**KEY**

**Programming language –** mysql command interpreter language

**To run –** run the files in numerical order as “source” arguments

**Data source -** aSQL table of herbarium specimen records: one row per specimen, with a minimum of two columns: 1) a specimen identifier and 2) a collector’s specimen description written in Spanish, Portuguese, English, French, German, or Dutch. The description field, referred to below as the “collector’s notes”, must be in text format.

**File-by-file documentation**

* **File1**

Inspects the first word of collector’s notes for habit keywords; bins hits by habit, one bin for each of 12 habits. Removes hits from source table.

* **File2**

Same action as File1, but inspects the second word. After binning, removes “bad” hits from the bins: a hit is bad if its keyword is part of a prepositional predicate phrase.

Files 2-4 act on records remaining in the original source table; they do not inspect the contents of bins from previous files.

Each of Files 1-4 creates a separate sets of bins, so when File4 has finished, there are as many as 12X4 = 48 bins; File7 will consolidate these 12X4 bins into 12 bins. Scripts create bins only if there are hits to fill them, so when File4 finishes, number of bins may be less than 48.

* **File3, File4**

Same as File2, but inspect third and fourth word, respectively

* **File5**

Cleans up initial inspections: inspects contents of selected bins and returns hits from this inspection to the source table. This step is a post-hoc fix and affects very few records.

* **File6**

Returns all bad hits to source table.

* **File7**

Consolidates the bins (up to 48 bins) into 12, one for each habit.

**Form of data when scripts have finished -** the original table is now subdivided into thirteen tables, twelve of them filled with records of putative habit types (a different type for each table), and the last filled with records with no putative habit type. No records have been annotated or changed, only regrouped.

**Note on text character requirements –** To prepare the text fields, replaceEuropean characters (vowels such as á or consonants such as ñ) with their nearest English counterparts.

**Note on annotations in the scripts –** annotations begin with “--###” if they were written for use with this Key. Other annotations predate this key.