

(v. 2) Module 1 - Take an overview of a topic model

Survey Flow

ReferenceSurvey: (Ref. Survey) Info about this exercise

Block: WHAT THIS MODULE DOES (1 Question)

Standard: TAKE A FIRST LOOK AT MODEL (3 Questions)

Standard: OBSERVE SOME TOPICS (4 Questions)

Standard: Observe some clusters & trends (4 Questions)

Standard: Summarize your observations of the model so far (4 Questions)

ReferenceSurvey: Ref. Survey Research Takeaways

ReferenceSurvey: (Ref. Survey) Survey end

Page Break

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.

Trial surveys do not have to be completed. (However, you have to rush through them with empty answers all the way to the finish to clear them from the system. Otherwise, next time you open the same survey, Qualtrics will put you back where you were in the same survey.)

Actual surveys should be completed within one week, after which they will be automatically recorded as finished.

Check this field if this is just a trial: (1)

OPEN RESEARCH POLICY

Statement of open research policy

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"](#) statement). This aim is related to current practices of ["open science"](#) and ["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](#)).

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".

Team names and members can be looked up on the [Team Dashboard](#). (Note: at the end of this survey, you can also enter an email address if you wish to receive a copy of the survey with your answers.)

Your name or team name. (If team, then use the format "Team1", "Team2", etc. Please do not use a space): (1) _____

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): (2) _____

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 letters for each member's stage--e.g., "2AL" (for Alan Liu), "2LT" (for Lindsay Thomas), etc. . (3) _____

Today's date in the format YYYY-MM-DD (e.g., 2019-07-06) (4) _____

The **present** Interpretation Protocol module you are using (e.g., 1, 2, 3a, 4a, etc.) --This is needed to generate an exercise ID. For example, if this is Module 3a, then enter 3a (5) _____

G DRIVE INFO Your team folder in the project Google team working space.

The shared WE1S working space for AM teams in summer 2019 for storing notes, reports, materials, etc. related to interpretation work is in this [Google Drive folder](#). Go there to find the subfolder your team created for your work and enter its URL here.

Please enter below the URL for your team's subfolder in the common AM Teams workspace:

Page Break



EXERCISE ID Exercise ID

The ID for the present exercise (generated from the information you just provided) is as follows. (Do not change. If you see a mistake, please go "back" in this survey and correct the information that generated the Exercise ID.)

EXERCISE DESCRIP 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.")

TOPIC MODEL **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](#).

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/](#). (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" (1) _____

"Start page" URL (2) _____

RESEARCH QUESTION **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 (1) _____

Operationalized form of question (2)

Research question ID (use the format "team2-q1", representing in this example team 2's first research question): (4) _____

PREVIOUS MODULES Previous Interpretation Protocol modules you used to address your research question (if any)

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.

- Step 1 - Module used (1) _____
- Step 2 - Module used (2) _____
- Step 3 - Module used (3) _____
- Step 4 - Module used (4) _____
- Step 5 - Module used (5) _____
- Step 6 - Module used (6) _____
- Step 7 - Module used (7) _____
- Step 8 - Module used (8) _____
- Step 9 - Module used (9) _____
- Step 10 - Module used (10) _____

Page Break _____

R1: exercise info

Reminders - Info about this Exercise: **Your name or team name:** `{EXERCISE INFO/ChoiceTextEntryValue/1}` **`{G DRIVE INFO/ChoiceTextEntryValue}`**
target="_blank">Link to your team's folder on the WE1S project Google team drive.

