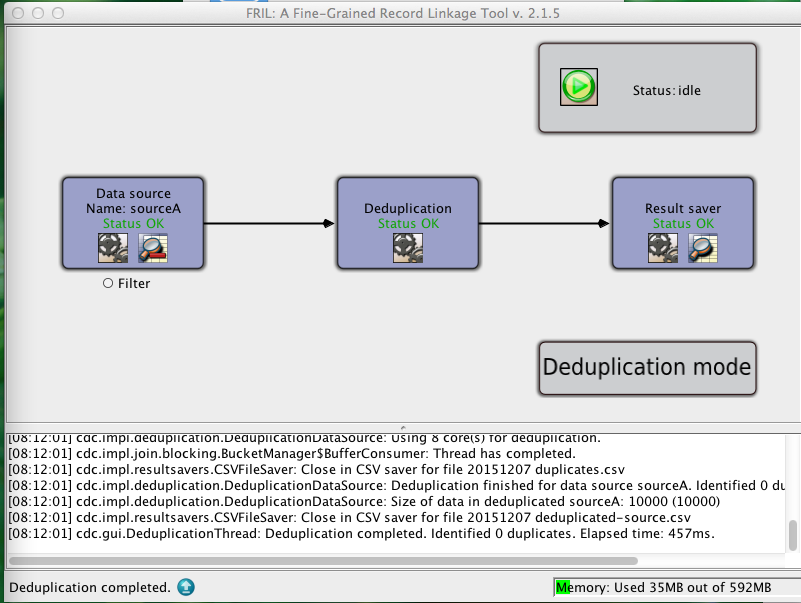
FRIL screenshots for deduplication of the dataset RLdata10000

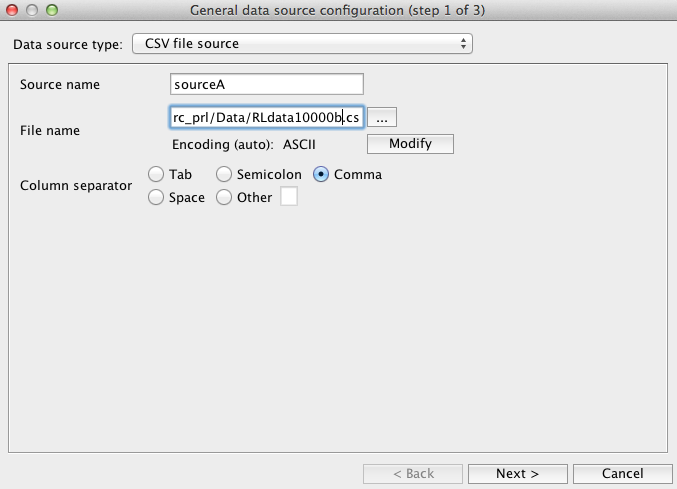
References:

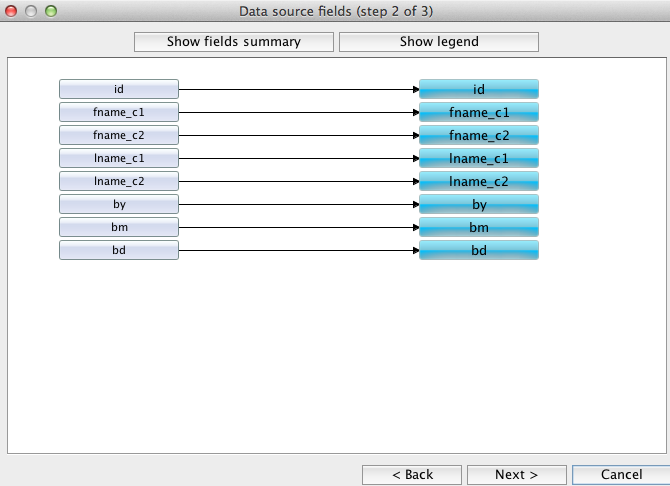
* FRIL v 3.2 Tutorial, <http://fril.sourceforge.net/FRIL-Tutorial-3.2.pdf>
* Choi, Sou-Cheng T., Yongheng Lin, and Edward Mulrow. "Comparison of Public-Domain Software and Services for Probabilistic Record Linkage and Address Standardization,” Towards Integrative Machine Learning and Knowledge Extraction*,* SpringerLNAI 10344, 2017. *To appear.* PDF available at http://tinyurl.com/ydxbjww4

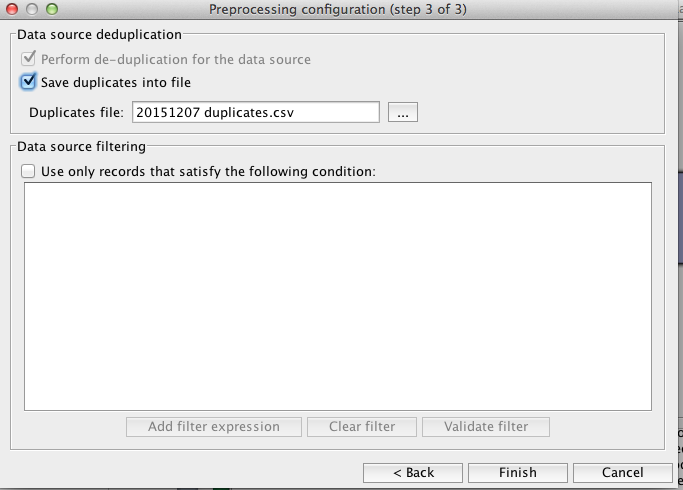
The main FRIL screen (on MAC OS):



