(v. 2) Module 4.a - Compare sets of topics (multiple topics)

Start of Block: INFO ABOUT YOU & EXERCISE

# IS THIS A TRIAL?

Before you begin, please check the choice below if this is just a trial or experimental exercise that should be ignored in collecting outputs from this module.

[ ]  Check this field if this is just a trial.

# OPEN RESEARCH POLICY STATEMENT

The modules of the WE1S Interpretation Protocol create records of research that support the project's aim of advancing "open, generalizable, and replicable digital humanities methodology" (see project ["About"](http://we1s.ucsb.edu/about/) statement). This aim is related to current practices of ["open science"](https://en.wikipedia.org/wiki/Open_science) and ["open-notebook science."](https://en.wikipedia.org/wiki/Open-notebook_science) The basic idea is that the materials, methods, steps, and findings of research should be transparent so that other researchers know how conclusions were reached and can test or reproduce the process--an ideal that is especially needed for understanding such machine-learning methods as topic modeling (see [WE1S bibliography on machine-learning interpretability](https://we1s.ucsb.edu/research/we1s-bibliography/bibliography-interpretability-and-explainability/)).

In completing this and other modules of the WE1S Interpretation Protocol you are creating an "open notebook" that in principle can be shared with others as a public record of research. *Please be sure that all descriptions, observations, and notes you write in this module are consistent with being part of public record--that is, appropriate in tone, respectful to others, and not in violation of privacy or copyright restrictions*.

*However please also do not let the ideal of open-notebook science inhibit you from speculative or tentative in-progress commentary*. It is understood that research, open or otherwise, is a *process* of finding knowledge with many stages and levels of confidence. (At the conclusion of this and other modules of the Interpretation Protocol, you will be asked to rate your confidence level in the current exercise.)

*In addition, lab notebooks are allowed to be messy as a part of the research process.* It is not expected that notes written into these modules be copyedited as if for formal publication. (The time for polishing is later when writing up a report synthesizing results from one or more modules of this Interpretation Protocol.)

# EXERCISE INFO

**Please enter the following the following information, which will generate an "Exercise ID".** (Tab from input field to input field to force calculation of the Exercise ID.)

* **Your name or team name**. (If team, then use the format "Team1", "Team2", etc. Please do not use a space): **[Name]**
* **Interpretation project number** (e.g., "4," indicating that this is the fourth interpretation project for you or your team that the exercise with this module is part of): **[Project#]**
* **Which stage of your present interpretation project does this module represent?** (E.g., if this module is the second you have used in the sequence of modules for the same interpretation project, then write the number "2" here. If multiple team members are working in parallel using various modules, then follow the convention of adding name-abbreviation—e.g., 2AL: **[Stage#]**
* **The *present* Interpretation Protocol module you are using is set as follows**: **4a**(This is part of your auto-generated exercise ID below, and cannot be altered in the present module document.)
* **Today's date in the format YYYY-MM-DD (e.g., 2020-07-06)** (use the date-picker calendar): **[Date]**

**Your team folder in the project Google team working space.**
This refers to your team working folder for notes, reports, materials, etc. related to interpretation work in the WE1S Google Team drive’s [\_Team Working Space](https://drive.google.com/drive/u/1/folders/1er05JpagMNlfdnMIvcCVgwbweGkGeSua) folder. Go there to find the folder your team created for your work.

**[FolderURL]**

# Exercise ID

The ID for the present exercise (generated from the information you just provided) is as follows. (If you see a mistake, please go "back" in this survey and correct the information that generated the Exercise ID.)

[Name]**\_** [Project#]\_ [Stage#]\_4a\_ [Date]

# Exercise Description

To assist in remembering what the exercise using this module is about, please enter below **a very brief description of what you are currently doing** (e.g., "We are now going to compare the keywords 'humanities' and 'sciences' in the 20190621\_2132\_us-humanities-top-newspapers---topics200 model.")

Click or tap here to enter text.

# Topic model you are studying

For the topic model you are studying, please enter below its **"topic model name"** and "**start page"** URL. A **"topic model name"** looks like this: *"20190621\_2132\_us-humanities-top-newspapers---topics200"*. You can locate this name in the [Registry of WE1S Topic Models](https://docs.google.com/spreadsheets/d/1n9nCzHut5aZEYRiSx5QD_yiiaLO9TeeQgwJJPd1ATt0/edit#gid=0). Note that this is the name of the specific granularity of the model you are studying (number of topics). The **"start page"** is the URL of the overall topic model (including all its granularities and their visualizations) listed on this page on the Harbor 10002 server: [Index of / projects/](http://harbor.english.ucsb.edu:10002/projects/). (You can also find this URL in the Registry of WE1S Topic Models.) A start page URL looks like this: http://harbor.english.ucsb.edu:10002/projects/20190621\_2132\_us-humanities-top-newspapers/

**Topic model name**: Click or tap here to enter text.

**"Start page" URL**: Click or tap here to enter text.

# Research question you are addressing

Please enter below information about the research question you are addressing in this exercise. You can find this information in the "Registry of WE1S Research Questions." (If you are taking an overview of a model prior to working on a specific research question, just enter "Taking an overview")

**Research question**:

Click or tap here to enter text.

