

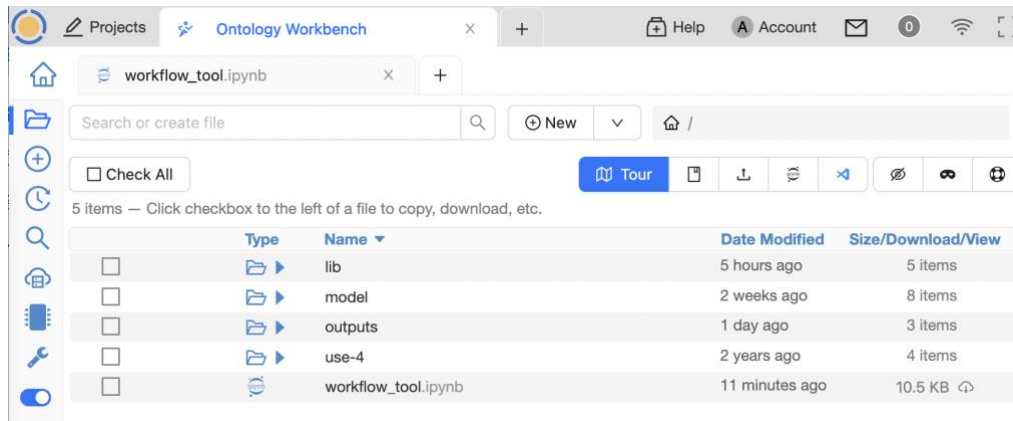
Ontology Enrichment Methodology¹

Ashley Moran

The Comparative Constitutions Project (CCP) ontology enrichment workflow conveys how we combine automation and human expertise to systematically discover new topics and assess existing topics in the CCP ontology. The methodology documented below conveys how to we implement this workflow, using the ontology enrichment tool to add and revise topics in the CCP ontology.

Login to the Ontology Tool

1. Open CoCalc, sign up for an account, and open the Ontology Workbench project. That should bring you here:



2. In the Name column, click on the “workflow_tool.ipynb” file. This will open a new page with the title “CCP Ontology Workflow Tool” (shown below). Under the title, you’ll see: **text cells**, with nothing to the left of the cell, which are the cells with our notes to walk you through the process of running the cell; **input cells**, with “In []:” to the left of the cell, which are the cells you’ll run to produce the desired outputs; and **output cells**, with “Out []:” to the left of the cell, which is where the results of running the input cells will appear. Since this is a collaborative platform, you’ll see the results of the last person who used it. So, your initial view may look slightly different from the screenshots below, but the basic structural elements will be the same.

CCP Ontology Workflow Tool

```
In [7]: __author__ = 'Roy Gardner'
__copyright__ = 'Copyright 2023, Roy and Sally Gardner'

%run ./lib/packages.py
%run ./lib/utilities.py
%run ./lib/clusters.py

try:
    n = len(model_dict)
    pass
except:
    encoder, model_dict, choice_dict = initialise()
```

Step 1: Select search parameters

This step will define the parameters for your topic search in the CCP ontology and in national constitutions. Run the cell below to generate an interface for selecting the following search parameters:

- User name
 - Enter your first name here
- Search threshold
 - Sets the minimum score a constitution section must meet to qualify as a match to your candidate topic and be included in the search results in Step 3
 - Too low and you’ll get too many results—many of which will be off-topic

Select Your Search Parameters

3. Scroll to “Step 1: Select search parameters” in the tool for guidance on selecting parameters for the search you’re about to do. Then scroll to the next input cell (shown below) and click on it. Click “Run” in the navigation bar.

