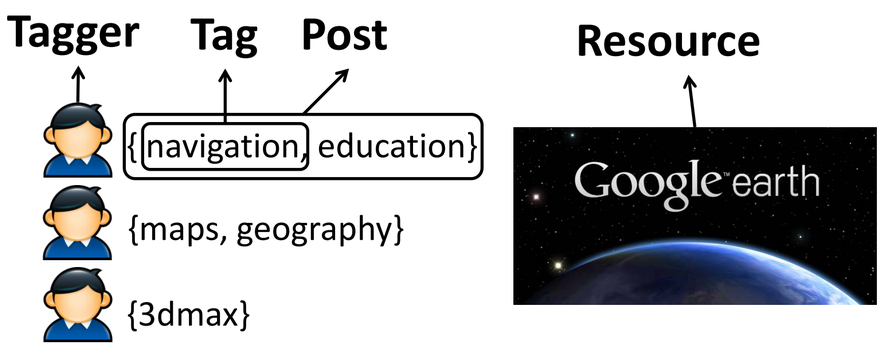
THE UNIVERSITY OF HONG KONG
DEPARTMENT OF COMPUTER SCIENCE

iTag: Incentive-Based Tagging

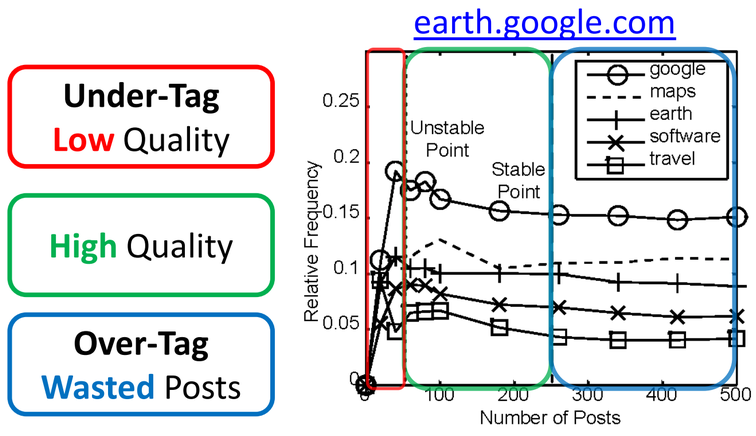
Siyu Lei, Xuan S. Yang, Luyi Mo, Silviu Maniu, Reynold Cheng
Department of Computer Science, The University of Hong Kong

Introduction

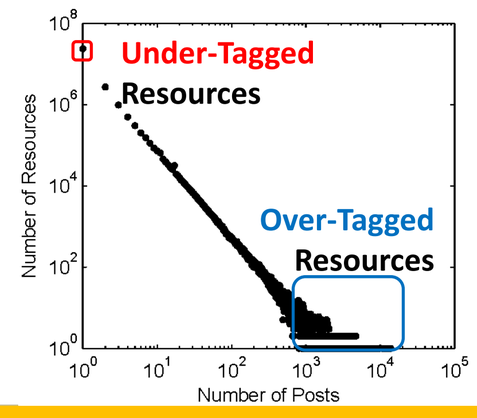
❑ ***Collaborative Tagging Data:*** facilitate many applications



❑ Resource needs ***sufficient number*** of posts to get high-quality tag data.



❑ **Under-Tag** and **Over-Tag** in Collaborative Tagging Systems



❖***Problem***: *how to improve the quality for collaborative tagging data?*

Approach

Tag Quality

❑ ***Relative Frequency Distribution (rfd)*** $\vec{F}_i(k)$**:** Normalized number of occurrence of each tag, after resource r_i has k posts.

❑ ***Stability*** $m_i(\omega, k)$ **:** Average similarity of ***rfds'*** within window $[k-\omega, k]$.

❑ ***Stable Point:*** When stability score surpasses a threshold τ.

❑***Over –Tagging:*** Posts given to resources that has passed stable point.

❑ ***Tag Quality***:

- For resource r_i: $q_i(k)$ defined on stability score.
- For resource set ***R:*** $q(R, \vec{k}) = \frac{1}{n}\sum_{i=1}^{n} q_i(k_i)$

Incentive-Based Tagging

❑ ***Intuition***: Find the optimal ordering of the resources to achieve the best tagging quality.