**Operationalized form of question**:

Click or tap here to enter text.

**Research question ID** (use the format "team2-q1", representing in this example team 2's first research question): Click or tap here to enter text.

# Previous Interpretation Protocol modules you used to address your research question (if any) [optional]

Enter in sequence below any Interpretation Protocol modules you have already completed in previous steps of your current interpretation project. For example, if for steps 1-3 preceding this exercise (where the current exercise is now step 4 in the sequence) you previously used modules 3a, 3b, and 3c in that order, then enter those module numbers in that sequence.

 **Previous module sequence**: Click or tap here to enter text.

End of Block: INFO ABOUT YOU & EXERCISE (v. 2, created 9 June 2019, rev. 7 July 2019)

Start of Block: WHAT THIS MODULE DOES

What module 4.a does

(Module 4.a, v. 2, created 22 June 2019; last rev. 23 June 2019.)

The purpose of module 4.a of the WE1S topic model Interpretation Protocol is to study one or more *sets* of topics (i.e., mulltiple topics together). The "sets" are ones you demarcate and populate with representative topics according to some criteria that makes sense for your research question. Some examples might be: one set of topics related to "humanities"; one set of topics in which both the words "humanities" and "science" are prominent; two different sets of topics, where one is about "humanities" and the other is about "science"; etc.

 Note that a *possible* use of this module is to compare different source categories in your corpus (i.e., sets of topics associated with distinct publication sources--e.g, one set of topics from mainstream newspapers, and another from student newspapers). However, if the source categories you are interested in are ones for which WE1S has pre-assigned metadata labels (e.g., a tag for all "student newspapers"), you would do better instead to use Interpretation Protocol module 5 (under construction), which is designed for comparing such pre-tagged source sets. (However, there are a limited number of such pre-categorized tagged sets for use in module 5.)

End of Block: WHAT THIS MODULE DOES

Start of Block: IDENTIFY SETS OF TOPICS

Step 1. Identify the set, or sets, of topics you are studying.

In this module, a "set" of topics is defined as any conceptual grouping of multiple topics you are interested in.

* A set can be conceptually ***homogeneous***--e.g., focused on similar topics (e.g., politics) or on topics associated with one part of the corpus (e.g., a particular group of media sources or time range).
* Or a set can be conceptually ***heterogeneous***--e.g., including topics on two issues (e.g., science, humanities) or related to two different sources. What makes a group of heterogeneous topiics a single "set" is that their issues, sources, or features are joined by some common factor (e.g., all "science" and "humanities" topics *above a threshold of proportional weight*, *sharing specific top keywords*, *sourced from a particular publication*, etc.).

You can study up to four sets of topics in this module (in combinations of homogeneous and heterogeneous sets as needed). If you are working with more than one set, then your summary observations at the end of this exercise will give you a chance to think about them together.

**Label the set(s) of topics you want to study**.

Below, identify by means of a descriptive label (a few words at most) each set of topics you want to study (up to a maximum of 4 sets). Examples might be, one set of topics on "science and humanities," or two different sets on "science" and "humanities.":

|  |  |
| --- | --- |
|  | **Labels for set(s) of topics you want to study:** |
| Set 1: |  |
| Set 2: |  |
| Set 3: |  |
| Set 4: |  |

**Criteria for choosing topics that make up your set(s).**

For each set that you identified, please explain your criteria for forming the set by answering the relevant question below (only one answer per set).

|  |  |  |
| --- | --- | --- |
|  | **HOMOGENEOUS TOPICS**:You want to look at homogeneous topics that are related to one or more *individual* concepts or parts of the corpus (e.g. "politics"). | **HETEROGENEOUS TOPICS**:You want to look at heterogeneous topics representing similarities or contrasts between two concepts (e.g., "politics" \_and\_ universities") that share some common factor (e.g., top words, weight, source, date, etc.). |
| **Set 1** |[ ] [ ]
| **Set 2** |[ ] [ ]
| **Set 3** |[ ] [ ]
| **Set 4** |[ ] [ ]

**Now you will start examining the set(s) of topics you identified.** Going to the next section will advance you to the exploration of the first set (Set 1) you designated. If you designated multiple sets, this module will iterate through those before bringing you to the conclusion of this interpretation exercise.

The analysis of each set branches into either an analysis of “homogeneous” or “heterogeneous” topics (according to the definitions above), depending on which kind of analysis is your goal. You should only fill out *one* of the branches for a set—*either* for “homogeneous” or “heterogeneous” topics. (*Note*: In the Qualtrics survey version of this Interpretation Protocol module, conditional branching log

ic directed users automatically to different parts of the survey depending on the choices they designated for the sets they were examining and whether they were studying homogeneous or heterogeneous topics.)

