Wine price table extraction workflow

Red = R Script

Blue = Output

Green line = done by run_wine_database.R

Folder of images, e.g. UCD_Lehmann_0011.jpg

run_wine_database.R

run_wine_price_tables.R

Extract prices and corresponding names from each image

run_parse_items.R

Match names to fields such as region, producer and color in a Kaggle database

run_wine_evaluate.R

Internal and external (vs. truth) validation of price tables

Complication of truth data

.RDS files (R data type), e.g. UCD_Lehmann_0011.RDS

of prices and names detected

Price List	page.cols items with all price info	Name List (Re-OCR)	Name List (Original OCR)

parse_folder_sample.RDS

ENTRY_PAGE

ENTRY_PRICE

NAME_MATCH

ENTRY_NAME

truth_all.csv

output_summary_internal.csv
output_summary_internal_singlestat.png
summary_vs_truth.csv

name_summary_global_stats.csv
(summary of name dictionary hits)

PRICE_NAME