Scrabble: Converting Unstructured Metadata into *Brick* for Many Buildings

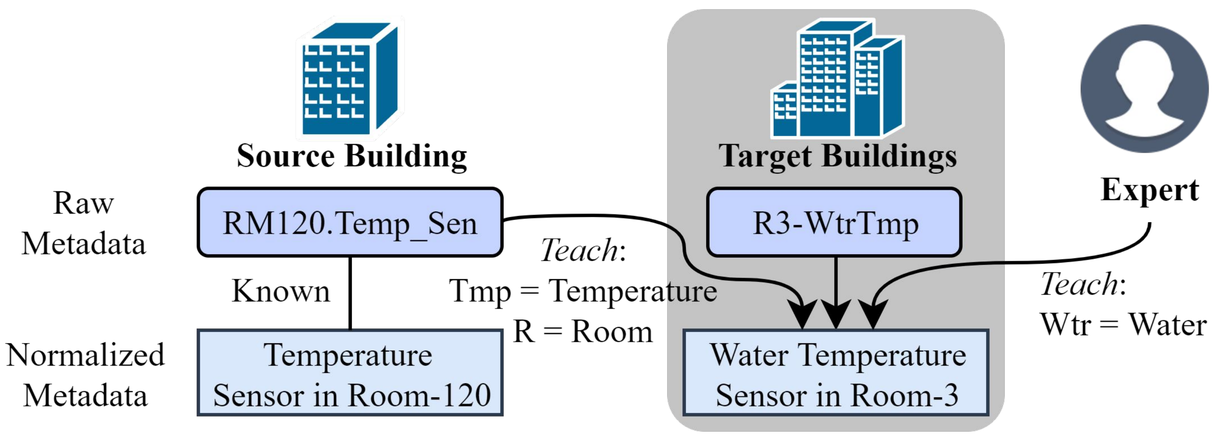
Jason Koh[1], Bharathan Balaji[2], Dhiman Sengupta[1], Julian McAuley[1], Yuvraj Agarwal[3], Rajesh Gupta[1]
University of California, San Diego[1], University of California, Los Angeles[2], Carnegie Mellon University[3]

Why Metadata for Buildings?

- The major bottleneck to deploy modern building applications is *the large human effort to map "metadata" into a usable format*.
- Vendor-given metadata commonly contains:
  ○ Point Type ○ Location ○ Equipment Name ○ Network Interface
- Unstructured Metadata Examples:

Vendor-Given	Expert's Interpretation		
Raw metadata	Sensor Type	Location	Equipment
ENG.CRAC-1.TEMPSETF	Zone_Temp_Setpoint	N/A	CRAC-1
SC-CRAC-1-MIG-008.Tmp	Temp_Sensor	N/A	CRAC-1
SC.3FLW-HALL.ZN-T	Zone_Temp_Sensor	Floor-3, W-Hall	N/A

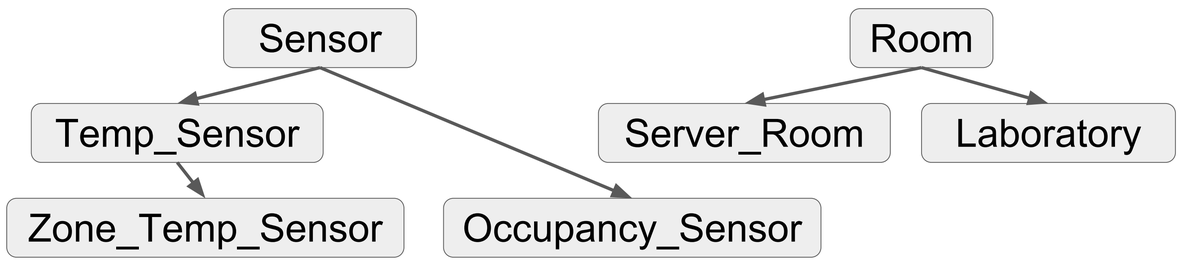
- Reduce the human effort by *existing buildings already normalized.*



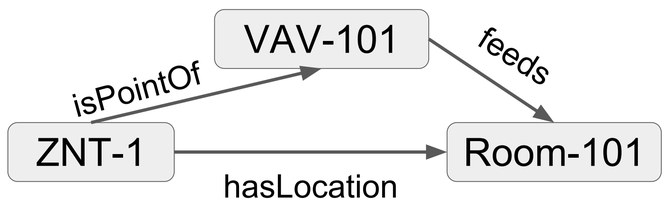
What is *Brick*?