End of Block: IDENTIFY SETS OF TOPICS

Start of Block: SET 1 -- HOMOGENEOUS

# SET 1 (HOMOGENEOUS TOPICS) --

# Examine homogeneous topics in Set 1

*Reminder*: You have identified **Set 1** as conceptually **homogeneous**--meaning that it is focused on similar topics (e.g., politics) or on topics associated with one part of the corpus (e.g., a particular group of media sources or time range).

(*Note*: First use the cluster-analysis tools in the Topic Model Observatory (see [TMO Guide](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/)) to see if **Set 1** may actually be a topic "cluster." If so, you may wish to examine **Set 1** using Interpretation Protocol 3.b for topic clusters instead of the present module.)

|  |
| --- |
|  |

**List below by their topic numbers up to 10 topics that you think can serve as representative of topic set 1.** This will require you to search for topics based on one or more criteria (e.g., association with a keyword, association with a source, proportional weight in the corpus, coherence of top keywords, etc.) and to engage in some cursory examination of candidate topics.

Use the general-purpose and special-purpose visualization interfaces in the Topic Model Observatory as needed to examine and select topics. (For example, recall that TopicBubbles is especially useful for viewing together on a single screen the words (and other info) of several topics at once. (See *TMO Guide*, [chapter 2](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/), section 3 for instructions on this feature of TopicBubbles.)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Topic # (1) | Topic # (2) | Topic # (3) | Topic # (4) | Topic # (5) | Topic # (6) | Topic # (7) | Topic # (8) | Topic # (9) | Topic # (10) |
| **Topics representative of topic set 1:**  | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. |

**Make a note about the grounds on which you chose these topics as representative of Set 1 (homogeneous)**:

**Running note for Set 1** (live editable at any point where this reappears in this Word document)

Enter in the field below your notes on Set 1, focusing on the grounds for which you chose these topics as representative of the set:

[RunningNote1]

**After examining the topics you identified as representative of Set 1 (homogeneous), please label and add any brief notes you wish for each topic**. (A topic label should be very brief, typically 1-3 words, and cannot include punctuation.)

|  |  |  |
| --- | --- | --- |
| **Topic #** | **Label for topic** | **Notes on topic** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
|  |

**Please write an observation on Set 1 (homogeneous) of topics by adding to the bottom of your running note on this set**. An “observation” should have two parts: a more-or-less objective description in which you have high confidence; and a thesis, hypothesis, interpretation, or suggestion for which you have less confidence before you study the model more closely but that might bear on your research question.

**Running note for Set 1** (live editable at any point where this reappears in this Word document)

Add your observations to the bottom of the note on set 1 you previously started:

[RunningNote1]

**This concludes your examination of Set 1 (“homogeneous” topics analysis). If you designated more sets of topics, continue to the Set 2 analysis. (Do not complete the Set 1 “heterogeneous” analysis, since you would logically be doing either a homogeneous or heterogeneous analysis, not both.) If you have completed the sets you wish to examine, then skip all subsequent parts of this module and go directly to the CONCLUDING OBSERVATIONS STEP in this module.**

End of Block: SET 1 -- HOMOGENEOUS

Start of Block: SET 1 -- HETEROGENEOUS

# SET 1 (HETEROGENEOUS TOPICS) --

Examine heterogeneous topics in Set 1

*Reminder*: You have identified **Set 1** as conceptually **heterogeneous**--e.g., including topics on two issues (e.g., science, humanities) or related to two different sources. What makes a group of heterogeneous topics a single "set" is that their issues, sources, or features are joined by some common factor (e.g., all "science" and "humanities" topics above a threshold of proportional weight, sharing specific top keywords, sourced from a particular publication, etc.).

A heterogeneous set of topics may be thought of as being associated with two different categories of issues, parts of the corpus, etc. For example, the issues of such a set could be categorized as "humanities," on the one hand, and "science," on the other.

**Identify through a brief label the two categories of issues or parts of the corpus in this set**.

|  |  |
| --- | --- |
|  | **Brief label for your categories**: |
| **Category 1** of issues or parts of corpus: |  |
| **Category 2** of issues or parts of corpus: |  |

List below by their topic numbers up to 10 topics that you think can serve as representative of topic **set 1 (heterogeneous)**. Also identify the category of the set they are associated with (or if you think they are intermediary between the categories). This will require you to search for topics based on one or more criteria (e.g., association with a keyword, association with a source, proportional weight in the corpus, coherence of top keywords, etc.) and some cursory examination of candidate topics.

