

Category Specific Task Extraction

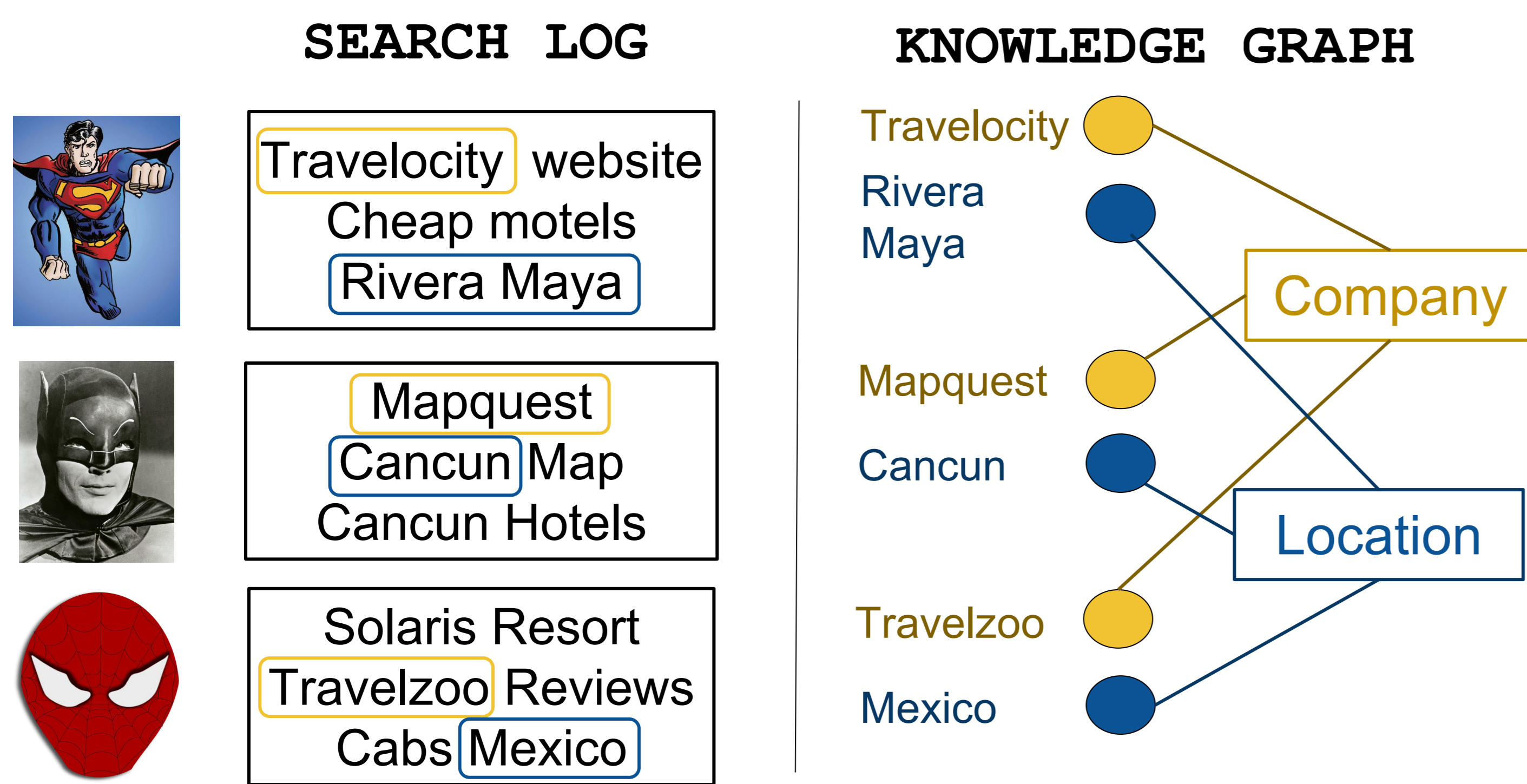
Manisha Verma
manisha.verma.13@ucl.ac.uk

Emine Yilmaz
emine.yilmaz@ucl.ac.uk



MOTIVATION

- Search queries are an outcome of some **task**.
- And **tasks** revolve around **real world objects** i.e. **Entities**.
- Can **entities** and their categories **be used to extract Tasks**?



CONTRIBUTIONS

- Exploit **entities** to **extract tasks** from search logs.
- Use **entity type (category)** information to group related tasks.
- Use **implicit** and **explicit** means to evaluate our approach with **existing task extraction** baselines.

