**Multi-Function File Utility — GUI Version (Release 2025-08-28-a)**

Last Updated: Aug 28, 2025

Author: Marc DeJong

Environment: Python 3.12+, Tkinter; pandas & dask (for selected features)

## Overview

A modular, GUI-based toolkit for working with structured CSV/TXT files. It supports both low-memory streaming operations and full-file (Dask) operations, selected per function. The main entry point is:

python file\_util\_gui\_0x.py

Note: the prior ReadMe (2025-06-14) listed #9 as “split\_file\_by\_condition”. This release formalizes Option 9 — Split by Key (Match / No Match) and documents UI/routing improvements.

## Supported Functions (GUI)

|  |  |  |
| --- | --- | --- |
| # | Function name | What it does |
| 1 | filter\_omit | Remove records matching rules |
| 2 | filter\_select | Keep records matching rules |
| 3 | replace\_rec\_contents | Replace values via rules |
| 4 | add\_rec\_stub\_(fixed) | Add a column with a fixed value |
| 5 | add\_rec\_stub\_(var\_from\_rec\_contents) | Add a column derived from other fields |
| 6 | Scissors — Column Chooser | Pick/arrange columns; writes report |
| 7 | sort\_records | Sort by one or more fields |
| 8 | dedupe\_records | Remove or extract duplicates |
| 9 | split\_by\_key (Match / No Match) | NEW: Split MAIN by presence in KEY |
| 10 | split\_by\_composite\_condition | Split by multiple field conditions |
| 11 | concatenate\_files | Combine files vertically (fast-path) |
| 12 | merge\_by\_key | Merge/Roll-in by key (fast-path) |
| 13 | compare\_two\_files\_gui | Compare two files; write compare report |

## What’s New in this Release (2025-08-28-a)

### Option 9 — Split by Key (Match / No Match)

Modules: key\_split\_9x.py, key\_split\_ui\_9x\_gui.py

Summary: Build a key set from a KEY file, then stream MAIN once, writing each row to MATCHED or NONMATCHED outputs with no column modifications.

Features:

• Header-aware preview and manual field mapping (MAIN ↔ KEY).

• Normalization options: case-insensitive, whitespace strip.

• Auto zero-pad widths learned from KEY (or provide explicit widths).

• Encoding probe (UTF-8 fallback to cp1252).

• Outputs: <main>\_MATCHED.csv, <main>\_NONMATCHED.csv, <main>\_SPLIT\_COUNTS.txt (totals + mapping).

Performance notes:

• MAIN is streamed one pass; memory use is primarily the KEY set size.

• Works well with very large MAIN when KEY is modest.

### Routing & UX updates

• Fast-path launches when only #11 or #12 is selected (as before, now documented).

• Option 9 integrates into the 0x controller: if selected with others, it runs first, then remaining operations continue.

• UI polish: reliable label rendering on Windows, delimiter auto-sniff during preview, centered window with on-screen clamping.

## How to Run

Ensure all scripts are in the same folder.

Start the main GUI:

python file\_util\_gui\_0x.py

Option 9 quick start:

• Pick Main file and Key file.

• Click Preview headers.

• Select matching fields on each side and Add Pair (repeat for composite keys).

• Optional: override output filenames.

• Click Run Now. A summary dialog shows totals; the control report is saved beside MAIN.

## Outputs & Logs

Each function writes results to the chosen directory; most use MAIN’s directory by default. A run-level log (file\_util\_log.txt) is written next to the launcher and includes Option 9 totals.

## Folder & Module Structure (selected)

|  |  |
| --- | --- |
| File | Purpose |
| file\_util\_gui\_0x.py | Main controller (routing + fast-paths) |
| key\_split\_9x.py | Option 9 engine (streamed split by key) |
| key\_split\_ui\_9x\_gui.py | Option 9 GUI (mapping, options, run) |
| merger\_gui\_10x.py | Option 12 GUI/engine |
| concatenate\_ui\_11x\_gui.py | Option 11 GUI |
| Two\_File\_Compare\_gui.py | Option 13 standalone compare GUI |

## Requirements

Python 3.12+; Tkinter (bundled). For selected full-file operations:

pip install pandas dask

## Known Limits & Tips

• Delimiter mismatch: If a field preview shows a single long header cell, use Auto-detect or re-run Preview; the UI also attempts auto-sniffing.

• Large KEY sets: Memory scales with KEY unique combinations. Use minimal columns for matching.

• Headers: For accurate mapping, ensure “has header” is set correctly for both files.

## Release History

• 2025-08-28-a — New: Option 9 — Split by Key (Match / No Match); 0x integration & UI fixes.

• 2025-06-14 — Initial GUI ReadMe and function table up to #13 (compare).