Use the general-purpose and special-purpose visualization interfaces in the Topic Model Observatory as needed to examine and select topics. (For example, recall that TopicBubbles is especially useful for viewing together on a single screen the words (and other info) of several topics at once. (See *TMO Guide*, [chapter 2](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/), section 3 for instructions on this feature of TopicBubbles.)

|  |  |  |
| --- | --- | --- |
| **Set 1 Topics** | **TOPIC #** | **CATEGORY:**Enter “1” or “2” for the categories you identified above. (On enter “1.5” for topics you think are intermediate between you categories” |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |

Make a note about the grounds on which you chose these topics as representative of **Set 1 (heterogeneous)**:

**Running note for Set 1** (live editable at any point where this reappears in this Word document)

Enter in the field below your notes on Set 1, focusing on the grounds for which you chose these topics as representative of the set:

[RunningNote1]

**After examining the topics you identified as representative of Set 1 (heterogeneous), please label and add any brief notes you wish for each topic**. (A topic label should be very brief, typically 1-3 words, and cannot include punctuation.)

|  |  |  |
| --- | --- | --- |
| **Topic #** | **Label for topic** | **Notes on topic** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
|  |

**Please write an observation on Set 1 (heterogeneous) of topics by adding to the bottom of your running note on this set**. An “observation” should have two parts: a more-or-less objective description in which you have high confidence; and a thesis, hypothesis, interpretation, or suggestion for which you have less confidence before you study the model more closely but that might bear on your research question.

**Running note for Set 1** (live editable at any point where this reappears in this Word document)

Add your observations to the bottom of the note on set 1 you previously started:

[RunningNote1]

**This concludes your examination of Set 1 (“heterogeneous” topics analysis). If you designated more sets of topics, continue to the Set 2 analysis. If you have completed the sets you wish to examine, then skip all subsequent parts of this module and go directly to the CONCLUDING OBSERVATIONS STEP near the end of this module.**

End of Block: SET 1 -- HETEROGENEOUS

Start of Block: SET 2 -- HOMOGENEOUS

# SET 2 (HOMOGENEOUS TOPICS) --

# Examine homogeneous topics in Set 2

*Reminder*: You have identified **Set 2** as conceptually **homogeneous**--meaning that it is focused on similar topics (e.g., politics) or on topics associated with one part of the corpus (e.g., a particular group of media sources or time range).

(*Note*: First use the cluster-analysis tools in the Topic Model Observatory (see [TMO Guide](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/)) to see if **Set 2** may actually be a topic "cluster." If so, you may wish to examine **Set 2** using Interpretation Protocol 3.b for topic clusters instead of the present module.)

|  |
| --- |
|  |

**List below by their topic numbers up to 10 topics that you think can serve as representative of topic set 2.** This will require you to search for topics based on one or more criteria (e.g., association with a keyword, association with a source, proportional weight in the corpus, coherence of top keywords, etc.) and to engage in some cursory examination of candidate topics.

Use the general-purpose and special-purpose visualization interfaces in the Topic Model Observatory as needed to examine and select topics. (For example, recall that TopicBubbles is especially useful for viewing together on a single screen the words (and other info) of several topics at once. (See *TMO Guide*, [chapter 2](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/), section 3 for instructions on this feature of TopicBubbles.)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Topic # (1) | Topic # (2) | Topic # (3) | Topic # (4) | Topic # (5) | Topic # (6) | Topic # (7) | Topic # (8) | Topic # (9) | Topic # (10) |
| **Topics representative of topic set 2:**  | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. |

**Make a note about the grounds on which you chose these topics as representative of Set 2 (homogeneous)**:

**Running note for Set 2** (live editable at any point where this reappears in this Word document)

Enter in the field below your notes on Set 2, focusing on the grounds for which you chose these topics as representative of the set:

[RunningNote2]

**After examining the topics you identified as representative of Set 2 (homogeneous), please label and add any brief notes you wish for each topic**. (A topic label should be very brief, typically 1-3 words, and cannot include punctuation.)

|  |  |  |
| --- | --- | --- |
| **Topic #** | **Label for topic** | **Notes on topic** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
|  |

**Please write an observation on Set 2 (homogeneous) of topics by adding to the bottom of your running note on this set**. An “observation” should have two parts: a more-or-less objective description in which you have high confidence; and a thesis, hypothesis, interpretation, or suggestion for which you have less confidence before you study the model more closely but that might bear on your research question.

**Running note for Set 2** (live editable at any point where this reappears in this Word document)

Add your observations to the bottom of the note on set 2 you previously started:

[RunningNote2]

**This concludes your examination of Set 2 (“homogeneous” topics analysis). If you designated more sets of topics, continue to the Set 3 analysis. (Do not complete the Set 2 “heterogeneous” analysis, since you would logically be doing either a homogeneous or heterogeneous analysis, not both.) If you have completed the sets you wish to examine, then skip all subsequent parts of this module and go directly to the CONCLUDING OBSERVATIONS STEP in this module.**

End of Block: SET 2 -- HOMOGENEOUS

Start of Block: SET 2 -- HETEROGENEOUS

# SET 2 (HETEROGENEOUS TOPICS) --

Examine heterogeneous topics in Set 2

*Reminder*: You have identified **Set 2** as conceptually **heterogeneous**--e.g., including topics on two issues (e.g., science, humanities) or related to two different sources. What makes a group of heterogeneous topics a single "set" is that their issues, sources, or features are joined by some common factor (e.g., all "science" and "humanities" topics above a threshold of proportional weight, sharing specific top keywords, sourced from a particular publication, etc.).

