.17	0.33	0.34	0.14
Korea	0.01	0.00	0.04	0.00	0.20	0.00	0.37	0.27	0.01	0.64	0.77	0.93	0.13	1.00	0.79	0.34	0.21	0.18	0.35	0.79	0.20	0.03	0.00	0.96
Sweden	0.35	0.17	0.20	0.34	0.49	0.14	0.74	0.46	0.34	0.48	0.51	0.85	0.23	0.79	1.00	0.69	0.19	0.15	0.50	0.88	0.65	0.18	0.35	0.86
Ireland	0.87	0.26	0.58	0.87	0.66	0.16	0.43	0.51	0.87	0.39	0.36	0.34	0.37	0.34	0.69	1.00	0.28	0.22	0.36	0.50	0.51	0.24	0.87	0.37
France	0.13	0.81	0.03	0.11	0.85	0.80	0.21	0.77	0.07	0.70	0.54	0.12	0.22	0.21	0.19	0.28	1.00	0.99	0.78	0.46	0.44	0.84	0.12	0.22
Spain	0.09	0.87	0.04	0.07	0.83	0.86	0.13	0.81	0.03	0.61	0.44	0.13	0.20	0.18	0.15	0.22	0.99	1.00	0.82	0.47	0.38	0.90	0.08	0.22
Hungary	0.15	0.92	0.08	0.14	0.84	0.91	0.36	0.96	0.10	0.38	0.25	0.48	0.22	0.35	0.50	0.36	0.78	0.82	1.00	0.80	0.56	0.92	0.15	0.51
Netherlands	0.17	0.51	0.10	0.16	0.61	0.49	0.56	0.73	0.14	0.49	0.49	0.89	0.21	0.79	0.88	0.50	0.46	0.47	0.80	1.00	0.56	0.52	0.16	0.91
Slovakia	0.20	0.42	0.09	0.20	0.73	0.40	0.93	0.47	0.17	0.42	0.25	0.26	0.17	0.20	0.65	0.51	0.44	0.38	0.56	0.56	1.00	0.44	0.20	0.30
Italy	0.15	0.98	0.06	0.14	0.83	0.99	0.15	0.92	0.09	0.28	0.11	0.13	0.33	0.03	0.18	0.24	0.84	0.90	0.92	0.52	0.44	1.00	0.15	0.18
Canada	1.00	0.19	0.43	1.00	0.47	0.08	0.06	0.39	1.00	0.07	0.05	0.02	0.34	0.00	0.35	0.87	0.12	0.08	0.15	0.16	0.20	0.15	1.00	0.04
Austria	0.05	0.17	0.05	0.03	0.29	0.16	0.43	0.43	0.04	0.50	0.61	0.99	0.14	0.96	0.86	0.37	0.22	0.22	0.51	0.91	0.30	0.18	0.04	1.00

Part 3

Mission-Oriented Innovation Policies (MOIP) and strategies

1. MOIP related policy documents

Using direct keywords related to MOIP ('MOIP', 'mission-oriented', 'mission oriented'), only 9% of documents entered in the group MOIP. Therefore, to classify the policy documents into MOIP or NO-MOIP related, we run a keyword searching by using the following words.

moip_set = ['MOIP', 'mission-oriented', 'mission oriented', 'Participatory agenda', 'bottom-up', 'targeted missions', 'portfolio of solutions', 'problem-solution(s) space', 'problem-solution(s) constellation', 'decomposition of societal problems', 'big problems decomposition', 'societal subproblem(s)', 'societal sub-problem(s)', 'transdisciplinary'].

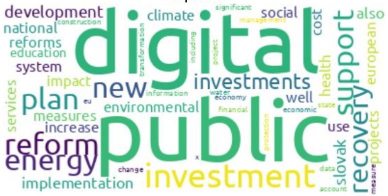
Results: 58 documents (18.5%) can be classified as policies related to MOIP characteristics.

Adding more keywords ['Societal challenge', 'challenge-based', 'challenge-oriented', 'top-down', 'proactive'], we found that 177 documents can be considered as MOIP related (56.5%).

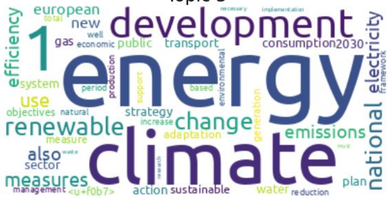
2. MOIP topic modeling

Following the same approach that Part 2, we found 10 topics within the MOIP related policy documents.

