

Database Dishes

Nathalie Cooke, McGill University; Nora Shaalan, McGill University

NUTRITION INFORMATION

Aimed at students in their junior year of a liberal arts degree, this lesson acquaints students with various digital libraries. Students receive a menu in riddles, in the style of popular riddle menus from the 19th and 20th century that were distributed during events called “conundrum suppers.” Students are asked to taste test one of four database dishes—but before they can begin, they must solve the riddles describing the dish to discover which database they have been served! Once students identify their assigned database, they will test it on the topic of “conundrum supper” and present their results. During the second half of class, the instructor will show exploratory tools that allow students to analyze the primary source material they gather. This lesson allows students to acquire knowledge of the relative strengths of each database for constructing primary source corpora and augment their skill sets for analyzing primary source materials.

LEARNING OUTCOMES

By the end of the session, students will be acquainted with various primary source databases and able to conduct

efficient and effective searches for primary source materials. In addition, students should be able to assess the appropriateness of each database for their research needs. Through this exercise, students will become more comfortable with understanding and analyzing the contents of primary source corpora.

RELEVANT RBMS/SAA JOINT GUIDELINES

2B, 3A, 5A

COOKING TIME

Cooking time is one session of 90 minutes. A taste test may take students approximately 75 minutes (to be completed independently).

NUMBER SERVED

A minimum of 8 students (2 students per database), with more students assigned to each database in order to accommodate a larger group.

INGREDIENTS

- Copies of the menu and worksheet for each student (see figures 1 and 2).
- Access to the four databases is required (substitutions may be

Figure 1. Database Dishes, a menu handed out to students. Each group of students is assigned a “dish” (or database) to taste test. A solution set is provided below.

Database Dishes: Identifying Which Databases Generate the Best Content Sets

1st Dish

My first is the empire on which the sun never sets;
My second is black and white, and read all over;
My last is a repository.

2nd Dish

To find my first you must look to the place from
which you borrow your books.
On this final part of this plate is where men
and women legislate.

3rd Dish

My first may be a beanie, a fedora or a beret;
My second, the beginning of the end;
My last, if betrayed is broken;
My whole, a two-word database.

Dessert

I’m not a light breeze, but I am a squall;
I am a cyclone, a very strong gust;
I measure high on the Beaufort scale.

Database Dishes: Solution Set

1st Dish

British Newspaper Archives

2nd Dish

Library of Congress

3rd Dish

Hathi Trust

Dessert

Gale

Figure 2. Worksheet.

Each student is given a handout, which guides their experience with the database. The first section contains a brief overview of Boolean operators and other techniques for effective searches. The second section of the worksheet asks them to keep a record of the salient features of the database with which they are working. It will include the following questions, formatted in such a way that there is space for them to write their responses on the sheet itself.

Boolean Operators

- Using **AND** between keywords will give you results that contain both terms. It is a way to narrow your search.
- Using **OR** between keywords will give you results that contain either of your keywords. It is a way to broaden your search.
- Using **NOT** between two keywords will give you results that contain the first term, as long as it does not contain the second term. It is a way to narrow your search.
- Using **quotation marks** around more than two terms will give you results that use that exact phrase, in that order.
- Using an **asterisk (*)** at the end of the keyword will give you your keyword with various endings. Example: *delight** will give you results for *delight*, *delighted*, *delights* and *delightful*.
- Using a **question mark (?)** within your keyword allows you to allow for differences in spelling. Example: “*organi?ed*” will give you results that contain both *organized* and *organised*.

Database: Questions to guide your assessment of the database

1. Does your database keep a record of your searches? If not, create a crumb trail of the searches you have performed below. Can you select documents to be saved in a content set?
2. Does your database allow you to filter results? What filter options are you given?
3. Is your database nation-specific? Are the primary sources you find predominantly from a single country? In a single language?
4. What kinds of documents does your database use? Periodicals or newspapers? Books and monographs? Does it include unpublished manuscripts and handwritten materials?
5. Does your database provide transcribed text for the primary sources? For example, can you see a typed version of the text that was originally handwritten? (This process is usually described as OCR or “optical character recognition.”) If so, does the database measure the accuracy of the OCR?
6. What file download options are you given? Can you download PDF versions of your primary sources?
7. What metadata is available for the primary materials? In other words, what descriptive information is available for each item you find on the database?

made depending on the availability of ingredients):

- British Newspaper Archives
- Gale Primary Sources
- Hathi Trust
- Library of Congress Chronicling America
- Open source data visualization resources such as Voyant tools (see <https://voyant-tools.org/>).

PREPARATION**Introduction (25 minutes)**

- Hand out Database Dishes menu and worksheets and introduce the exercise to students. Separate the students into groups of equal size, with each group assigned to one “dish” or database (see figure 1).
- Each group of students will be responsible for researching a single topic, “co-nundrum supper” in this case, through one database. Depending on the size of the group, students can break down the task by keyword, date range, or other parameters.
- Allow students some time to solve the riddle assigned to their group. If students are stuck, reveal the names of the databases to facilitate the task.
- Briefly review Boolean operators and other effective search techniques with students, using the worksheet as an aid (see figure 2).

Search (25 minutes)

- Ask students to keep a record of the various features of the database and the limitations they find, as well as the quantity and quality of source material they are able to mine. *Note:* The last 5 minutes could be taken as a short break.

Discussion (25 minutes)

- After students have had sufficient time to perform searches and to aggregate results, come back together as a group to discuss the efficiency of each database.
- Ask each group to present the information they gathered on their worksheet. The instructor can take this opportunity to provide some explanatory information on the databases they have selected and discuss ways of analyzing large swathes of primary source material.

Experimenting with Digital Tools (15 minutes)

- In the remaining time, the instructor can introduce some open access and user-friendly digital tools, such as those provided by the Gale Digital Scholar Lab or, if instructors do not have access to these databases and their tools, open source resources such as Voyant tools. These text-analysis tools produce visualizations, such as graphs, n-grams, and word clouds, that illustrate characteristics of the corpus (i.e., most common words or word frequencies) and that will give students insight into the latent patterns

within the corpora they build. This will provide an alternative angle from which students can analyze their corpora.

TASTE TEST

Using the lesson plan as a guide, ask students to repeat the same steps independently with a topic of their own choosing. For this assignment, students are required prepare a preliminary data visualization with a short explanation and a brief paragraph that analyzes the primary source material they have uncovered. Students will be assessed using the following rubric:

- Selecting a database, conducting an effective search, and completing or printing a crumb trail of their search strategy (/5).
- Synthesis paragraph of 150 words on the content set or corpus created, with justification for database selection and search terms used (/5).
- Creating a visualization using Voyant and providing a paragraph of 150 words that explains what insight such a visualization gives (/5).