A heterogeneous set of topics may be thought of as being associated with two different categories of issues, parts of the corpus, etc. For example, the issues of such a set could be categorized as "humanities," on the one hand, and "science," on the other.

**Identify through a brief label the two categories of issues or parts of the corpus in this set**.

|  |  |
| --- | --- |
|  | **Brief label for your categories**: |
| **Category 1** of issues or parts of corpus: |  |
| **Category 2** of issues or parts of corpus: |  |

List below by their topic numbers up to 10 topics that you think can serve as representative of topic **set 2 (heterogeneous)**. Also identify the category of the set they are associated with (or if you think they are intermediary between the categories). This will require you to search for topics based on one or more criteria (e.g., association with a keyword, association with a source, proportional weight in the corpus, coherence of top keywords, etc.) and some cursory examination of candidate topics.

Use the general-purpose and special-purpose visualization interfaces in the Topic Model Observatory as needed to examine and select topics. (For example, recall that TopicBubbles is especially useful for viewing together on a single screen the words (and other info) of several topics at once. (See *TMO Guide*, [chapter 2](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/), section 3 for instructions on this feature of TopicBubbles.)

|  |  |  |
| --- | --- | --- |
| **Set 2 Topics** | **TOPIC #** | **CATEGORY:**Enter “1” or “2” for the categories you identified above. (On enter “1.5” for topics you think are intermediate between you categories” |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |

Make a note about the grounds on which you chose these topics as representative of **Set 2 (heterogeneous)**:

**Running note for Set 2** (live editable at any point where this reappears in this Word document)

Enter in the field below your notes on Set 2, focusing on the grounds for which you chose these topics as representative of the set:

[RunningNote2]

**After examining the topics you identified as representative of Set 2 (heterogeneous), please label and add any brief notes you wish for each topic**. (A topic label should be very brief, typically 1-3 words, and cannot include punctuation.)

|  |  |  |
| --- | --- | --- |
| **Topic #** | **Label for topic** | **Notes on topic** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
|  |

**Please write an observation on Set 2 (heterogeneous) of topics by adding to the bottom of your running note on this set**. An “observation” should have two parts: a more-or-less objective description in which you have high confidence; and a thesis, hypothesis, interpretation, or suggestion for which you have less confidence before you study the model more closely but that might bear on your research question.

**Running note for Set 2** (live editable at any point where this reappears in this Word document)

Add your observations to the bottom of the note on set 2 you previously started:

[RunningNote2]

**This concludes your examination of Set 2 (“heterogeneous” topics analysis). If you designated more sets of topics, continue to the Set 3 analysis. If you have completed the sets you wish to examine, then skip all subsequent parts of this module and go directly to the CONCLUDING OBSERVATIONS STEP near the end of this module.**

End of Block: SET 2 -- HETEROGENEOUS

Start of Block: SET 3 -- HOMOGENEOUS

# SET 3 (HOMOGENEOUS TOPICS) --

# Examine homogeneous topics in Set 3

*Reminder*: You have identified **Set 3** as conceptually **homogeneous**--meaning that it is focused on similar topics (e.g., politics) or on topics associated with one part of the corpus (e.g., a particular group of media sources or time range).

(*Note*: First use the cluster-analysis tools in the Topic Model Observatory (see [TMO Guide](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/)) to see if **Set 3** may actually be a topic "cluster." If so, you may wish to examine **Set 3** using Interpretation Protocol 3.b for topic clusters instead of the present module.)

|  |
| --- |
|  |

**List below by their topic numbers up to 10 topics that you think can serve as representative of topic set 3.** This will require you to search for topics based on one or more criteria (e.g., association with a keyword, association with a source, proportional weight in the corpus, coherence of top keywords, etc.) and to engage in some cursory examination of candidate topics.

Use the general-purpose and special-purpose visualization interfaces in the Topic Model Observatory as needed to examine and select topics. (For example, recall that TopicBubbles is especially useful for viewing together on a single screen the words (and other info) of several topics at once. (See *TMO Guide*, [chapter 2](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/), section 3 for instructions on this feature of TopicBubbles.)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Topic # (1) | Topic # (2) | Topic # (3) | Topic # (4) | Topic # (5) | Topic # (6) | Topic # (7) | Topic # (8) | Topic # (9) | Topic # (10) |
| **Topics representative of topic set 3:**  | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. |

**Make a note about the grounds on which you chose these topics as representative of Set 3 (homogeneous)**:

**Running note for Set 3** (live editable at any point where this reappears in this Word document)

Enter in the field below your notes on Set 3, focusing on the grounds for which you chose these topics as representative of the set:

[RunningNote3

**After examining the topics you identified as representative of Set 3 (homogeneous), please label and add any brief notes you wish for each topic**. (A topic label should be very brief, typically 1-3 words, and cannot include punctuation.)

|  |  |  |
| --- | --- | --- |
| **Topic #** | **Label for topic** | **Notes on topic** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
|  |

**Please write an observation on Set 3 (homogeneous) of topics by adding to the bottom of your running note on this set**. An “observation” should have two parts: a more-or-less objective description in which you have high confidence; and a thesis, hypothesis, interpretation, or suggestion for which you have less confidence before you study the model more closely but that might bear on your research question.