METHODOLOGY

A single user goal can be represented by **several queries**.
Cluster queries in each category into **distinct tasks**.

Entity Tagging and Query Aggregation

- Identify all the entities in a query.
- Aggregate queries as per their entity-type.

Task Extraction Algorithm

Input: Category hierarchy \mathcal{H} , Queries \mathcal{L} , min_{sp} , min_{sc}

Output: Category hierarchy \mathcal{H}' with tasks

```

1: for  $q_i, q_j \in \mathcal{L}$  do
2:    $sim[q_i][q_j] = \sum_k w_k \cos(f_{ki}, f_{kj})$ 
3: end for
4: for category  $C_i \in \mathcal{H}$  do
5:   for parent  $P_{ij} \in par(C_i)$  do
6:     if  $\frac{|C_i \cap P_{ij}|}{|C_i \cup P_{ij}|} > min_{sp}$  then
7:       merge( $C_i, P_{ij}$ )
8:     end if
9:   end for
10: end for
11: for category  $C_i \in \mathcal{H}$  do
12:    $C'_i = dpmeans(C_i, sim)[8]$ 
13: end for
14: for category  $C'_i \in \mathcal{H}'$  do
15:   for queries  $q_{ij} \in C'_i$  do
16:     if  $sc(q_{ij}) < min_{sc}$  then
17:       merge with nearest cluster  $\in \{par(C'_i) \cup child(C'_i)\}$ 
18:     end if
19:   end for
20: end for

```

Category based operations:

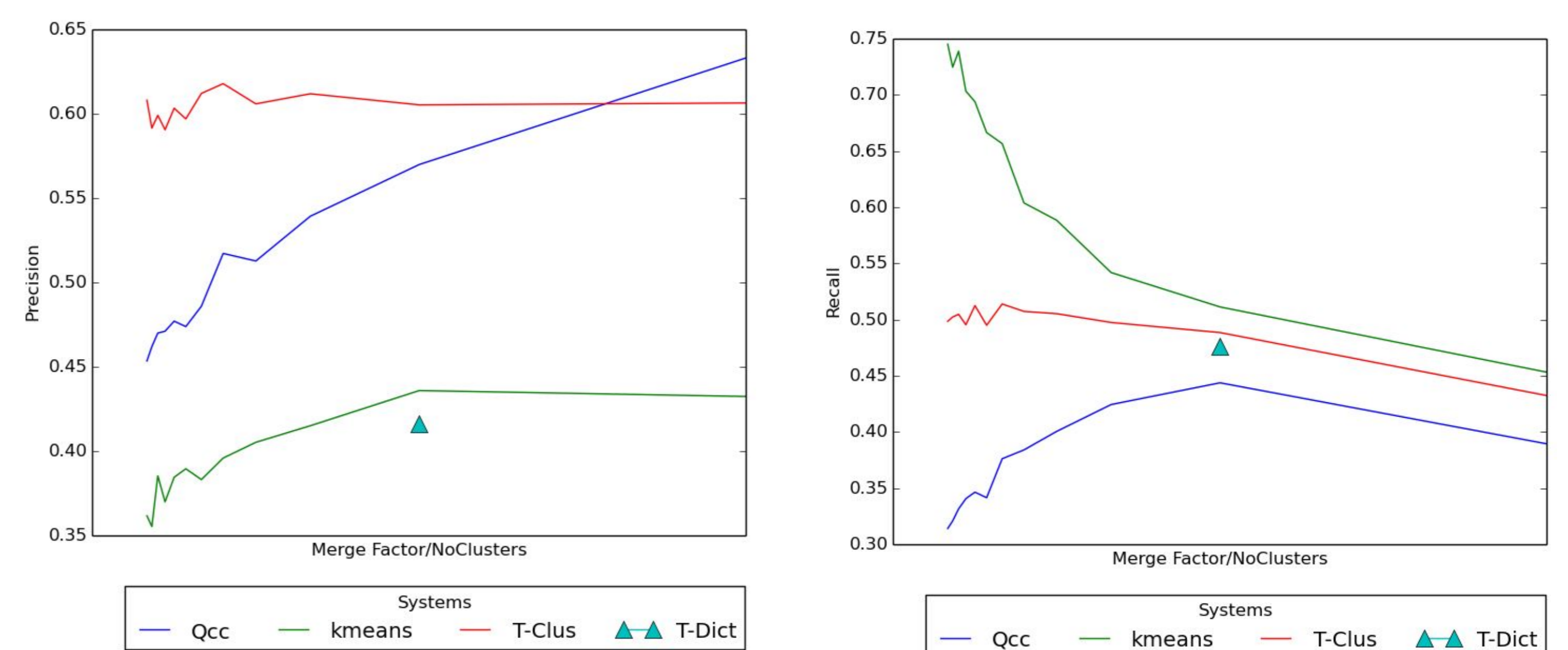
- Merge queries of **low resource categories with parent node**.
- Queries may be **incorrectly mapped** to a task cluster.
- **Re-assign** such queries with low cluster scores to tasks on parent or child nodes.