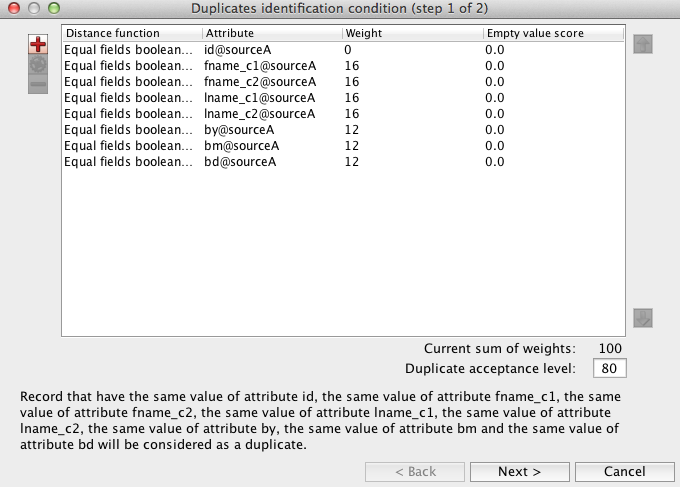
Click the icon in the box “Data source Name: sourceA”

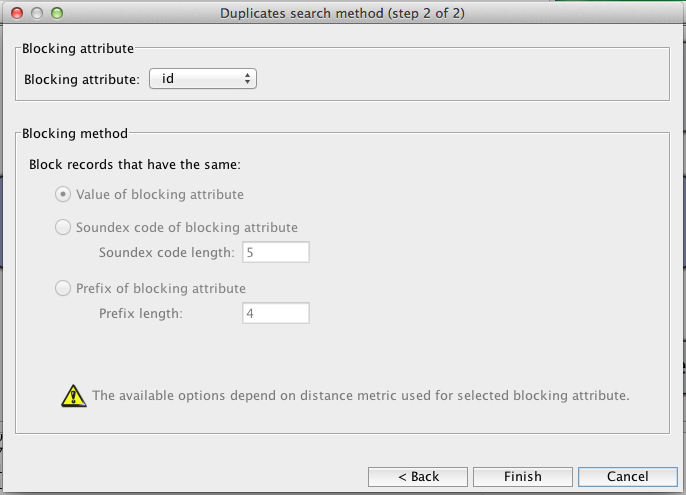




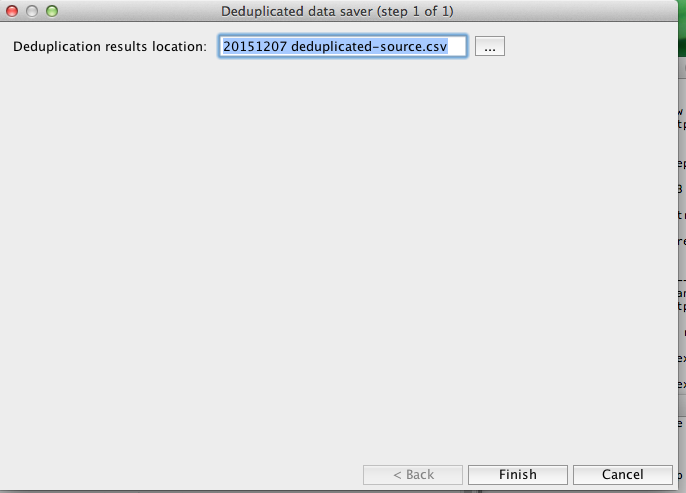


Click **Deduplication**:



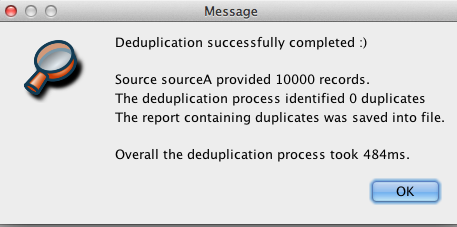


Click **Results saver** icon:

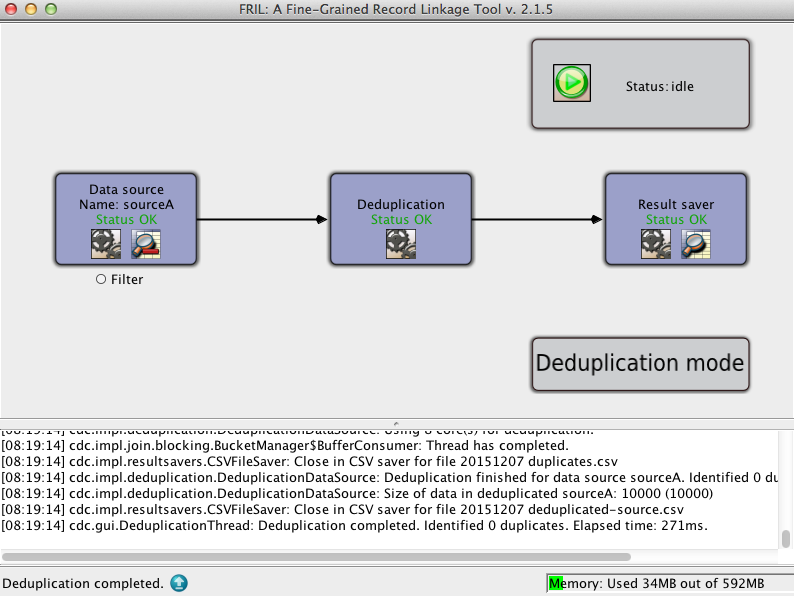


Click “**Status:idle**” button:

Finally, the following message box summarizes the results and time taken:



The main screen has a log that reports the memory usage:



I obtained a csv file from the process, in a text file “20151207 deduplicated-source.csv”.

However, no duplicates are found in the above process.

I saved the above configuration in “20151207frildedup.xml” to record the workflow and in the future, I could simply open the XML file and rerun the process.