**Running note for Set 3** (live editable at any point where this reappears in this Word document)

Add your observations to the bottom of the note on set 3 you previously started:

[RunningNote3

**This concludes your examination of Set 3 (“homogeneous” topics analysis). If you designated more sets of topics, continue to the Set 4 analysis. (Do not complete the Set 3 “heterogeneous” analysis, since you would logically be doing either a homogeneous or heterogeneous analysis, not both.) If you have completed the sets you wish to examine, then skip all subsequent parts of this module and go directly to the CONCLUDING OBSERVATIONS STEP in this module.**

End of Block: SET 3 -- HOMOGENEOUS

Start of Block: SET 3 -- HETEROGENEOUS

# SET 3 (HETEROGENEOUS TOPICS) --

Examine heterogeneous topics in Set 3

*Reminder*: You have identified **Set 3** as conceptually **heterogeneous**--e.g., including topics on two issues (e.g., science, humanities) or related to two different sources. What makes a group of heterogeneous topics a single "set" is that their issues, sources, or features are joined by some common factor (e.g., all "science" and "humanities" topics above a threshold of proportional weight, sharing specific top keywords, sourced from a particular publication, etc.).

A heterogeneous set of topics may be thought of as being associated with two different categories of issues, parts of the corpus, etc. For example, the issues of such a set could be categorized as "humanities," on the one hand, and "science," on the other.

**Identify through a brief label the two categories of issues or parts of the corpus in this set**.

|  |  |
| --- | --- |
|  | **Brief label for your categories**: |
| **Category 1** of issues or parts of corpus: |  |
| **Category 2** of issues or parts of corpus: |  |

List below by their topic numbers up to 10 topics that you think can serve as representative of topic **set 3 (heterogeneous)**. Also identify the category of the set they are associated with (or if you think they are intermediary between the categories). This will require you to search for topics based on one or more criteria (e.g., association with a keyword, association with a source, proportional weight in the corpus, coherence of top keywords, etc.) and some cursory examination of candidate topics.

Use the general-purpose and special-purpose visualization interfaces in the Topic Model Observatory as needed to examine and select topics. (For example, recall that TopicBubbles is especially useful for viewing together on a single screen the words (and other info) of several topics at once. (See *TMO Guide*, [chapter 2](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/), section 3 for instructions on this feature of TopicBubbles.)

|  |  |  |
| --- | --- | --- |
| **Set 3 Topics** | **TOPIC #** | **CATEGORY:**Enter “1” or “2” for the categories you identified above. (On enter “1.5” for topics you think are intermediate between you categories” |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |

Make a note about the grounds on which you chose these topics as representative of **Set 3 (heterogeneous)**:

**Running note for Set 3** (live editable at any point where this reappears in this Word document)

Enter in the field below your notes on Set 3, focusing on the grounds for which you chose these topics as representative of the set:

[RunningNote3

**After examining the topics you identified as representative of Set 3 (heterogeneous), please label and add any brief notes you wish for each topic**. (A topic label should be very brief, typically 1-3 words, and cannot include punctuation.)

|  |  |  |
| --- | --- | --- |
| **Topic #** | **Label for topic** | **Notes on topic** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
|  |

**Please write an observation on Set 3 (heterogeneous) of topics by adding to the bottom of your running note on this set**. An “observation” should have two parts: a more-or-less objective description in which you have high confidence; and a thesis, hypothesis, interpretation, or suggestion for which you have less confidence before you study the model more closely but that might bear on your research question.

**Running note for Set 3** (live editable at any point where this reappears in this Word document)

Add your observations to the bottom of the note on set 3 you previously started:

[RunningNote3

**This concludes your examination of Set 3 (“heterogeneous” topics analysis). If you designated more sets of topics, continue to the Set 4 analysis. If you have completed the sets you wish to examine, then skip all subsequent parts of this module and go directly to the CONCLUDING OBSERVATIONS STEP near the end of this module.**

End of Block: SET 3 -- HETEROGENEOUS

Start of Block: SET 4 -- HOMOGENEOUS

# SET 4 (HOMOGENEOUS TOPICS) --

# Examine homogeneous topics in Set 4

*Reminder*: You have identified **Set 4** as conceptually **homogeneous**--meaning that it is focused on similar topics (e.g., politics) or on topics associated with one part of the corpus (e.g., a particular group of media sources or time range).

(*Note*: First use the cluster-analysis tools in the Topic Model Observatory (see [TMO Guide](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/)) to see if **Set 4** may actually be a topic "cluster." If so, you may wish to examine **Set 4** using Interpretation Protocol 3.b for topic clusters instead of the present module.)

|  |
| --- |
|  |

**List below by their topic numbers up to 10 topics that you think can serve as representative of topic set 4.** This will require you to search for topics based on one or more criteria (e.g., association with a keyword, association with a source, proportional weight in the corpus, coherence of top keywords, etc.) and to engage in some cursory examination of candidate topics.