¹ This material is based upon work supported by the National Science Foundation under Grant Number 2315189. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

```
In [3]: choice_dict = init_choice_dict()
generate_corpus_interface(choice_dict,0.64,0.72)
```

- That will generate an interface in the output cell that allows you to select your search parameters (shown below). Enter your first name as your **user ID**, set your **search threshold** and **cluster threshold**, enter your candidate topic text in the **topic** field, and (if you want to export the results in addition to reviewing them in the tool) check the **export results** box and enter what you want the file name to be for these search results in the **export name** field. Click “Apply Choices.” Note: Do not run this cell after you enter the search parameters, or it will clear your search parameters.

```
In [5]: choice_dict = init_choice_dict()
generate_corpus_interface(choice_dict,0.64,0.72)
```

Out[5]:

User ID:

Set thresholds:

Search: 0.62

Cluster: 0.72

Topic:

☒ Export results

Export name:

Search the CCP Ontology

- Scroll to “Step 2: Search the CCP ontology” in the tool for guidance on comparing your candidate topic to current topics in the CCP ontology. Then click on the input cell in that section (shown below). Run the cell.

```
In [11]: if len(choice_dict['topic']) > 0:
print('Search topic:',choice_dict['topic'])
search_ontology(choice_dict,model_dict,encoder)
else:
alert('No topic selected.')
```

- That will generate results in the output cell (shown below). The results show the current CCP topics that are the closest match to your candidate topic text. Typically, good matches are found above 0.65. Use these results to determine (i) if your candidate topic is distinct from current CCP topics or (ii) if revisions are needed to your candidate topic to differentiate it from current CCP topics. If revisions are needed, run your revised candidate topic back through Steps 1-2 of the ontology workflow tool. If no revisions are needed, proceed to Step 3 of the ontology workflow tool.

```
In [11]: if len(choice_dict['topic']) > 0:
print('Search topic:',choice_dict['topic'])
search_ontology(choice_dict,model_dict,encoder)
else:
alert('No topic selected.')
```

Out[11]: Search topic: energy independence and energy policy

Score	Label	Ontology
0.6	Territory secession. Allows withdrawal or independence of groups or territory	ccp
0.59	Executive independence. Guarantees the independence of the executive; relates to the separation of powers	ccp
0.58	Natural resources ownership. Describes ownership, possession, and use of natural resources, including energy, minerals, water, forests, flora, fauna, air, land, or other natural resources; includes nationalization, privatization, and use of natural resources as a public good or consumer good	ccp
0.57	Judicial independence. Guarantees that judges are autonomous and the judiciary is independent	ccp
0.56	Political theorists and figures. Reference to teachings of historical figures, theorists, or philosophers	ccp
0.56	First chamber policy areas. Policy areas where the first chamber of the legislature has special privileges or restrictions	ccp

Search National Constitutions

7. Scroll to “Step 3: Search national constitutions” in the tool for guidance on the search you’re about to do. Then click on the input cell in that section (shown below). Run the cell.

```
In [6]: if len(choice_dict['topic']) > 0:
        print('Search topic:',choice_dict['topic'])
        print('Search threshold:',choice_dict['search_threshold'])
        print('Cluster threshold:',choice_dict['cluster_threshold'])
        run_search(choice_dict,model_dict,encoder)
    else:
        alert('No topic selected.')
```

8. A green dot will appear next to the input cell to indicate it is running, and progress notes will appear in the output cell below it (shown below). Don’t run anything else while the input cell is running.

```
In [●]: if len(choice_dict['topic']) > 0:
        print('Search topic:',choice_dict['topic'])
        print('Search threshold:',choice_dict['search_threshold'])
        print('Cluster threshold:',choice_dict['cluster_threshold'])
        run_search(choice_dict,model_dict,encoder)
    else:
        alert('No topic selected.')
```

Out[9]: Search topic: energy independence and energy policy
Search threshold: 0.62
Cluster threshold: 0.72

9. That will generate results in the output cell (shown below). The results show all sections of national constitutions that match your candidate topic at the search threshold you defined. It clusters results by the cluster threshold you defined.

