Experimenting with Boolean search

INTRODUCTION: Understanding structured searches

Imagine you are about to write a dissertation on the following topic:

Interactive data visualization of unstructured data

Text analytics can provide useful insights however it is often difficult for end users to interpret them. Visualization techniques can help present patterns in a way that supports interpretation and interactivity allowing effective exploration and knowledge discovery. This project will deliver a novel interactive data visualization tool for unstructured data derived from social media conversations, built on open-source toolkits such as D3.

How would you find relevant literature? A quick Google search may get you started, but is unlikely to be rigorous or systematic enough for a scientific dissertation. For a structured search task like this, we need to break it down into concepts or search facets, e.g.

1. Input (social media)
2. Transformation (text analytics)
3. Output (interactive data visualisation)

There are of course many other ways to think about and analyse an information need, but the above analysis is a reasonable start. You can use these facets to create a structured search strategy, such as:

```
"interactive data visualisation" AND "social media" AND "text analytics"
```

However, this search is unlikely to be very comprehensive. Let's expand each facet:

- Set1 = {"interactive data visualisation" OR "interactive data visualization" OR "information visualisation" OR "information visualization" OR "data visualisation" OR "data visualization"}
- Set2 = {"social media" OR "social network" OR facebook OR twitter}
- Set3 = {"text analytics" OR NLP}

Then combine them using Boolean logic:

- Set4 = Set1 AND Set2 AND Set3

With this method you can expend each individual facet with different synonyms and then merge them to create the final result when you think each individual facet captures that aspect of the information need. Our search query now looks like this:

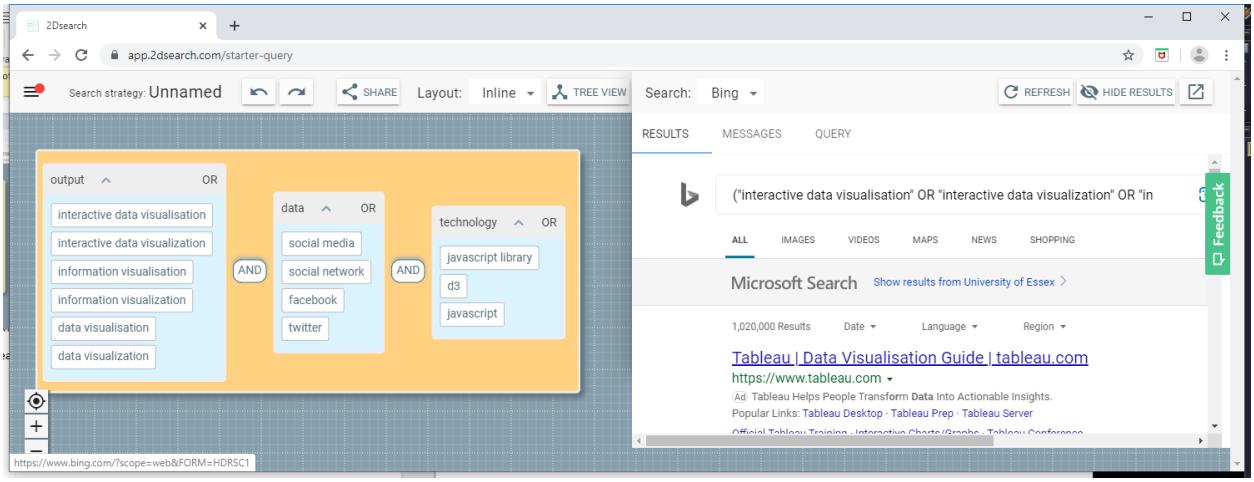
```
("interactive data visualisation" OR "interactive data visualization" OR "information visualisation" OR "information visualization" OR "data
```

```
visualisation" OR "data visualization") AND ("social media" OR "social
network" OR facebook OR twitter) AND ("text analytics" OR NLP)
```

However, when search strings get this long, errors can start to creep in and it becomes increasingly difficult to manage the query without breaking it.

EXERCISE: Explore visual approaches

It is possible to create structured searches using visual tools. Using 2Dsearch (app.2dsearch.com) it is possible to construct the search above by entering terms on a 2D canvas and then dragging and dropping them to form groups:



The query is then executed using the selected search engine and you can view the results in the righthand panel (select Bing for instant results or Google Scholar to see the results in a popout window).

