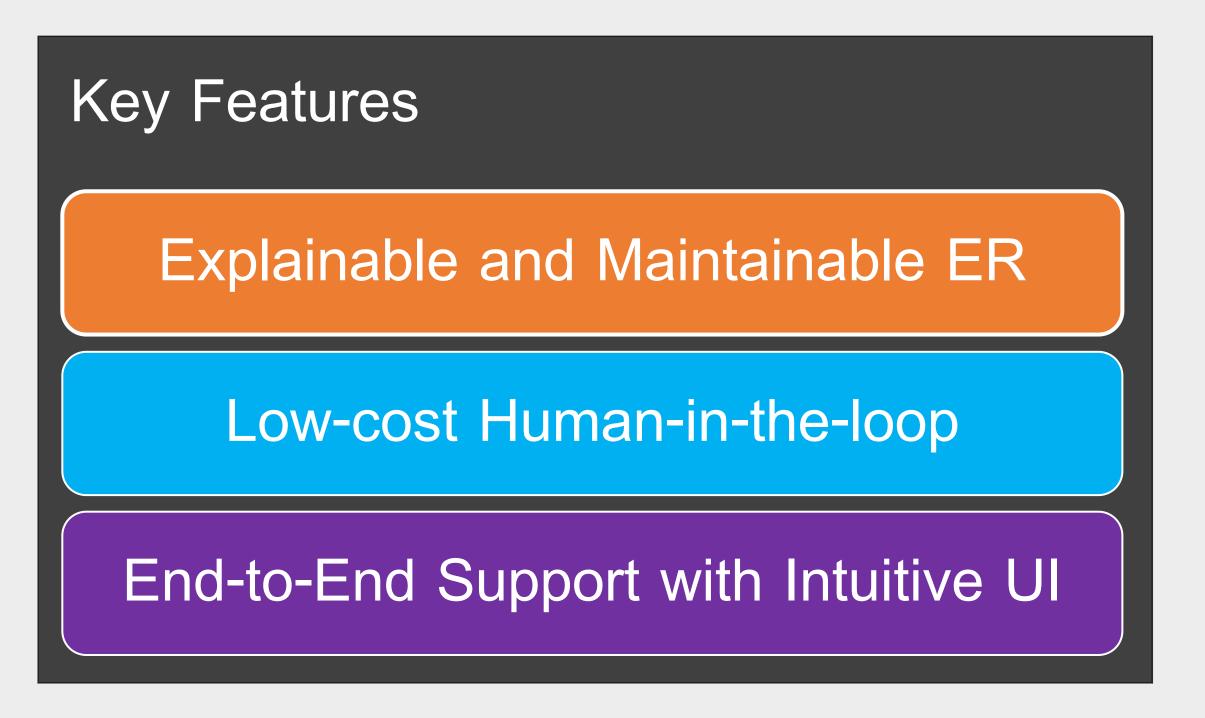
Pvpcpj BO: x Erj k9ck9cgb9illmPvpcpj doBumick ibBkcpcp Obplinctlk

