

电子表格克隆检测

张雅坤, 窦文生, 朱家鑫, 徐良, 周智勇, 魏峻, 叶丹

Learning to Detect Table Clones in Spreadsheets (ISSTA 2020)

联系方式: 张雅坤, zhangyakun18@otcaix.iscas.ac.cn, <http://www.tcse.cn/~zhangyakun18>

Table Clone

Table clone: Two tables share the similar computational semantics, and prescribe the similar business task.

Cloning is common in spreadsheets !

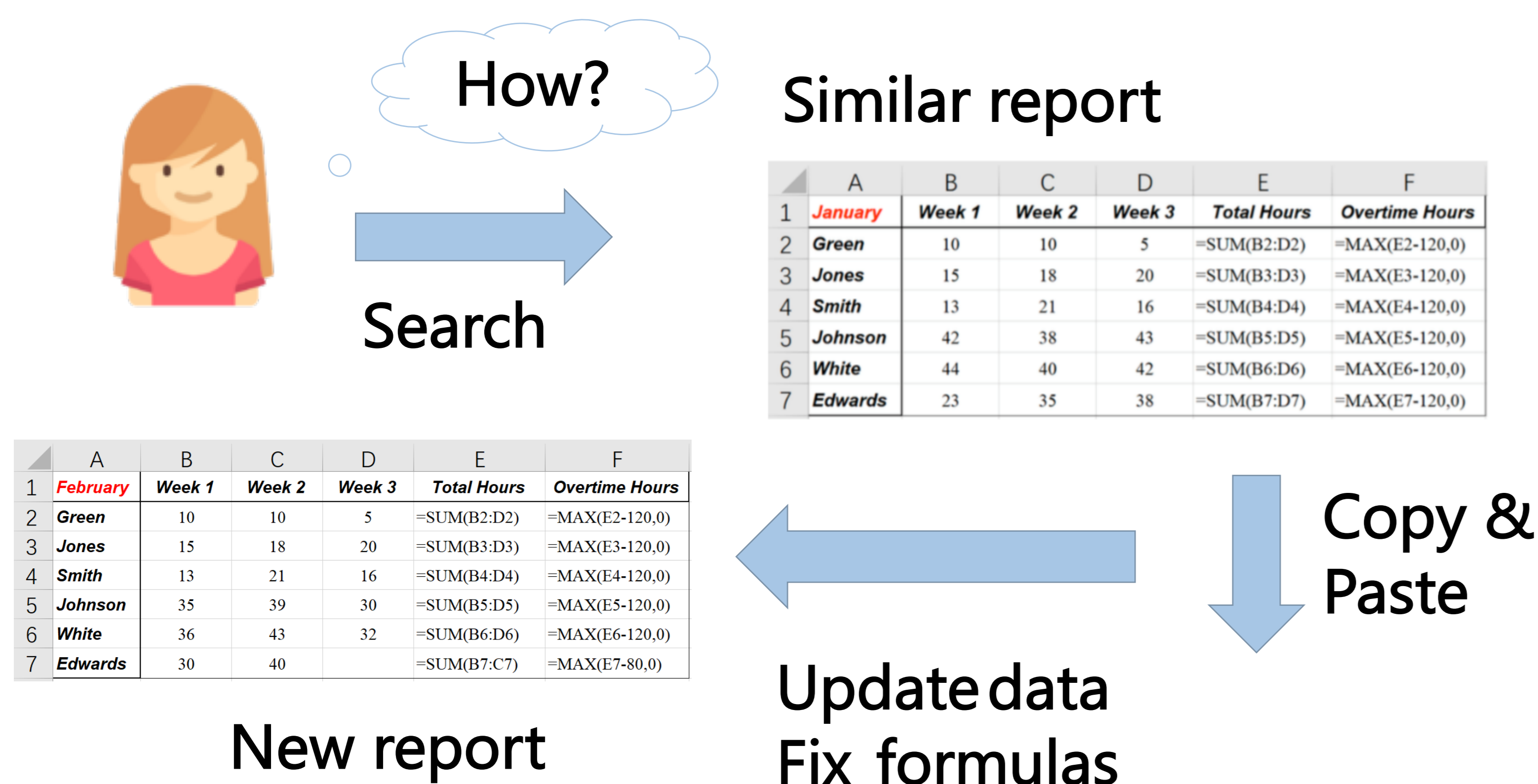
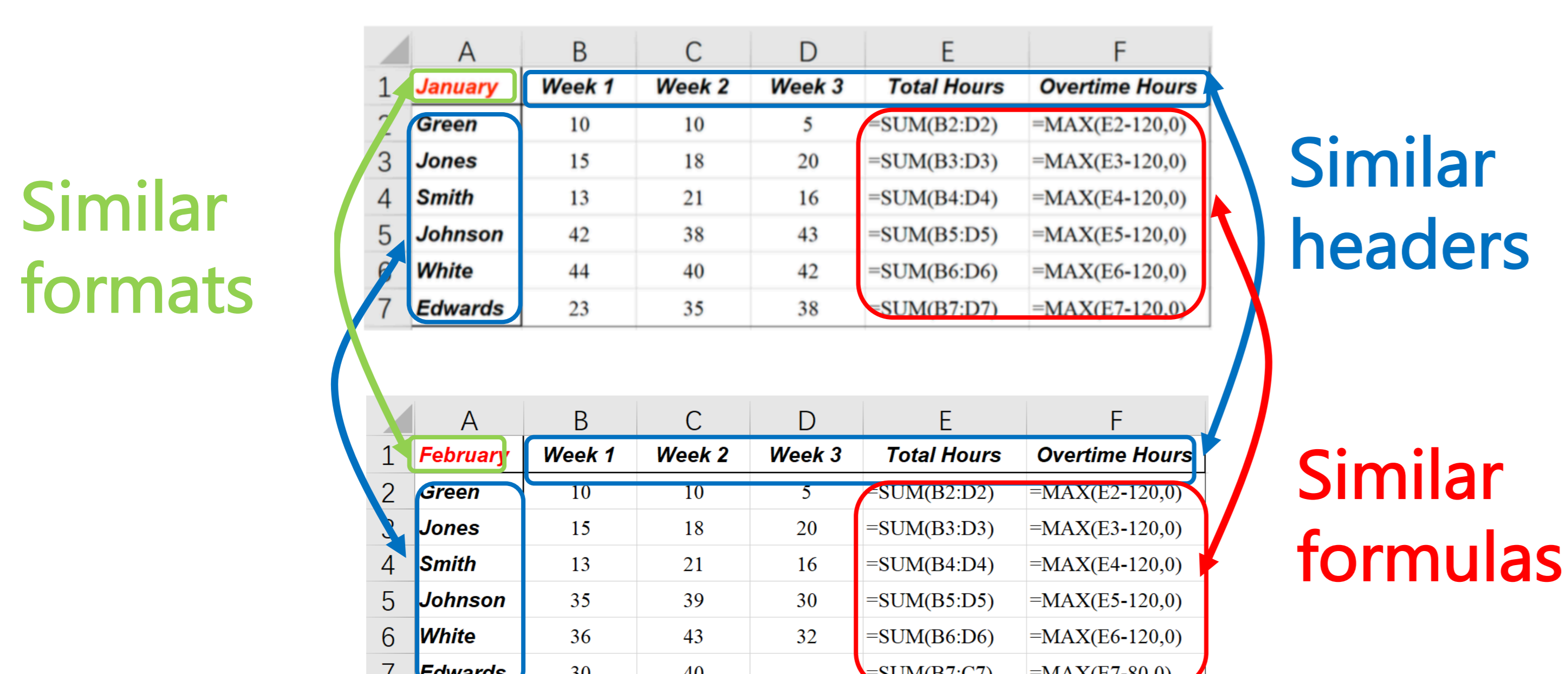


