

# BacNet v.1.0

## User manual

BacNet software tool builds a network from given:

- XML-file that contains abstracts downloaded from PubMed database
- File (files) that contain list of words to search for

Usage:

### *Step 1: Download abstracts from PubMed*

Go to <http://www.ncbi.nlm.nih.gov/pubmed> and run your query to that database. The query might look like this: "(vagina OR vaginal ) AND (lactobacillus)". Then push the "Search" button. Results should look similar to those in Fig. 1 below:

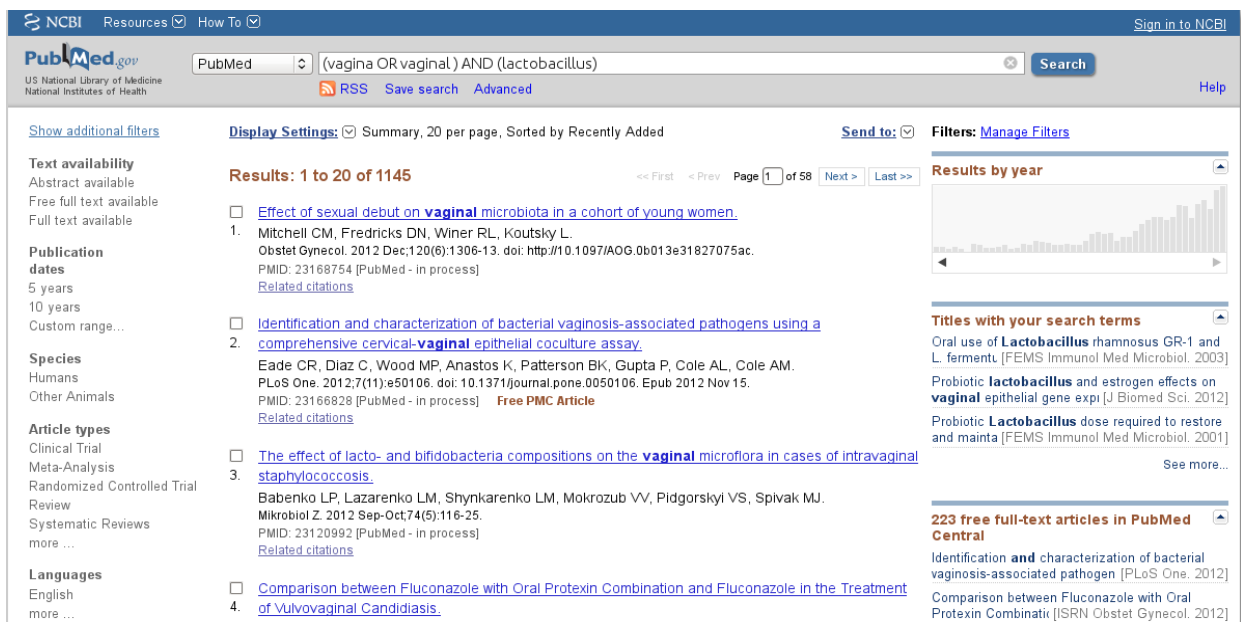


Fig. 1. Pubmed results from "(vagina OR vaginal ) AND (lactobacillus)" query.

Then download results. Save them into BacNet's root directory. The name of the file must be: "pubmed\_result.xml":

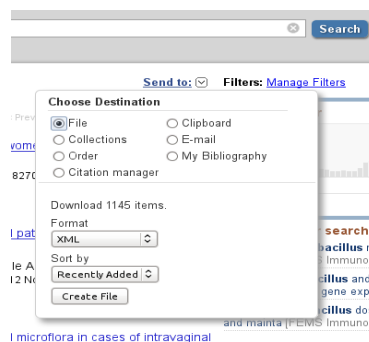


Fig. 2. Downloading results.

This **pubmed\_result.xml** file contains all abstracts from the output.