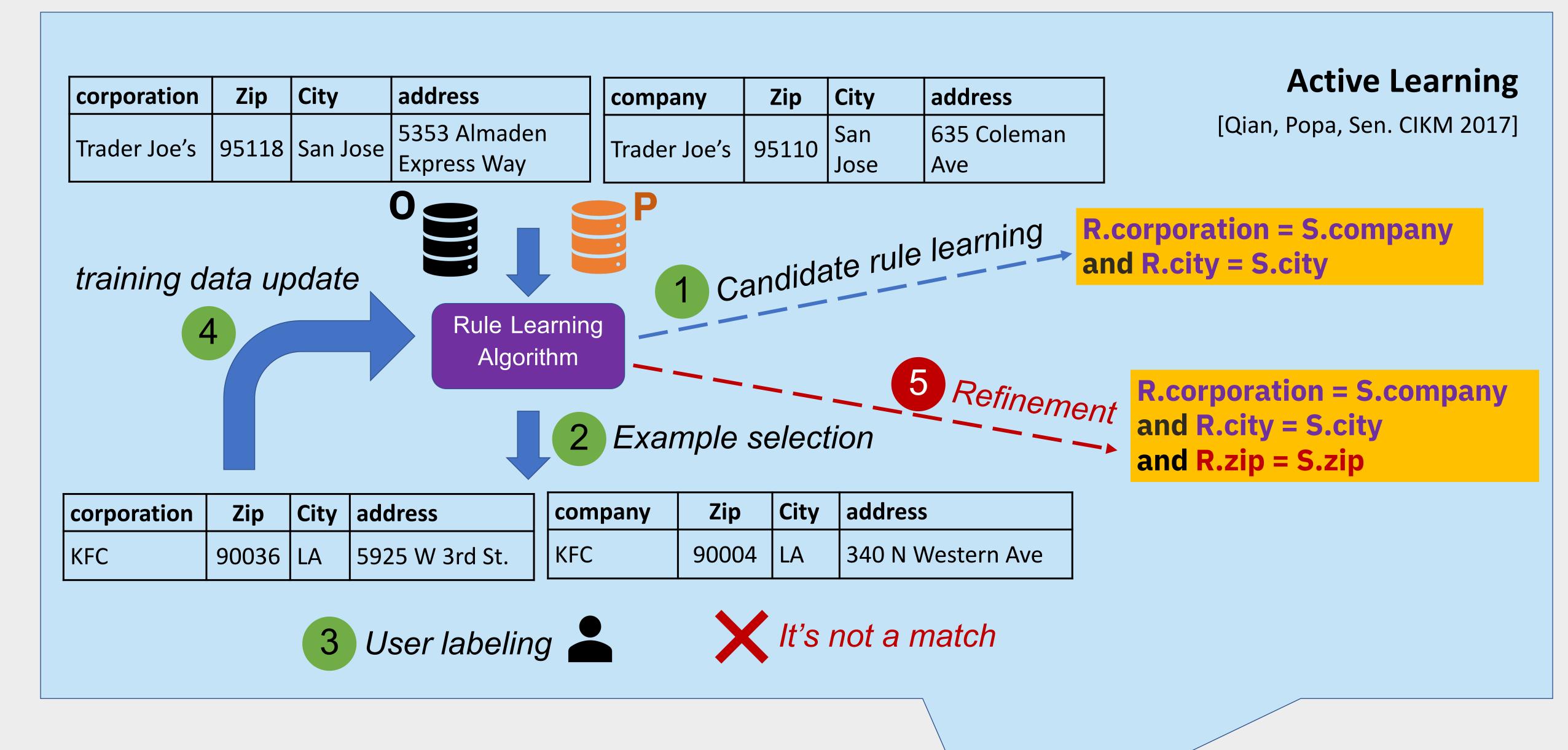


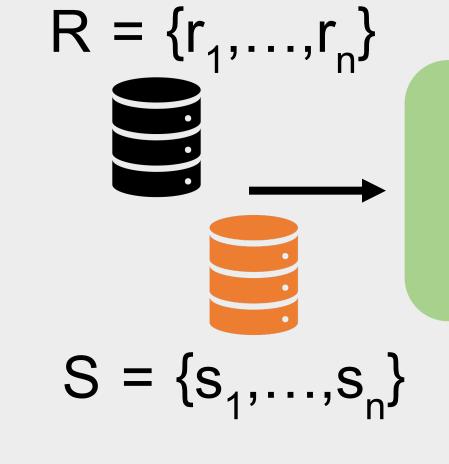
Kun Qian (qian.kun@ibm.com) Lucian Popa (lpopa@us.ibm.com) Prithviraj Sen (senp@us.ibm.com)



SystemER supports:

- CSV or newline delimited json files.
- Automatic schema inference.





Data Preparation

Training data is optional:

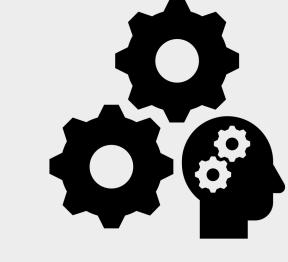
(r_x, s_y) ★ (r_p, s_q) ↓ 2

Training data creation

JaccardSim,
editDistance,
NormalizeName,

Feature
Engineering

Model Learning



Explainable & Customizable ER model

HIL language¹

SOOK THE

SystemER provides:

- a library of commonly used similarity functions/normalization/filtering functions
- a auxillary tool named LUSTRE for learning more complex normalization functions.

Labeled data is acquired dynamically during

active learning

- A heuristic is used to generate a candidate rule to start

¹[Hernandez et. al. EDBT 2013]