EVALUATION

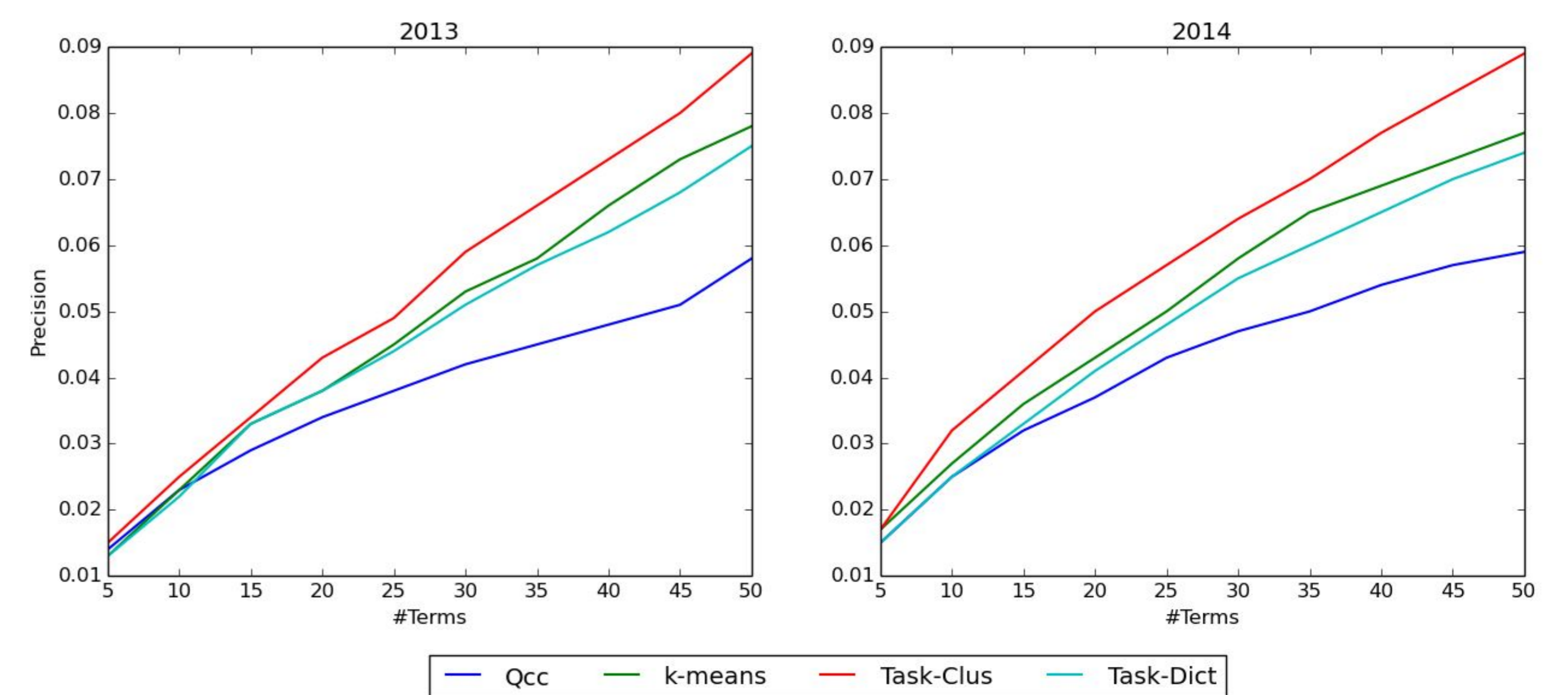
Intrinsic Evaluation

	QCC [1]	K-Means	T-Clus	T-Dict [2]
Intra-cluster	0.56	0.57	0.43	0.13
Inter-cluster	0.82	0.84	0.98	0.11