### *Step 2: Prepare input files*

Go to BacNet's root directory. Prepare a list of words to look for. Save it into separate file (in BacNet root directory), name it wherever you want. The file must contain lines, each single line represents a word (or a phrase).

Example:

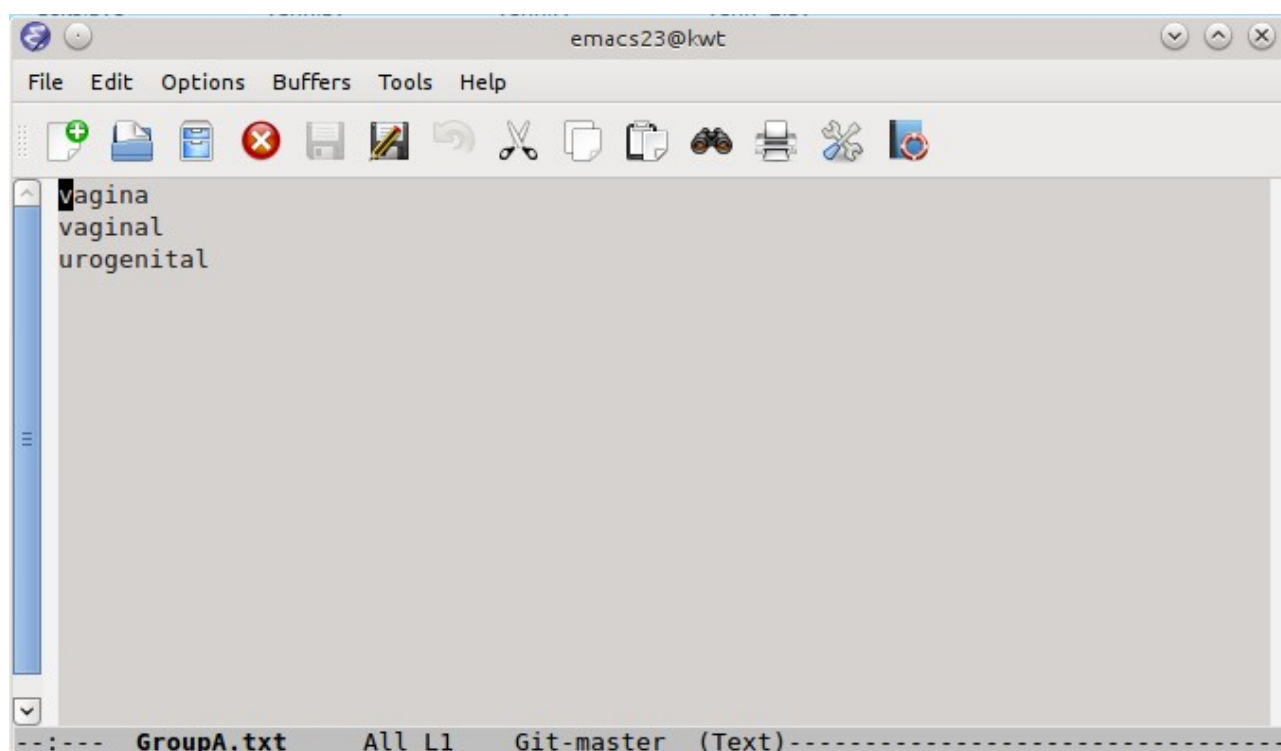


Fig. 3. Example of the set of words in the input file.