Exercise ID: `{EXERCISE ID/ChoiceTextEntryValue}` **Exercise**
Description: `{EXERCISE DESCRIP/ChoiceTextEntryValue}` **Topic model name:**
`{TOPIC MODEL/ChoiceTextEntryValue/1}` `{TOPIC MODEL/ChoiceTextEntryValue/2}`"
target="_blank">Start page of model **Research question (and**
ID): `{RESEARCH QUESTION/ChoiceTextEntryValue/1}` (research question ID #`{RESEARCH QUESTION/ChoiceTextEntryValue/4}`) **Operationalized form of question:** `{RESEARCH QUESTION/ChoiceTextEntryValue/2}` **Previous sequence of interpretation modules used for this question (if any):** `{PREVIOUS MODULES/ChoiceTextEntryValue/1}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/2}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/3}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/4}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/5}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/6}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/7}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/8}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/9}`, `{PREVIOUS MODULES/ChoiceTextEntryValue/10}`

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

Start of Block: WHAT THIS MODULE DOES

MODULE PURPOSE **What module 1 does**

(Module 1, v. 2, created 18 June 2019; last rev. 28 June 2019.) The purpose of this module 1 of the WE1S topic model Interpretation Protocol is to take you through the first stage of analyzing a topic model in a standardized and reproducible way. Module 1 asks you to assess the quality of a model as a whole and reconnoiter its topics, clusters, and other features.

Tools: In this module you will be working primarily with three general-purpose visualization interfaces for exploring topic models: Dfr-Browser, TopicBubbles, and pyLDAvis. You will also be using a few more specialized interfaces. and WE1S Topic Bubbles. (For explanations of and best practices and in using these tools, consult the WE1S [Topic Model Observatory Guide \(TMO Guide\)](#)).

End of Block: WHAT THIS MODULE DOES

Start of Block: TAKE A FIRST LOOK AT MODEL

R1 exercise info `{R1: exercise info/QuestionText}`

STEP 1 Step 1. Assess basic quality of model at this granularity

Using the general-purpose visualization interfaces Dfr-browser, Topic Bubbles, and pyLDAvis in the WE1S Topic Model Observatory (see [TMO Guide](#) for best practices in taking an overview in the general-purpose interfaces), quickly reconnoiter the model you are studying. Then answer the following quality-assessment question.

	1 (worst) (1)	2 (2)	3 (3)	4 (4)	5 (best) (5)
How satisfactory (coherent, useful) do you think this topic model is at the topic-number granularity you are studying? (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



OTHER GRANUALRITY?

If your answer to the above question falls lower than a satisfaction score of “3,” examine different topic-number granularities of the model (e.g., 150 instead of 200 topics) and enter below the granularity of the model that seems best to you (identified by number of topics, e.g., “150”). You may wish to consider changing to the other granularity at this point. (If you are working in an interpretation team, this would be the time to have a team discussion about which granularity to use.

Best granularity of model if not the one you are currently studying:

End of Block: TAKE A FIRST LOOK AT MODEL

Start of Block: OBSERVE SOME TOPICS

STEP 2 Step 2. Initial Inspection of Model -- Observe some topics

Now you will begin reconnoitering the model--a little like going to a real-estate open house and

first of all walking around all the rooms. Essentially, you are developing an overall sense of the model. As you inhabit a model over a sequence of Interpretation Protocol modules, you will gain increasing familiarity with the model.

R1 exercise info [\\${R1: exercise info/QuestionText}](#)

LOOK AT TOPICS Using the general-purpose visualization interfaces (Dfr-browser, TopicBubbles, and pyLDAvis), look at some topics in the model. Here is a checklist of topics to sample: Topics at a variety of proportional weights (size of their circles in these models or their ranked "proportion of corpus" in the Dfr-browser "Overview > List" view) Topics that appear densely clustered together in the "scaled" or "intertopic distance map" views of the general-purpose interfaces. *(However, you will later use more specialized cluster analysis tools in this exercise to detect clusters more accurately.)* Topics that seem to be outliers in the "scaled" or "intertopic distance map" views of the general-purpose interfaces Topics in which the word "humanities" (or some other word you are interested in) is among the top words.



RUNNING NOTE 1 If you noticed anything of interest in observing topics that you might want to follow up on in pursuing research questions in more detailed Interpretation Protocol modules later, use the "Running Note" field below to record your thoughts.

RUNNING NOTE: (the content of this note will be carried forward to future sections of this survey to help with remembering what you have done and maintaining overall cognitive grasp of your current research task).