Table Clone Types

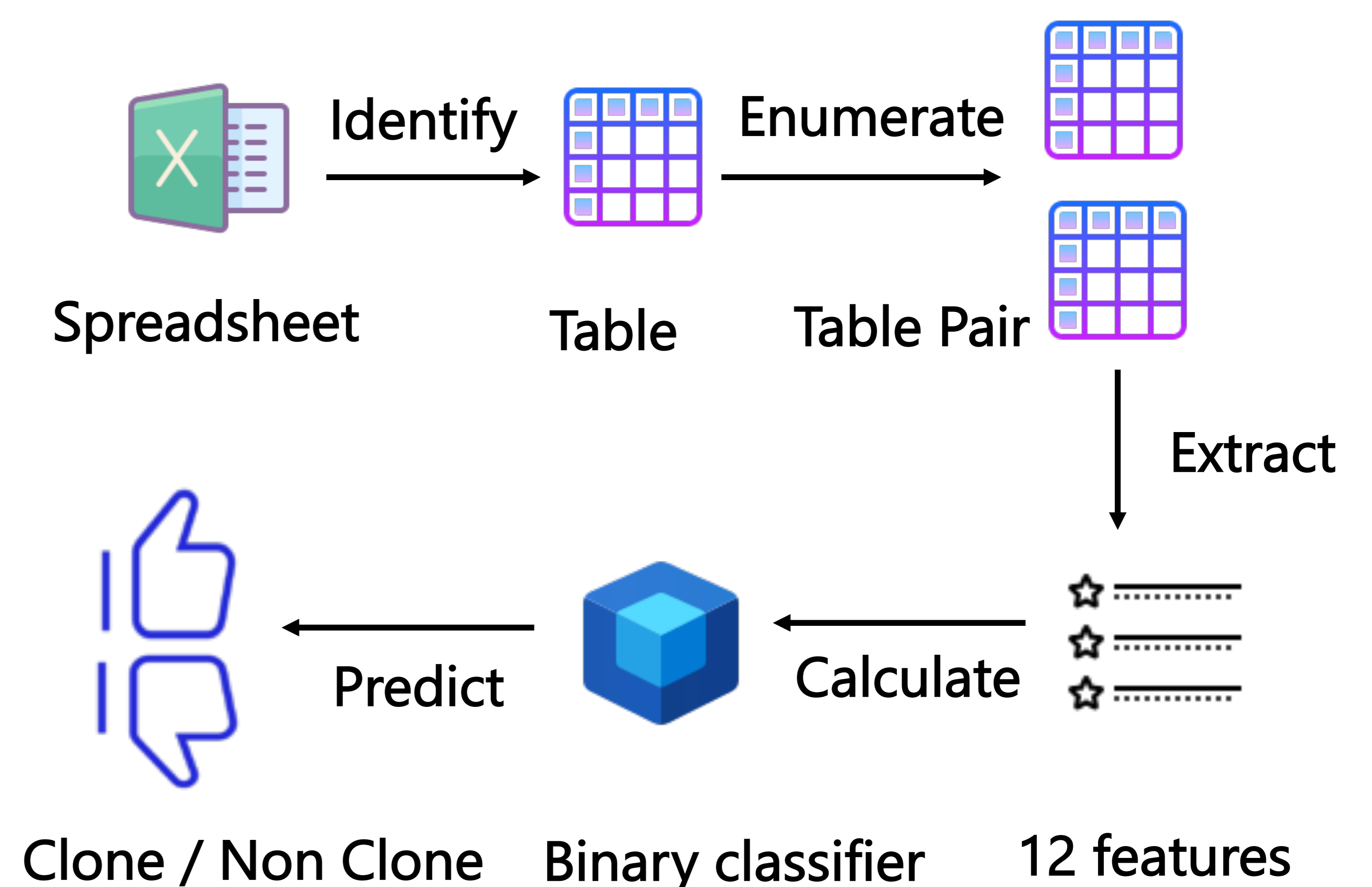
- ❑ Type-1: Identical tables allowing variations in **formats**.
- ❑ Type-2: Table clones have variations on **data and formulas**, and also variations in Type-1.
- ❑ Type-3: Table clones have variations on **table headers**, and also variations in Type-2.
- ❑ Type-4: Table clones have variations on **rows / columns**, and also variations in Type-3.

Identify Table Clones

Key observation: Table clones share **the similar structures and formats**, even though they contain various variations.



Approach Overview



We propose a learning-based approach, LTC, to detect table clones with or without structure changes.

- ✓ Identify Tables
- ✓ Extract 12 features about structures and formats
- ✓ Computer feature similarities for every table pair
- ✓ Predict table clones by Leveraging a binary classification on feature similarities

LTC Evaluation

LTC can detect table clones with and without structure changes precisely.

	TableCheck	LTC
Precision	37.5%	97.8%
Recall	11.1%	92.1%
F1-measure	17.1%	94.9%

Spreadsheet	Intra-Clone	Inter-Clone
Corpus	Precision	Precision
EUSES	97.0%	86.0%
Enron	93.0%	94.0%
FUSE	98.0%	97.0%
Total	96.0%	92.3%