### *Step 4: Run the BacNet's search engine*

BacNet software tool generates adjacency matrices and saves them into .csv format. Adjacency matrix is needed to build a network. Run BacNet by executing the following command:

**python bacnet.py <required options>**

<required options> :

**-1 <filename>** - Looks for co-occurrence of any of the words in a *set\_of\_words.txt* with other words in the same set excluding themselves. The output of this is the adjacency matrix named "adj\_matrix\_11.csv"

Example: `python bacnet.py -1 GroupB.txt`

**-1 <filename> -2 <filename>** - Co-occurrence of any of the words in *set\_of\_words1.txt* with any of the words in *set\_of\_words2.txt*. The output of this is the adjacency matrix named "*adj\_matrix\_12.csv*"

Example: `python bacnet.py -1 GroupA.txt -2 GroupB.txt`

In order to do this you have to prepare two separate files. Look at the figures below:

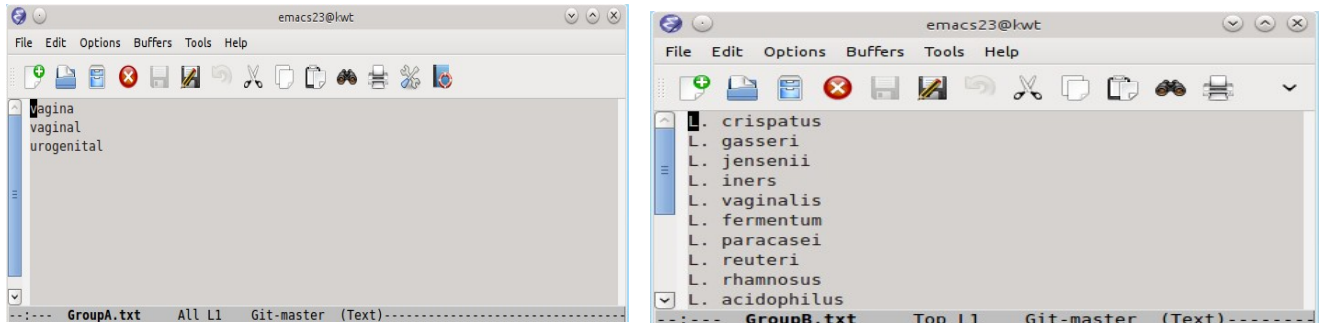


Figure 4. Example of input files.

**--kw <filename>** - Keyword approach. If this option is set, you have to provide a list of words, assuming that "*pubmed\_result.xml*" is already presented. The output of that is adjacency matrix named "*adj\_matrix\_strong.csv*". For more detailed explanations of this method, please see the article.

Example: `python bacnet.py --kw set_of_words.txt`

**--tm <filename>** - Timing approach. Initial goal was to see how the network structure changes with time. Output of this command is a large set of PDF files saved in the BacNet root directory. Each of those PDF files represent a network generated for single year.

Example: `python bacnet.py --tm GroupA.txt`

### Step 5: Visualization

Open the R-script "*network\_visualizator.R*" (saved in the BacNet's root directory) and edit it if necessary. Run this script to see the network structure.

Example:

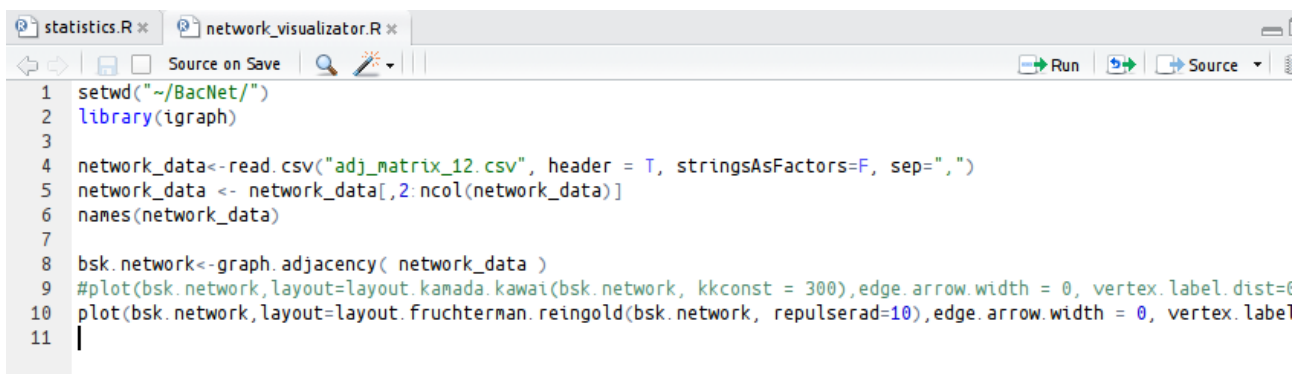


Fig. 5. R-script.