Out[8]: Search topic: energy independence and energy policy
Search threshold: 0.62
Cluster threshold: 0.72
Number of search results 63
Number of clusters 6

0

Section text	Constitution	Section ID
ensures realization of the financial credit and monetary policy;	Azerbaijan_2016	950
the definition and the implementation of monetary policy;	Democratic_Republic_of_the_Congo_2011	1019
determine monetary policy; and	Zambia_2016	3113
to formulate and implement monetary policy.	Zimbabwe_2017	4406
formulating monetary policy;	South_Sudan_2013	2111
the harmonization of the economical, financial and monetary policy;	Niger_2017	785

1

Section text	Constitution	Section ID
Energy policy.	Uganda_2017	4480
energy policy.	Kenya_2010	5059
energy policy;	Sweden_2012	2190

2

Section text	Constitution	Section ID
Energy policy including electricity and gas reticulation and energy regulation.	Kenya_2010	5075
electricity and gas reticulation and energy regulation.	Kenya_2010	5177
energy (including electricity and power generation and distribution);	Papua_New_Guinea_2016	4521
Energy and hydro electricity	Zambia_2016	3925
electricity and the establishment of electric power stations;	Nigeria_2011	4112
electricity and the establishment in that State of electric power stations;	Nigeria_2011	4124
Electricity; gas and gas works; and other works for the production and distribution of power and energy.	Malaysia_2007	3425

3

Section text	Constitution	Section ID
The Andorran People, with full liberty and independence, and in the exercise of their own sovereignty,	Andorra_1993	2
the defence of independence and sovereignty;	Mozambique_2007	49
the independence, the sovereignty and the unity of the State;	Mozambique_2007	2190
consolidating and strengthening the independence and sovereignty of our state,	Mongolia_2001	3
safeguarding and consolidating national sovereignty and independence;	Algeria_2020	67
Maintain the UAE's independence and sovereignty.	United_Arab_Emirates_2009	34
national independence and unity of the State;	East_Timor_2002	1391

National independence and the unity of the state;	Portugal_2005	2795
Respect for sovereignty and national independence;	Angola_2010	92
National independence, territorial integrity and unity;	Angola_2010	2271

4

Section text	Constitution	Section ID
Nuclear energy, atmosphere and space related	Nepal_2016	3736
mineral resources necessary for the generation of nuclear energy;	Pakistan_2018	3398
the production of nuclear fuels and the generation and use of nuclear energy;	Pakistan_2018	3400
Nuclear energy and nuclear installations, including—	United_Kingdom_2013	10683
Nuclear energy and nuclear installations, including—	United_Kingdom_2013	20062

5

Section text	Constitution	Section ID
to promote energy conservation and cost-effective use of energy, as well as to develop and support the production and use of alternative energy to enhance sustainable energy security.	Thailand_2017	292
The Confederation shall establish principles on the use of local and renewable energy sources and on the economic and efficient use of energy.	Switzerland_2014	708
The Confederation shall legislate on the use of energy by installations, vehicles and appliances. It shall encourage the development of energy technologies, in particular in the fields of saving energy and the renewable energy sources.	Switzerland_2014	710

unclustered

Section text	Constitution	Section ID
establish and pursue policy,	Serbia_2006	615
rocks, minerals, fossil fuels and other sources of energy;	Kenya_2010	4735
Basic regulation of mining and energy.	Spain_2011	1038
The State encourages all acts of protection of the environment, development and use of new energy and recycled energy.	Socialist_Republic_of_Vietnam_2013	343
To ensure food and energy sovereignty.	Ecuador_2021	1975
State Land and state Natural Resources;	South_Sudan_2013	2539
Financial and budgetary policies;	Gabon_2011	564
Policies governing natural resources;	Gabon_2011	565
Taxes on mineral oil, natural gas and minerals for use in generation of nuclear energy.	Pakistan_2018	3470
Non-conventional energy sources.	India_2016	6135
To protect forest reserves and expand the plant coverage, the State encourages the development of local sources of energy: solar, wind and others.	Haiti_2012	1329
Financial independence;	Guatemala_1993	996
The State submits itself [s'attelle] to develop its energy potential in view of achieving energy sovereignty, [and] access to energy and to build a dynamic and competitive industrial, mining, oil and a gas sector, oriented to satisfy the national necessities and the requirements of development.	Niger_2017	685
The Confederation shall legislate on the transport and the supply of electrical energy.	Switzerland_2014	719

The candidate topic above—*energy independence and energy policy*—matched 63 constitution sections, organized in 6 clusters. Matches scoring 0.65 and above are typically fairly good matches.