- ***Building Metadata Schema* designed for *Portable Applications*.**
- Describe all knowledge in a directed graph. Everything is a node (or an entity) that have relationships with each other.
  E..g, ZNT-1 is an instance of
- Tags constitute TagSets.

- TagSets have hierarchy for different levels of specification and categories of TagSets.

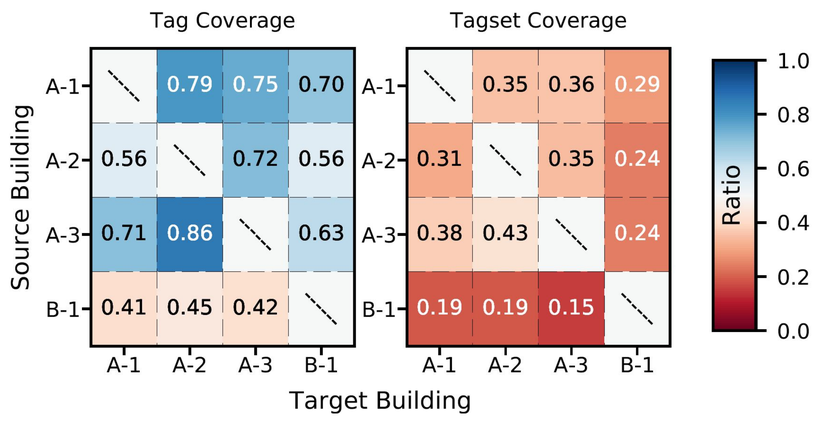


- Entities can have relationships with each other defined in Brick.

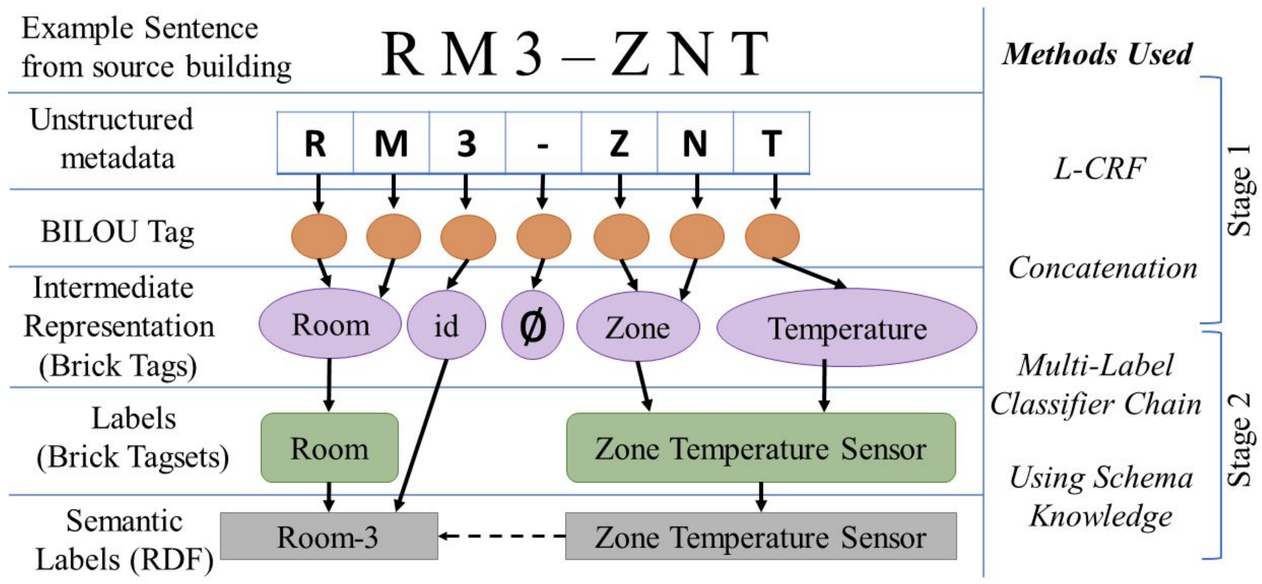


Characteristics of Different Buildings

- We compare 3 buildings from UCSD and 1 building from CMU.
  ○ Tags are more common than TagSets across different buildings.