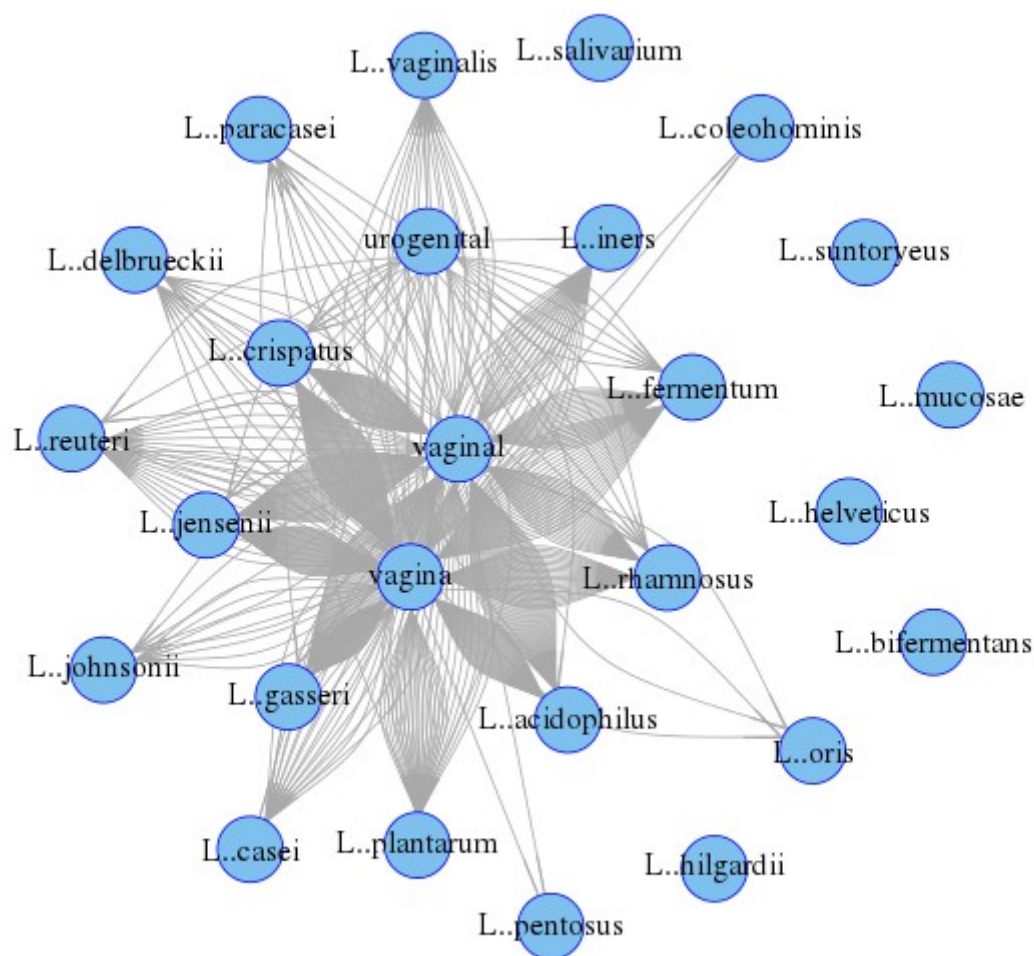


Fig. 6. Example of network.

## Contacts

Ilya Zhbannikov, [zhba3458@vandals.uidaho.edu](mailto:zhba3458@vandals.uidaho.edu). Any feed backs are greatly appreciated!