Instructions for using KNIME to find Diverticulosis cases and controls for eMERGE:

1. Download both Diverticulos-workflows-shared.zip (KNIME workflow) & diverticulosis-libs.zip (Java NLP libraries) files from phekb.org
2. Install & Configure KNIME
   a. Download & Install KNIME from http://knime.org/knime
   b. Open KNIME & Install extensions:
      i. From File menu select Install KNIME extensions
      ii. Under "KNIME Labs Extensions," check:
         1. **Required**: "KNIME Python Scripting" (see screen shot below)
         2. Optional: If you are going to read in your data that is stored as a SAS data set (SAS7BDAT fmt.), then also check "KNIME SAS7BDAT Reader (Windows Only)"
      iii. Optional: If you are going to read in your data that is stored in an Excel file (.xls file), then under "KNIME & Extensions" also check "KNIME XLS Support"
      iv. Click Next
      v. Click radio button to Accept terms of license agreement, click Next
      vi. Click Finish
      vii. You will be prompted to close & re-open KNIME
3. Import workflow into KNIME & connect to Java NLP libraries:
   a. Unzip the library files to a location you can access from the computer where KNIME is installed
   b. In KNIME, choose Import Workflow from File menu, Browse to the Diverticulosis-workflows.zip file you downloaded, click Next & Finish
   c. Open the 1ˢᵗ workflow: "1 -diverticulosis cases, controls, and ages" under Diverticulosis-workflows-shared
   d. Open (double-click) the "NLP" metanode & follow the instructions on that screen*
4. Execute the workflow (from Node menu select Execute All) with the included test data & check the output: there should be 5 diverticulosis cases, of which 2 also have diverticulitis, 2 controls, and 1 pt. that is neither a case nor a control.

5. Add your data. The Table Creator nodes are configured with the desired column names, types, and values, so please use these column names, types, and text type values when collecting your data. There are 4 ways to add your data (listed in order of most to least direct): configure & then connect 1 of these nodes to the Cache node (for the proc. data), and another to the Ages metanode (for the birth dates) & re-execute the workflow:
    a. Use a Database Reader node that returns the data directly from your database.   If your patient IDs are strings, connect from the DB node to & configure the String to Number node, which then connects to the Cache node
    b. Use a SAS (SAS7BDAT file format) Reader node to read directly from your SAS data set.
    c. Use a File Reader pointing to a file that contains data exported from your database.
    d. Use a XLS Reader pointing to an .xls (not .xlsx, save as older Excel version) file that contains the same data exported from your database.
    e. The Table Creator node can be opened & data can be pasted directly into it, but this method is <u>recommended only for testing a subset of data</u> (as copy/paste is error prone).

*Steps which are listed in KNIME to configure the Java Snippet NLP node:

1. On the Java Snippet (simple) - NLP node, right click & click Configure
2. Select Additional Libraries tab
3. Click Add button
4. Point to the location of the library files extracted earlier.
5. Select all and add all the files
6. Click OK

Screen shot of KNIME extension(s) to install, only the required extension is shown as checked: if you are going to import from a SAS or Excel file, be sure to check those extensions as well.