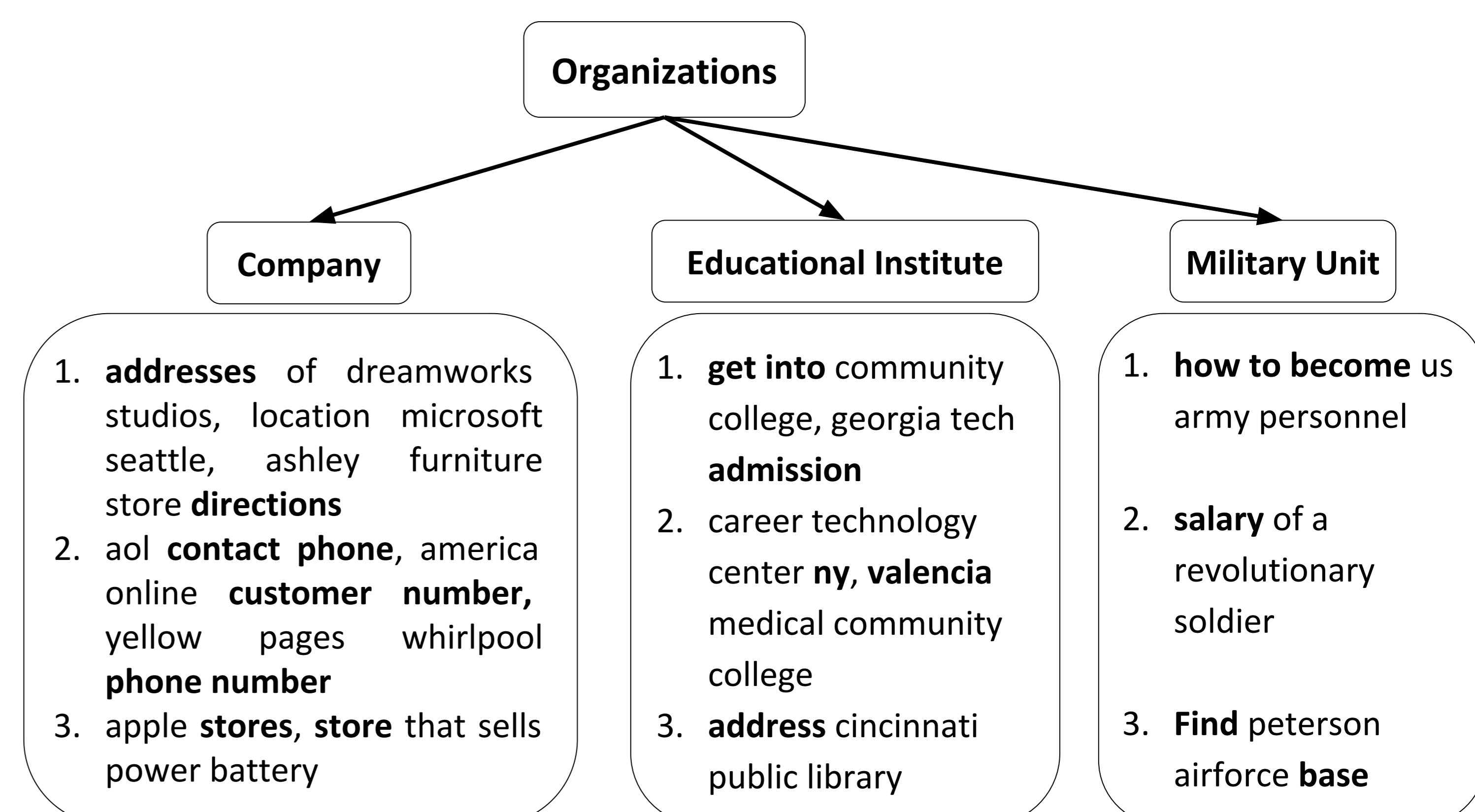
Precision - Recall Curves



Term Prediction



EXAMPLE OUTPUT



References

- C. Lucchese, S. Orlando, R. Perego, F. Silvestri, and G. Tolomei. Discovering tasks from search engine query logs. *ACM Trans. Inf. Syst.*, 2013.
- M. Verma and E. Yilmaz. Entity oriented task extraction from query logs. In *CIKM 2014*.