The search above used a relatively low **search threshold** of 0.62 to produce more matches; this helps ensure the search won't miss relevant matches, but it also increases the number of irrelevant matches. The search used a relatively high **cluster threshold** of 0.72, meaning topics in each cluster are likely to be very closely related; this helps identify groups of constitution sections that are and aren't related to the candidate topic.

In the search above, constitution sections in clusters 1, 2, 4, and 5 relate to energy policy, electricity and gas, nuclear energy, and renewable energy, respectively, and thus align with the candidate topic of *energy independence and energy policy*. Constitution sections in clusters 0 and 3, however, relate to monetary policy and general state independence, and thus are not relevant to the candidate topic. The “unclustered” group above includes constitution sections that are potentially related to the candidate topic but not to other constitution sections; these must be reviewed individually. For example, unclustered segments on fossil fuels and recycled energy are relevant to the candidate topic, but those on financial policies and financial independence are not.

Review Your Results

- Review your results in “Step 3: Search national constitutions” to assess if any changes are needed to your search parameters to produce an optimal set of matches.
- If a search produces few matches, lower the search threshold in Step 1 (and remember to click “Apply Choices” in Step 1) and re-run Step 3. You should keep lowering it as long as the lowered threshold continues to produce additional good matches and starts to produce incorrect matches. Once lowering the search threshold *only* produces more incorrect matches, that's a sign that the search threshold is low enough.
- If a search produces few *good* matches, try rephrasing or reformulating the candidate topic. This could include: trying broad vs specific search terms, reordering the search terms to put the most important part of the topic first, elaborating the topic with a description, or using existing matches or known constitutional

provisions on this topic to help think through different language that might capture more constitution sections on the topic. Be sure your topic formulation doesn't have extraneous or contradictory text that could throw off your search results. Re-run Steps 1-3 with these new formulations to see which captures the most accurate matches.

13. If a search produces good matches *with distinct clusters*, try each cluster topic as a candidate topic. It may be that cluster topics could provide good sub-topics or related topics to capture constitution sections in a more nuanced way. Re-run Steps 1-3 with these cluster topics to see which captures the most accurate matches.
14. Where applicable, use the CCP Characteristics of National Constitutions data (aka "CCP data") to check that the results are capturing instances of the topic we *know* are in constitutions (since the CCP data are tagged by humans).

For example, since the new *recall* topic aims to capture "voter recall of elected officials," we can use CCP topics that capture some aspects of voter recall to check that our *recall* topic formulation captures provisions tagged by those existing CCP topics. The existing CCP topic tracking who can propose to dismiss the head of state (*hospdiss*) includes an option for the public to propose to dismiss the head of state (*hospdiss_5*). This is shown in the CCP codebook as follows:

```
v98. Who can propose a dismissal of the Head of State? (Asked only if HOSDISS is answered
1)
1. Head of Government – [HOSPDISS_1]
2. First (or only) chamber of the Legislature – [HOSPDISS_2]
3. Second Chamber of the Legislature – [HOSPDISS_3]
4. Both Chambers of the Legislature are required – [HOSPDISS_4]
5. Public (e.g., recall) – [HOSPDISS_5]
6. Public Prosecutor – [HOSPDISS_6]
7. Court – [HOSPDISS_7]
8. Constitutional Council or Constitutional Court – [HOSPDISS_8]
9. the Government/Cabinet – [HOSPDISS_9]
90. left explicitly to non-constitutional law – [HOSPDISS_90]
96. other, please specify in the comments section – [HOSPDISS_96]
97. Unable to Determine – [HOSPDISS_97]
98. Not Specified – [HOSPDISS_98]
```

We can then look for *hospdiss_5* in the CCP data for the most recent year available (2021) to see which countries have such a provision allowing the public to propose to dismiss the head of state. In the CCP data, this is shown as a 1 if the provision is present, and a 0 if not. There are only four countries with a 1 for this variable:

country	year	hospdiss_5
Bolivia	2021	1
Ecuador	2021	1
Liechtensteir	2021	1
Venezuela	2021	1

Then we can check if these provisions in the constitutions of Bolivia, Ecuador, Liechtenstein, and Venezuela came up in the search results produced by the ontology tool. If not, we could lower the search threshold and/or revise the topic formulation used in the ontology tool to try to capture these provisions, perhaps by adding language similar to these constitutional provisions.