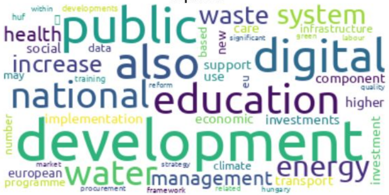
Topic 1


digital
public
investment
investments
support
recovery
new
plan
reform
energy
development
national
reforms
education
system
impact
measures
increase
implementation
environmental
climate
social
cost
health
european
also
well
use
slovak
projects
services
economic
change

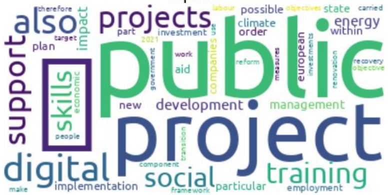
Topic 3


energy
climate
development
change
renewable
national
electricity
efficiency
emissions
european
new
gas
public
transport
consumption
2030
system
use
objectives
natural
strategy
adaptation
measure
also
sector
measures
management
action
sustainable
water
reduction
plan
<u+f0b7>

Topic 8


development
public
education
national
digital
also
water
management
energy
waste
system
infrastructure
health
social
data
increase
support
use
component
higher
implementation
economic
investments
european
programme
transport
climate
care
may
training
investment

Topic 9


public
project
projects
also
support
skills
digital
social
training
development
management
new
impact
plan
companies
possible
state
energy
climate
order
within
european
implementation
particular
employment
aid
work
reform
recovery

0. Emissions
1. Public investment
2. National innovation system
3. Energy
4. AI and sharing data
5. National development
6. Support & funding
7. Interoperability
8. Education
9. Projects and training

* Further development:

➔ Add more stopwords to refine the topic modeling process.
➔ Comparisons between topics of MOIP and topics of NO-MOIP.
➔ Relations between MOIP principles and topics found.
➔ Cosine similarity between characteristic vectors of both groups of policies.
➔ Re-run Part 1 for each group and go deeply into the analysis.

Complementary section

Additional potential applications explored

東京大学
THE UNIVERSITY OF TOKYO

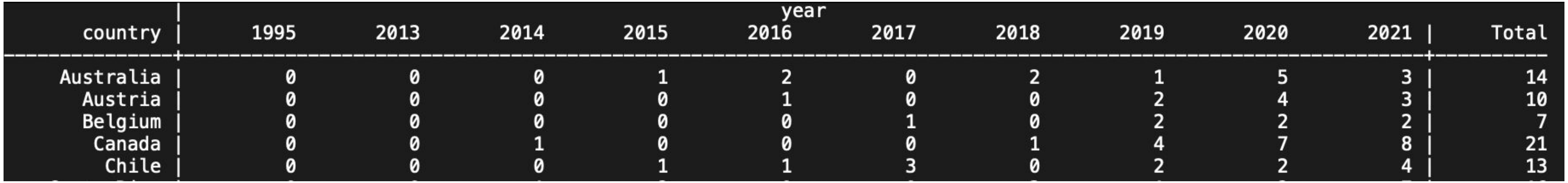
東京大学大学院
工学系研究科
SCHOOL OF ENGINEERING
THE UNIVERSITY OF TOKYO

1. Combining STI policy & other sources

Quantitative data obtained from policy documents can be analyzed in combination with other data [1]. For example, consider the plausibility of the following statement:

The higher the GDP per capita, the more inert the country is towards the enactment of STI policies.