End of Block: OBSERVE SOME TOPICS

Start of Block: Observe some clusters & trends

STEP 3 Step 3. Initial Inspection of Model -- Observe some clusters and trends

Next you will take a closer look at possible topic clusters in the model as well as see if you can recognize any temporal trends in topics.

R1 exercise info [\\${R1: exercise info/QuestionText}](#)

CLUSTERS & TOPICS

Clusters: Using the specialized "clustering analysis" tools named Clusters7D and DendrogramViewer in the WE1S Topic Model Observatory (see [TMO Guide](#) for how-to's and best practices), look at more closely at what might be clusters of topics (and also outlier topics) in the model.

Trends: Using the "stacked" view in Dfr-browser (see [TMO Guide, Chapter 1](#), section 1.c), see if you can spot any apparent trends or other temporal changes when a topic suddenly thickens or thins in proportional weight in the model. If you find any topic that is interesting in its temporal behavior, look it up by in any of the general-purpose visualization interfaces (including Dfr-browser's other "list" and topic views) to see what it is about.



RUNNING NOTE 2 If you noticed anything of interest in reconnoitering clusters and trends that you might want to follow up on in pursuing research questions in more detailed Interpretation Protocol modules later, use the "Running Note" field below to record your thoughts.

RUNNING NOTE:

End of Block: Observe some clusters & trends

Start of Block: Summarize your observations of the model so far

STEP 4 Step 4 (final step). Summarize your observations of the model

R1 exercise info [\\${R1: exercise info/QuestionText}](#)



RUNNING NOTE 3 REMINDER -- YOUR RUNNING NOTE:

Here is what you have observed so far in taking an overview of your topic model:



SUMMARY OBSERVATION Please look over your "running note" above and correct or edit as you wish. Then below write a brief summary or synthesis of highlights, reflections, or speculations for further research about this model based on your initial overview of it:

End of Block: Summarize your observations of the model so far



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. **Reminder:**

The folder for your team ([\\${EXERCISE INFO/ChoiceTextEntryValue/1}](#)) in the WE1S Google team drive is [\\${G DRIVE INFO/ChoiceTextEntryValue}](#)" target="_blank">here. The main "AM Team Working Space" for all the teams in the WE1S Google team drive is [here](#). 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 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](#) web annotation tool to highlight or annotated them in your browser for private or shared use.



COMPANION Q's? **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? **If 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:

	Null Result (no confidence) (1)	(4)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	Valid Result (full confidence) (13)
Confidence level: (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

SUGGESTIONS? Suggestions about this module?

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

EMAIL A COPY? Before you conclude this module of the WE1S Interpretation Protocol, do you wish to have a report of your answers emailed to you?

Choose "yes" if you have completed the module successfully as part of an interpretation project. The plain-text, HTML, and PDF copies of your survey should be kept as part of the documentation for answering a research question.

- Yes (4)
- No (5)

Display This Question:

If Before you conclude this module of the WE1S Interpretation Protocol, do you wish to have a report... = Yes



EMAIL ADDRESS Please provide your email address to receive a copy of this survey.

You will receive by email a text transcript of the questions and answers in this survey, plus links to HTML and PDF versions. You should keep these as documentation of your interpretation project. **Caution:** The online HTML and PDF versions linked in the email will expire after 31 days.)

Good practice is to copy the emailed transcript as text into a Google Doc in your team folder (in a subfolder called "Qualtrics survey transcripts") and also to download and store there the PDF

version. Name the transcripts in the following format (for example): "Module3a (2019-07-11) - history (Tarika)" (In this example, "history" has been added because it was the keyword that Module3a was used to explore; and "(Tarika)" was added to identify the specific person completing the survey if the whole research team did not do so.)

Caution: The online HTML and PDF versions linked in the email you receive will expire after 31 days.

Your email:

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

IMPORTANT: To submit your answers and have Qualtrics mark your survey as completed, please press "Save à Next"

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