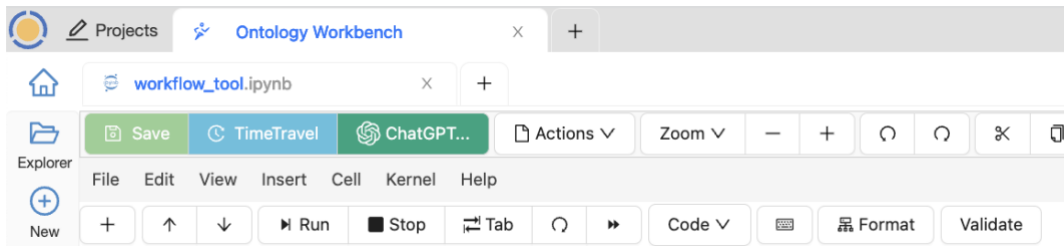
We can then do the same thing for other CCP topics that capture other aspects of voter recall, including *hosadiss_5*, *hogpdiss_6*, *hogadiss_5*, *cabdiss_7*, *rempro_1*, and *jrempro_9*.

We don't have CCP topics that are applicable to every proposed new topic. But, for those we do have, we want to be sure we use them to make the new topics as robust and complete as possible.

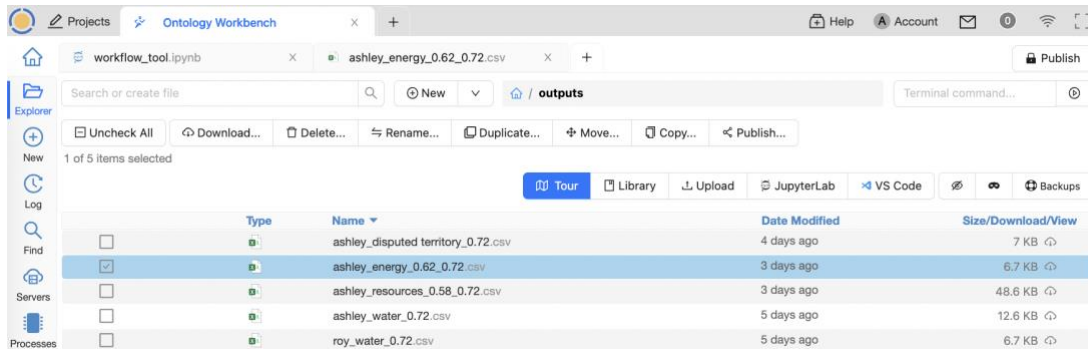
15. Repeat any or all of these steps to iteratively work toward what you feel is the best set of search results.

Download Your Search Results

16. Once you settle on your candidate topic text, and have completed Steps 1-3, download the results you exported in Step 1. To do so, click on the "Explorer" folder in the navigation menu (shown below).



17. That will show the Ontology Workbench folders/files menu you saw when first opening the workbench. In the Name column, click on “outputs.” Within the outputs folder, check the box next to your exported file (shown below). The exported file will be named as “user ID_export name_search threshold_cluster threshold.” Click download in the nav menu, then click download again when prompted.



18. To return to the ontology workflow tool, click the “workflow_tool.ipynb” tab near the top of the screen.

Record and Upload Your Analysis

19. In a Google doc or in a row of the topic file, explain why you recommend or don't recommend adding this topic to the CCP ontology. This should briefly note (i) whether you recommend adding this as a new topic in the CCP ontology, (ii) the closest existing CCP topics, (iii) why this new topic tag is/isn't useful, (iv) which clusters/issues in your results should/shouldn't receive that topic tag, (v) possible new subtopics that might fit under the new topic (if applicable), (vi) changes you'd propose to the CCP ontology to accommodate the new topic (if applicable), and (vii) what other topic formulations you tried.
20. Name your exported topic file as: name_topic_search threshold_cluster threshold_v1 (or whichever version it is). If you wrote your explanatory notes in a Google doc, name the file as: name_topic_notes_v1 (or whichever version it is).
21. Upload your exported topic file and explanatory notes in the submission folder for your topic.
22. Discuss your topic formulation, results, and analysis with the full team. Together we'll decide next steps before moving to final steps on the topic. In particular, this discussion will decide which types of provisions to **include** and **not include** in the set of provisions covered by this topic tag. This will be key in annotating your results in the next step.