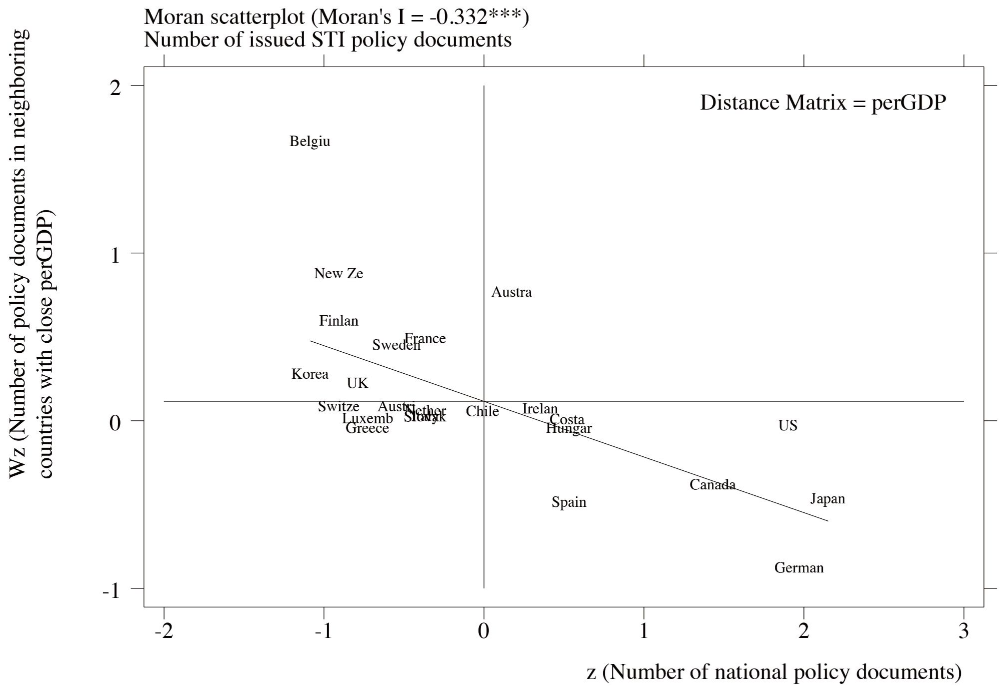
We constructed a matrix of economic distances for the 24 OECD countries in terms of GDP per capita (2011-2020, average in USD), as d_{ij}=1/abs(perGDP$_i$ - perGDP$_j$). Then, an example of the basic statistic of STI group by (country, year) is as follows:



```
              |                                        year
      country |      1995       2013       2014       2015       2016       2017       2018       2019       2020       2021 |     Total
--------------+-----------------------------------------------------------------------------------------------------------------+----------
    Australia |         0          0          0          1          2          0          2          1          5          3 |        14
      Austria |         0          0          0          0          1          0          0          2          4          3 |        10
      Belgium |         0          0          0          0          0          1          0          2          2          2 |         7
       Canada |         0          0          1          0          0          0          1          4          7          8 |        21
        Chile |         0          0          0          1          1          3          0          2          2          4 |        13
```

[1] https://data.worldbank.org.cn/indicator/NY.GDP.PCAP.CD?

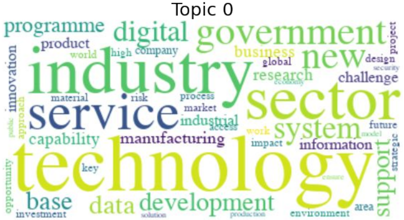
1. Combining STIP strategy database & other sources



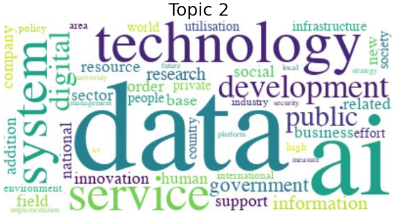
There is a significant negative effect of GDP per capita on the spatial relationship with the number of STI policy documents. That is, countries with higher GDP per capita have less incentive for STI policy to be published.

2. Combining STI policy & COVID-19 category

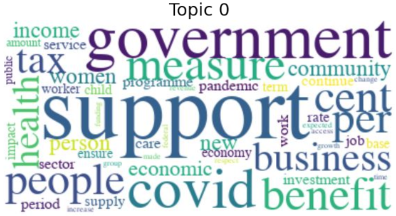
Statistically, because the STI dataset is evenly distributed over pre-covid (153) and intra-covid (160), it can be divided into two parts in order to know the impact of Covid-19 on STI policy making to some extent.



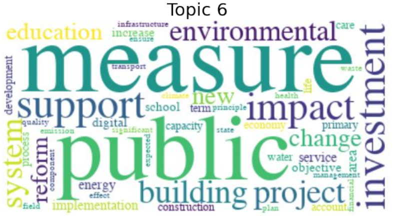
Industry



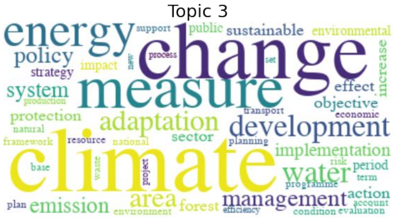
AI



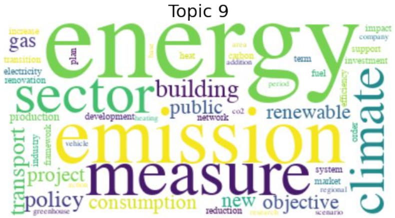
Governmental support



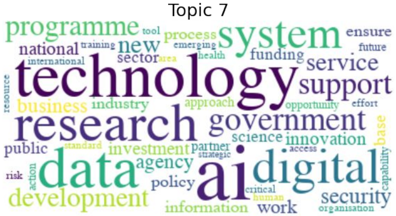
Public engagement



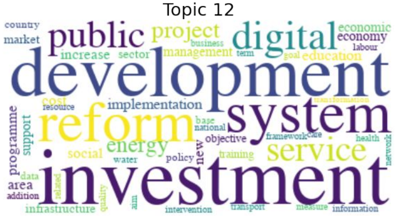
Climate change



Energy



AI



Investment

pre-covid

intra-covid