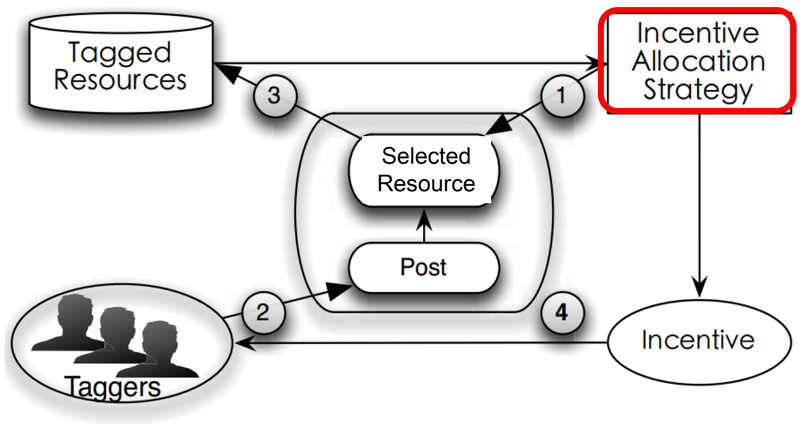
❑ ***Input***: A set of tagged resources and budget.

❑ ***Output:*** Incentive Allocation.

❑ ***Objective:*** Maximize Tag Quality.

❑ ***Optimal Solution:***

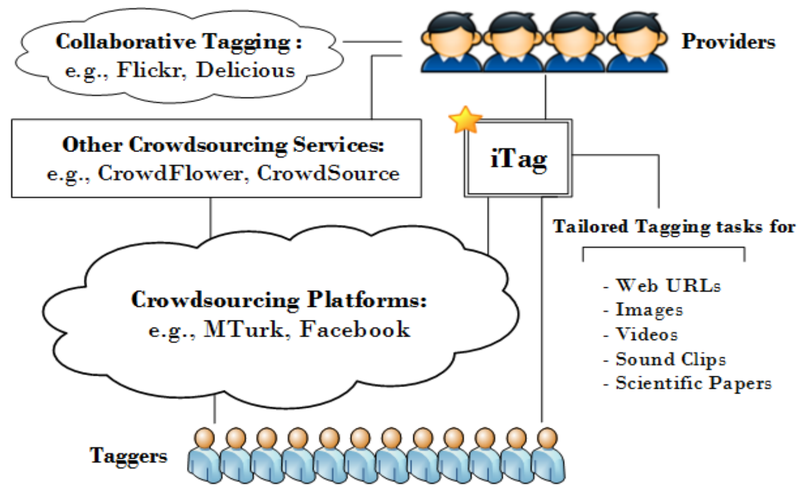
- Dynamic Programming
- Need to know the posts in the future.



Money, Online Storage, Free Membership

Reference: X. Yang, R. Cheng, L. Mo, B. Kao, and D. Cheung "On Incentive-Based Tagging," ICDE, 2013.

Overview



Incentive Allocation Strategies

❑ ***Random (R)***:
Randomly allocate resources to taggers to tag.

❑ ***Fewest Post First (FP)***:
Prioritize the under-tagged resources.

❑ ***Most Unstable First (MU)***:
Prioritize the most unstable resources ; window size ω .

❑ ***Hybrid (FP-MU):***
FP first, switch to ***MU*** when each resource has ω posts.

Results

❑ ***Dataset****:* 5000 urls and their posts from *del.icio.us*

Figure1: Quality vs. Budget

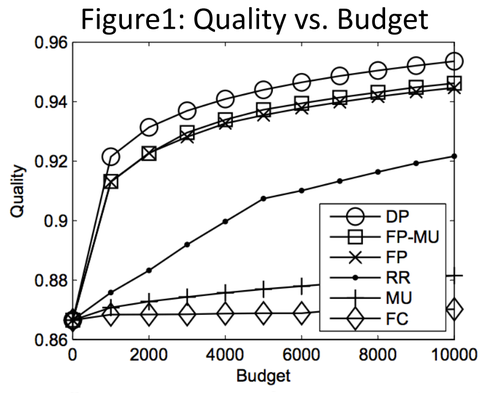
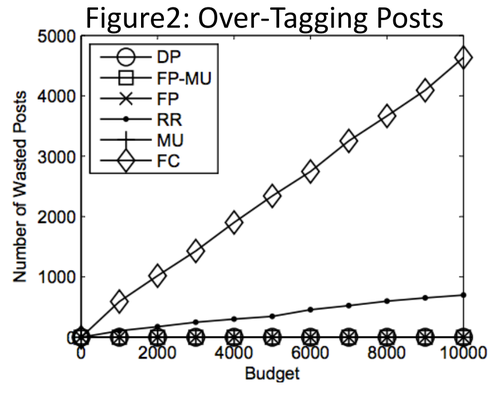


Figure2: Over-Tagging Posts



❑ ***FP*** & ***FP-MU*** close to optimal;
- ***FC*** hardly increases the quality.

❑***50%*** of the posts by FC are over-tagging.

❖***Conclusion***: *iTag can significantly improve tag data quality for providers with the least amount of money.*

Contact

Siyu Lei
sylei@cs.hku.hk