Annotate and Re-upload Your Exported Topic File

23. Once your topic is cleared for the final steps, complete the final analysis in your exported topic file (shown below).

The exported csv shows your search results in the following format. Rows at the top note your **user** name, candidate **topic**, **search threshold**, **number of results** from the search, **cluster threshold**, **number of clusters**, and **number of unclustered** constitution sections. Columns show the **cluster ID** (group of text sections), **section text** (constitutional text matching your candidate topic), **section ID** (section number of the matching constitutional text), **constitution name** (country and year of the matching constitutional text), **link** (deep link to that constitutional text on Constitute), and **accept result** (where you'll note whether or not to accept the text matched to your candidate topic).

User	ashley					
Topic	energy independence and energy policy					
Search threshold	0.62					
Number of results	63					
Cluster threshold	0.72					
Number of clusters	6					
Number unclustered	29					
Cluster ID	Section text	Section ID	Constitution name	Link	Accept result	Notes
0	ensures realization of the financial credit and monetar	950	Azerbaijan_2016	https://www.constituteproject.org/constitution/Azerbaijan_2016.html	No	
0	the definition and the implementation of monetary pol	1019	Democratic_Republic_of_Congo_2002	https://www.constituteproject.org/constitution/Democratic_Republic_of_Congo_2002.html	No	
0	determine monetary policy; and	3113	Zambia_2016	https://www.constituteproject.org/constitution/Zambia_2016.html	No	
0	to formulate and implement monetary policy.	4406	Zimbabwe_2017	https://www.constituteproject.org/constitution/Zimbabwe_2017.html	No	
0	formulating monetary policy;	2111	South_Sudan_2013	https://www.constituteproject.org/constitution/South_Sudan_2013.html	No	
0	the harmonization of the economical, financial and mo	785	Niger_2017	https://www.constituteproject.org/constitution/Niger_2017.html	No	
1	Energy policy.	4480	Uganda_2017	https://www.constituteproject.org/constitution/Uganda_2017.html	Yes	
1	energy policy.	5059	Kenya_2010	https://www.constituteproject.org/constitution/Kenya_2010.html	Yes	
1	energy policy;	2190	Sweden_2012	https://www.constituteproject.org/constitution/Sweden_2012.html	Yes	
2	Energy policy including electricity and gas reticulation	5075	Kenya_2010	https://www.constituteproject.org/constitution/Kenya_2010.html	Yes	
2	electricity and gas reticulation and energy regulation.	5177	Kenya_2010	https://www.constituteproject.org/constitution/Kenya_2010.html	Yes	
2	energy (including electricity and power generation and	4521	Papua_New_Guinea_2002	https://www.constituteproject.org/constitution/Papua_New_Guinea_2002.html	Yes	
2	Energy and hydro electricity	3925	Zambia_2016	https://www.constituteproject.org/constitution/Zambia_2016.html	Yes	
2	electricity and the establishment of electric power stat	4112	Nigeria_2011	https://www.constituteproject.org/constitution/Nigeria_2011.html	Yes	
2	electricity and the establishment in that State of electr	4124	Nigeria_2011	https://www.constituteproject.org/constitution/Nigeria_2011.html	Yes	
2	Electricity; gas and gas works; and other works for the	3425	Malaysia_2007	https://www.constituteproject.org/constitution/Malaysia_2007.html	Yes	
3	The Andorran People, with full liberty and independenc	2	Andorra_1993	https://www.constituteproject.org/constitution/Andorra_1993.html	No	
3	the defence of independence and sovereignty;	49	Mozambique_2007	https://www.constituteproject.org/constitution/Mozambique_2007.html	No	
3	the independence, the sovereignty and the unity of the	2190	Mozambique_2007	https://www.constituteproject.org/constitution/Mozambique_2007.html	No	
3	consolidating and strengthening the independence and	3	Mongolia_2001	https://www.constituteproject.org/constitution/Mongolia_2001.html	No	

First, use the team’s decisions on which types of provisions to include and not include to note which constitution sections should receive the topic tag in the “accept results” column. The “accept result” column auto-populates with “yes.” For clusters or individual constitution sections that match your candidate topic, leave the “yes” in this column. For clusters or constitution sections that do not match your candidate topic, change this to “no” (shown in red above). For those where you’re not sure, change this to “maybe.”