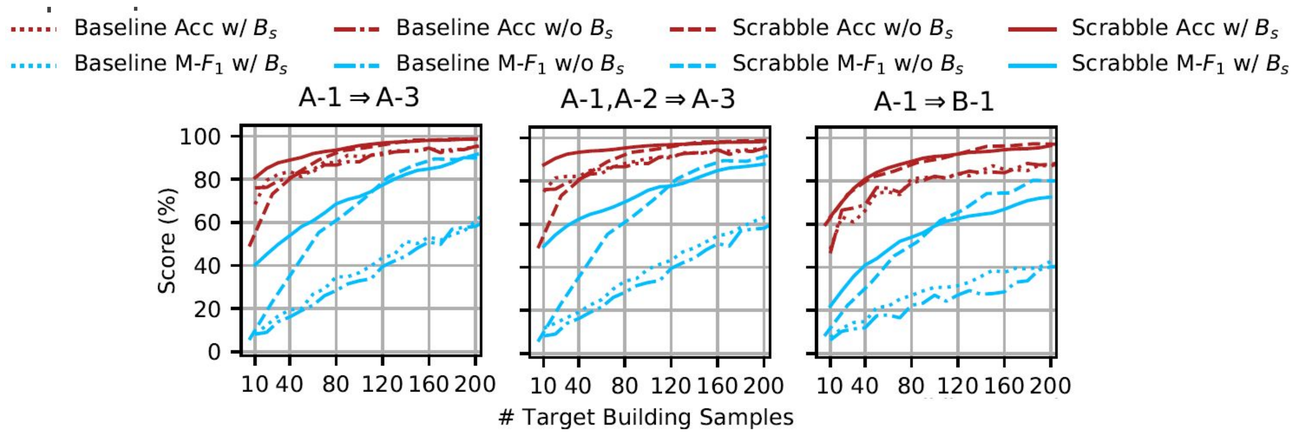
Scrabble Framework



- **Basic Idea:**
  ○ Character => Tags would be more reusable than TagSets
  E.g., If ZN=Zone is known, ZN in ZNT, ZN-T and ZN-1 can be known.
  ○ Tags => TagSets are given or easy to learn.
  E.g., obviously, {Zone, Temperature, Sensor}
  => Zone_Temperature_Sensor
  ○ Two stages learning will help reusing existing knowledge.
- **Character-level entity recognition using CRF.**
  ○ R -> Beginning of RM for "Room". M -> Inside RM for "Room" (BIO)
  ○ Provide character-level error resiliency.
  ○ No predefined delineation rule is required.
- **Mapping raw metadata to Tags as Intermediate Representation (IR)**
  ○ RM -> Room, ZN -> Zone, T -> Temperature
  ○ Mapping to IR is easier than to exact TagSets.
- **Mapping Tags to TagSets by a multilabel classifier.**
  ○ {Room, Zone, Temperature} -> {Room, Zone_Temperature_Sensor}
  ○ This layer is resilient to variations as mappings are somewhat known by Brick's structure.
  ○ *Structured Classifier Chain* is proposed for multi-label classification.
- **Select most informative samples to learn from experts.**
  ○ Metrics: confidence-based metric, raw metadata utilization metric.
  ○ Ask examples in the test set with low scores to experts.
  And iterate the entire process.

Evaluations

- Active learning setup: At each iteration, ask selected 10 samples to experts for labels.
- Baseline: BoW -> Multi label classification -> Entropy-based active learning



  ○ Scrabble outperforms the baseline in any cases.
  ○ The learning speeds between Scrabble w/ and w/o source samples converge around 100 samples, which should not be.
- Logical comparison with existing work.
  ○ The baseline: two different buildings share less features.
  ○ CNN for text classification: good accuracy but not suitable for active learning / transfer learning framework.
  ○ Zodiac: Limited to multi-class classification.
  ○ Bhattacharya et al.: rules are strict and not designed for transferring knowledge to other buildings.

UC San Diego