Use the general-purpose and special-purpose visualization interfaces in the Topic Model Observatory as needed to examine and select topics. (For example, recall that TopicBubbles is especially useful for viewing together on a single screen the words (and other info) of several topics at once. (See *TMO Guide*, [chapter 2](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/), section 3 for instructions on this feature of TopicBubbles.)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Topic # (1) | Topic # (2) | Topic # (3) | Topic # (4) | Topic # (5) | Topic # (6) | Topic # (7) | Topic # (8) | Topic # (9) | Topic # (10) |
| **Topics representative of topic set 4:**  | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. | Enter topic # here. |

**Make a note about the grounds on which you chose these topics as representative of Set 4 (homogeneous)**:

**Running note for Set 4** (live editable at any point where this reappears in this Word document)

Enter in the field below your notes on Set 4, focusing on the grounds for which you chose these topics as representative of the set:

[RunningNote4]

**After examining the topics you identified as representative of Set 4 (homogeneous), please label and add any brief notes you wish for each topic**. (A topic label should be very brief, typically 1-3 words, and cannot include punctuation.)

|  |  |  |
| --- | --- | --- |
| **Topic #** | **Label for topic** | **Notes on topic** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
|  |

**Please write an observation on Set 4 (homogeneous) of topics by adding to the bottom of your running note on this set**. An “observation” should have two parts: a more-or-less objective description in which you have high confidence; and a thesis, hypothesis, interpretation, or suggestion for which you have less confidence before you study the model more closely but that might bear on your research question.

**Running note for Set 4** (live editable at any point where this reappears in this Word document)

Add your observations to the bottom of the note on set 4 you previously started:

[RunningNote4]

**This concludes your examination of Set 4 (“homogeneous” topics analysis). (Do not complete the Set 3 “heterogeneous” analysis, since you would logically be doing either a homogeneous or heterogeneous analysis, not both.) Go next to the CONCLUDING OBSERVATIONS STEP in this module.**

End of Block: SET 4 -- HOMOGENEOUS

Start of Block: SET 4 -- HETEROGENEOUS

# SET 4 (HETEROGENEOUS TOPICS) --

Examine heterogeneous topics in Set 4

*Reminder*: You have identified **Set 4** as conceptually **heterogeneous**--e.g., including topics on two issues (e.g., science, humanities) or related to two different sources. What makes a group of heterogeneous topics a single "set" is that their issues, sources, or features are joined by some common factor (e.g., all "science" and "humanities" topics above a threshold of proportional weight, sharing specific top keywords, sourced from a particular publication, etc.).

A heterogeneous set of topics may be thought of as being associated with two different categories of issues, parts of the corpus, etc. For example, the issues of such a set could be categorized as "humanities," on the one hand, and "science," on the other.

**Identify through a brief label the two categories of issues or parts of the corpus in this set**.

|  |  |
| --- | --- |
|  | **Brief label for your categories**: |
| **Category 1** of issues or parts of corpus: |  |
| **Category 2** of issues or parts of corpus: |  |

List below by their topic numbers up to 10 topics that you think can serve as representative of topic **set 4 (heterogeneous)**. Also identify the category of the set they are associated with (or if you think they are intermediary between the categories). This will require you to search for topics based on one or more criteria (e.g., association with a keyword, association with a source, proportional weight in the corpus, coherence of top keywords, etc.) and some cursory examination of candidate topics.

Use the general-purpose and special-purpose visualization interfaces in the Topic Model Observatory as needed to examine and select topics. (For example, recall that TopicBubbles is especially useful for viewing together on a single screen the words (and other info) of several topics at once. (See *TMO Guide*, [chapter 2](https://we1s.ucsb.edu/research/we1s-tools-and-software/topic-model-observatory/tmo-guide/tmo-guide-topicbubbles/), section 3 for instructions on this feature of TopicBubbles.)

|  |  |  |
| --- | --- | --- |
| **Set 4 Topics** | **TOPIC #** | **CATEGORY:**Enter “1” or “2” for the categories you identified above. (On enter “1.5” for topics you think are intermediate between you categories” |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |
| **Topic representative of topic set 1:** |  |  |

Make a note about the grounds on which you chose these topics as representative of **Set 4 (heterogeneous)**:

**Running note for Set 4** (live editable at any point where this reappears in this Word document)

Enter in the field below your notes on Set 4, focusing on the grounds for which you chose these topics as representative of the set:

[RunningNote4]

**After examining the topics you identified as representative of Set 4 (heterogeneous), please label and add any brief notes you wish for each topic**. (A topic label should be very brief, typically 1-3 words, and cannot include punctuation.)

|  |  |  |
| --- | --- | --- |
| **Topic #** | **Label for topic** | **Notes on topic** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
|  |

**Please write an observation on Set 4 (heterogeneous) of topics by adding to the bottom of your running note on this set**. An “observation” should have two parts: a more-or-less objective description in which you have high confidence; and a thesis, hypothesis, interpretation, or suggestion for which you have less confidence before you study the model more closely but that might bear on your research question.