1. Start a new search

Go to https://app.2dsearch.com/, click "Get Started" and follow the onboarding Tour. You will see there is already a search query entered as an example. Watch the introductory videos:

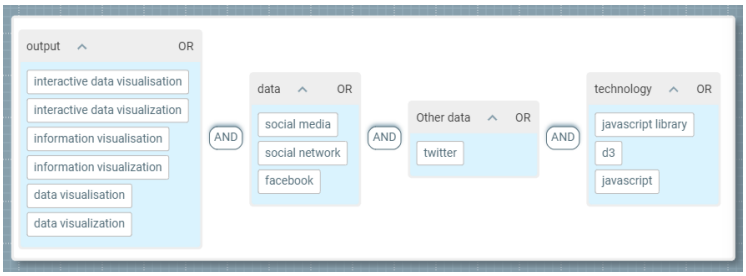
- https://youtu.be/tJ3fBRvszhQ
- https://youtu.be/-jI0AzgywgE

To start a new search strategy, open the menu in the top left and select the 'New' option in the menu for a blank canvas. Now start typing your terms, and drag and drop to create groups.

Name the groups so you know what each means conceptually. Select an "inline" layout. Rearrange your groups so that you have something similar to the screenshot above.

2. Create a new group

Right click on the canvas and select "Create group" and name it "Other data". Drag the keyword "Twitter" into this group. Now drag the new group box into the main query box so it joined with an AND operator. You query should look like this:



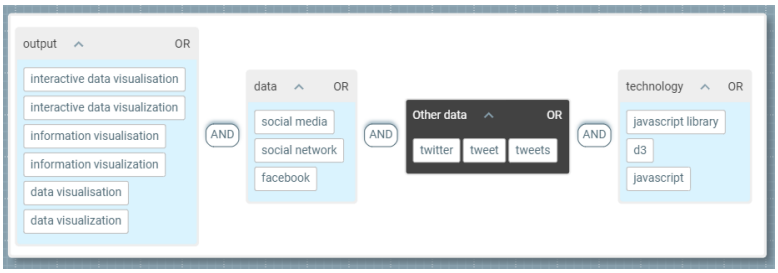
3. Apply negation

We are primarily interested in results from Facebook and not Twitter. Adjust the query to use AND-NOT for the term Twitter. To do this, you can right click on the group and select “Apply NOT”. Your query should now look like this:

Use the ‘Search’ drop down to select Bing as your target search engine and note the number of results.

4. Expand the terms

Right click on the term “Twitter” and select “Suggest terms…” and this will offer related terms based on lookup in various online resources and language models. Add the terms “tweet” and “tweets”. Your query should now look like this:

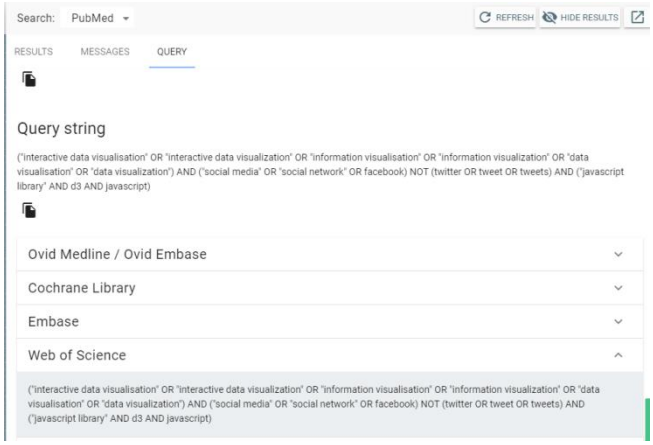


Your search results should now exclude any results that have the keywords “Twitter”, “tweet” and “tweets” in them. You should have fewer results than before (although web search engines tend to interpret structured searches in an inconsistent manner).

5. Copy the query string

Select the “Query” tab in the righthand panel and you will see your query string. If you switch your database to Pubmed, you will also see translations to various other databases. Your query should look something like this:

```
("interactive data visualisation" OR "interactive data visualization" OR "information visualisation" OR "information visualization" OR "data visualisation" OR "data visualization") AND ("social media" OR "social network" OR facebook) NOT (twitter OR tweet OR tweets) AND ("javascript library" OR d3 OR javascript)
```



Select the translation for Web of Science. This can now be copied and pasted into other databases. Note that the operator AND-NOT is abbreviated to NOT in this syntax.

6. Save and share your search strategy

Save your search strategy using Menu->Save. If necessary, create an account using your Google ID or use a custom username and password. Generate a link to your search (using the 'Share' button) which you will share in the next exercise.