We will use your line-by-line annotation to automatically apply an accepted new topic to these constitution sections in constitutions shown on [Constitute](#). If you need more context on a constitution section to determine if it matches the candidate topic or not, click on the section link on that row of the spreadsheet.

Second, add rows for the **topic tag** and **topic label** (as shown below). You’ll get this information from the CCP ontology assignments sheet. You can deviate from what is noted on the assignments sheet if needed, but just flag this so we can make those changes across other platforms.

Third, change the “topic” row to instead be labeled **topic semantic search text** (as shown below). This is the *internal* topic formulation you used in the ontology tool to find constitutional provisions. We will retain this search text so we can replicate the results in the future as needed.

Fourth, add a row for the **topic description** (as shown below). This is the *public-facing* topic formulation we’ll use to describe the topic on Constitute. It will be very similar to the topic semantic search text, but it could add additional text that is useful to Constitute users but was excluded from the topic semantic search text since it distracts the algorithm.

Lastly, add a row for **CCP notes** that lists the types of provisions that we INCLUDE and DO NOT INCLUDE under this topic tag.

User	catalina						
Topic tag	relsup						
Topic label	Public support for religion						
Topic description	Guarantees or allows public support for religion or religious institutions like houses of worship, schools, and political parties. This includes governmental recognition, assistance, or funding.						
Topic semantic search text	Public support for religion, religious institutions, religious purposes. Allows for government funding of religious institutions, including churches and K-12 schools. Secular.						
Search threshold	0.65						
Number of results	63						
Cluster threshold	0.7						
Number of clusters	3						
Number unclustered	12						
Intern notes	Recommend adding as a new CCP topic. The closest current CCP topic discusses the separation of church and state, and it has a relatively strong score (.69). The next closest topic includes tax status of religious organizations (.65). This new topic would apply to *20 constitution sections (4 results were "maybe") and is geographically distributed across Africa, Asia, Europe, Oceania, and South America. Adding key terms like "secular" and "religious purposes" , along with the entire government funding phrase, yielded more precise constitution sections. This is also an extremely relevant and debated topic, so it is necessary to be included in the ontology.						
CCP notes	INCLUDE references to recognizing, assisting, or funding any religious institutions. DO NOT INCLUDE segments on secular public education, separation of church and state, freedom of religious institutions, and tax-exempt status, as these are captured by existing CCP topics (seprel, freerel, rellax). DO NOT INCLUDE bans on public support for religion/religious institutions since that is also covered by topic description for seprel. For any constitution segments where the context (and thus the exact meaning) is not clear in the segment itself (e.g., when the segment is part of a list and the whole list is not included), look up the segment in Constitute to be sure your yes/no/maybe result accounts for the complete context/meaning of the segment.						
Cluster ID	Section text	Section ID	Constitution name	Accept result	Notes		
	0 Public education is secular.	102	Burkina_Faso_2015	No	seprel		
	0 Public education is secular and free of charge.	125	Chad_2018	No	seprel, edfree		
	0 Religious denominations have the right to establish	295	Kosovo_2016	No	freerel		
	0 Every religious community or denomination, and ev	516	Fiji_2013	Yes			
	0 Churches and religious communities shall be equal	175	Serbia_2006	No	freerel		
	0 Charitable institutions, churches and parsonages o	254	Philippines_1987	No	rellax		
	0 The Macedonian Orthodox Church, as well as the I	538	Macedonia_2011	No	freerel		
	0 Public funds shall be allocated to public schools, ar	4154	Brazil_2017	Yes			

24. Upload your edited csv and explanatory doc to the submissions folder for your topic.