**Running note for Set 4** (live editable at any point where this reappears in this Word document)

Add your observations to the bottom of the note on set 4 you previously started:

[RunningNote4]

**This concludes your examination of Set 4 (“heterogeneous” topics analysis). Go next to the CONCLUDING OBSERVATIONS STEP of this module.**

End of Block: SET 4 -- HETEROGENEOUS

Start of Block: CONCLUDING OBSERVATIONS

# CONCLUDING OBSERVATIONS:

# **Concluding steps of this exercise . . . .**

**Reminder**: Below are the running notes you earlier completed in this module for one or more sets of topics:

**Your notes on Set 1**:

[RunningNote1]

**Your notes on Set 2** (if you studied a second set of topics):

[RunningNote2]

**Your notes on Set 3** (if you studied a third set of topics):

[RunningNote3

**Your notes on Set 4** (if you studied a fourth set of topics):

[RunningNote4]

**Concluding observations**

Please write a concluding observation on the set(s) of topics you studied in this exercise, synthesizing and analyzing as needed in ways that bear on the research question you are trying to answer.

An “observation” should have two parts: a more-or-less objective description in which you have high confidence; and a thesis, hypothesis, interpretation, or suggestion for which you have less confidence before you study the model more closely but that might bear on your research question.

**Enter your concluding observations in the field below:**

[FinalNote]

End of Block: CONCLUDING OBSERVATIONS

Start of Block: RESEARCH TAKEAWAYS (v. 2, created 23 June 2019; rev. 27 June 2019)

|  |
| --- |
|  |

## TAKEAWAYS: Before you finish: Research Takeaways?

After studying your topic model in this exercise, are there any research takeaways you want to record, store, describe, or log (depending on the nature of the takeaway) to provide key evidence, materials, and data that can be used in writing up a report on your research question or for future humanities advocacy? For example, a research takeaway can be:

* Evidence in the form of an important topic(s) or relation between topic(s) and words, documents that you want to describe so that you can find it again;
* A screenshot of something you found while analyzing your topic model with one of the Topic Model Observatory visualization interfaces;
* Citations for or excerpts from key documents associated with a topic (e.g., three newspaper articles you want to use in a report);
* A text copy of words in a topic ranked by proportional weight;
* Etc.

If you have found such research takeaways, record, store, or describe them as appropriate in your team folder on the WE1S Google team drive. You may also want quickly to log or make a note about your takeaways in the field below in order to have a record of them as associated with this survey. That way, when you read over your survey results, you can remember what you took away and where you stored or annotated it. P.S. Some of the Topic Model Observatory interfaces may also allow you to use the [Hypothes.is](https://web.hypothes.is/) web annotation tool to highlight or annotated them in your browser for private or shared use.

## Companion questions?

"Companion questions" may be defined as questions that help provide context for the research question you are currently addressing or that are follow-on questions you might want to pursue later. WE1S asks you periodically to consider if companion questions have occurred to you because it's a good way to prevent the kind of "tunnel vision" that research can sometimes create.

This is best explained by way of example. Suppose that your research question concerns the degree and nature of discussion about the "humanities crisis." A tunnel-vision view of the problem would be to consider only discussion of the "humanities crisis." A contextually wider view would consider companion questions such as the following: what is the degree and nature of discussion about *other* kinds of crisis by comparison? For example, in what other contexts (political, economic, spiritual, etc.) does the word "crisis" tend to appear? Do mentions of "crisis" in all these contexts tend to spike up together, or be concentrated in certain kinds of media or sources? Etc.

Having advanced on your research question using the present module, have any "companion questions" occurred to you? I**f so, please set them down here so that they can be remembered for possible future use:**

End of Block: RESEARCH TAKEAWAYS (v. 2, created 23 June 2019; rev. 27 June 2019)

Start of Block: SURVEY END (v.2, created 20 June 2019, rev. 28 June 2019)

## CONFIDENCE? Final consideration -- Confidence assessment

It is possible after conducting an interpretation exercise to conclude that you have either high or low confidence in the results (with confidence descending all the way down to a null result).

Please rank your level of confidence in the results of this interpretation exercise on the following scale:

1 ----------------------------------------------------- to ------------------------------------------------- 10

 (1 = null result) (10 = full confidence)

Choose a confidence value.

## Suggestions about this module?

Do you have any suggestions about this module of the WE1S Interpretation Protocol (problems, improvements)?

## **Important**: Save this completed module in your team working folder.

Please save this completed module of the Interpretation Protocol in your team’s working folder:

**[FolderURL]**

This module is part of the documentary evidence trail for your future research reports and Key Finding cards.

END **This is the end of the current module of the WE1S Interpretation Protocol.**

End of Block: SURVEY END (v.2, created 20 June 2019, rev. 28 June 2019)