(v. 2) Module 3.a - Analyze a topic

Survey Flow

ReferenceSurvey: (Ref. Survey) Info about this exercise

Block: WHAT THIS MODULE DOES (1 Question)
Standard: IDENTIFY TOPIC TO EXAMINE (3 Questions)
Standard: EXAMINE TOPIC (7 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

<u>Statement of open research policy</u>
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 <u>"About"</u> statement). This aim is related to current practices of <u>"open science"</u> and <u>"open-notebook science."</u> 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 <u>WE1S bibliography on machine-learning interpretability</u>).

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:

_____

⊙

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"I (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 3.a does
(Module 3.a, v. 2, created 20 June 2019; last rev. 23 June 2019.)

 The purpose of this module 3.a of the WE1S topic model Interpretation Protocol is to examine a
specific topic.

End of Block: WHAT THIS MODULE DOES

Start of Block: IDENTIFY TOPIC TO EXAMINE

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

✳

STEP 1 **Step 1. Identify a topic to examine.**

Modules 1 (Overview) and 2 (Representative topics) in the WE1S Interpretation Protocol will previously have given you an overall sense of topics in the model you are looking at. Based on that initial understanding of the model, and using a combination of the tools in the WE1S Topic Model Observatory (see TMO Guide), look over the topics in the model and identify a particular topic that you want to examine closely to address your research question.

Enter the number of that topic below:

_____

TOPIC WEIGHT LEVEL Using the general-purpose tools in the Topic Model Observatory (Dfr-Browser, TopicBubbles, pyLDAvis), examine the proportional weight of the topic relative to other topics in the model. Would you rank this as a high-weight, middle-weight, or low-weight topic?

(Remember that you learned about "terraces" of high/middle/low weight topics in Module 2.)

The topic I identified is:

○ high weight (1)

○ middle weight (2)

○ low weight (3)

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

R data you entered
 Reminders - Data you entered above:
 The topic you are examining is **${STEP 1/ChoiceTextEntryValue},** which you assessed as **${TOPIC WEIGHT LEVEL/ChoiceGroup/SelectedChoices}**.

STEP 2 **Step 2. Examine your topic and take some notes**

(a) Using the general-purpose interface in the Topic Model Observatory (Dfr-Browser, Topic Bubbles, and pyLDAvis), examine the top words in your topic. (Remember that in pyLDAvis you can adjust the λ "lambda" metric for relevance of terms to a topic to fine-tune your sense of the relation of top words to a topic. The creators of this tool found in one test with a topic model that λ = 0.6 gave optimal results. See Sievert and Shirley.)

Screenshot: pyLDAvis - adjust λ (lambda) relevance metric

 Evaluate the coherence of the top words as a group and consider what you might label this topic based on the words. (A topic label should be a short descriptive phrase consisting usually of no more than two concepts related to each other--e.g., "college majors and careers." A label should not be in the form of a thesis argument--e.g., not something like "college majors make little difference in careers")

 (b) Using the special-purpose interfaces of Metadata-pyLDAvis and Metadata7D in the Topic Model Observatory (see *TMO Guide*, chapters 4 and 5), examine the relation of your topic to publication sources (to see, for example, if the topic is heavily associated with some particular source or set/kind/region of sources.

(c) Using the special-purpose interfaces of Clusters7D and DendrogramViewer in the Topic Model Observatory (see *TMO Guide*, chapters 7 and 8), see if your topic is part of a cluster of topics or an outlier. If it is part of what seems to be a cluster of topics, take a quick look at some of the neighboring topics.

(d) Using the Dfr-Browser topic view, take a look at the top documents associated with this topic. (You can also use Metadata-pyLDAvis to see the top documents assocated with your topic.)

⊙

RUNNING NOTE 1 **Your Notes**: Take some running notes below in regard to whatever you find that is of interest in regard to instructions #a, b, c, d above. Later in this module you can draw on these notes to make a summary observation.

_____

_____

_____

_____

_____

LABEL TOPIC **Label your topic**: Try to label your topic. (A label should be a word or short phrase representing only the objective part of your observation, or what the topic is "about". It should not be formed as a sentence, which tends to posit a thesis.)

_____

Page Break ————————————————————————————

R data you entered
 Reminders - Data you entered above:
 The topic you are examining is **${STEP 1/ChoiceTextEntryValue},** which you assessed
as **${TOPIC WEIGHT LEVEL/ChoiceGroup/SelectedChoices}**.

 You labeled this topic: "${LABEL TOPIC/ChoiceTextEntryValue}"

◎

OBSERVE TOPIC **Please write a summary observation on this topic**: Drawing on, revising,
and extending your running notes, write a summary observation on your topic designed to
address your research question.

For the purpose of addressing a research question, a good "observation" should have two parts:
a more-or-less objective description in which you have high confidence; and a thesis,
hypothesis, interpretation, or suggestion bearing on your research question.

_____

_____

_____

_____

_____

End of Block: EXAMINE TOPIC

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. **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 iin 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? I**f so, please set them down here so that they can be remembered for possible future use:**

_____

_____

_____

_____

_____

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)	○	◐	◐	◐	◐	◐	○	○	○	○

SUGGESTIONS? **Suggestions about this module?**

Do yo 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 ithe whole research team did not do so.)

  Caution: The online HTML and PDF versions ilinked 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"