# Tag Recommendations for StackOverflow Posts

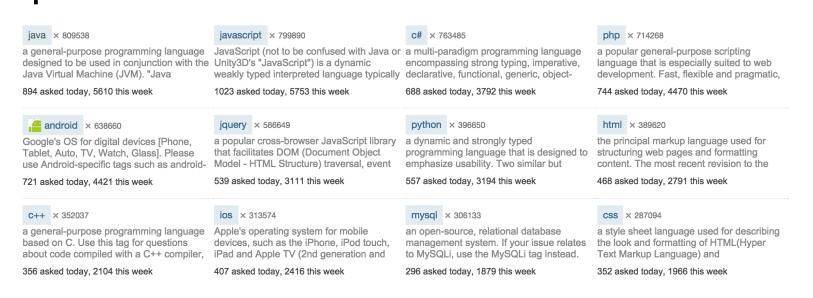
CSE 6240 Project Proposal

Ke Wang

Haochen Zhao

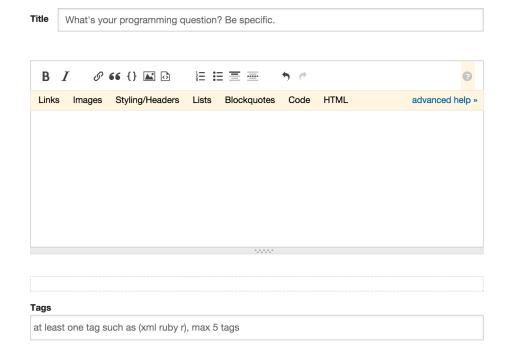
## Tags and StackOverflow

- Tag: a type of meta-information
- Usage: indexing, categories, search terms, personalized bookmarks



#### Problem

 Problem: When a user (maybe a newcomer) writes a new post, can we recommend some tags?



## Traditional Approaches

- TF-IDF
  - Extract possible keywords from the post
  - Compute TF-IDF, and return top terms
- Similarity-based
  - Compute similarity based on TF-IDF
  - Recommend terms to the new post
- Tag-term co-occurrence

• LDA 
$$P(c|d, \alpha, \beta) = \sum_{t=1}^{T} P(c|t, \beta) P(t|d, \alpha)$$

### More Approaches

- pLSI-based and LDA-based Tag Recommender
  - Main idea: compare documents' topic distribution; assign tags to similar documents.

#### Datasets

- StackOverflow Dumps: posts, post links, tags, users
- ECML PKDD Discovery Challenge 2009

#### References

- [1] Zhang, N., Zhang, Y., & Tang, J. (2009). A tag recommendation system based on contents. *ECML PKDD Discovery Challenge 2009 (DC09)*, 285.
- [2] Xia, X., Lo, D., Wang, X., & Zhou, B. (2013, May). Tag recommendation in software information sites. *In Proceedings of the 10th Working Conference on Mining Software Repositories (pp. 287-296)*. IEEE Press.
- [3] Wang, S., Lo, D., Vasilescu, B., & Serebrenik, A. (2014, September). EnTagRec: An enhanced tag recommendation system for software information sites. In *Software Maintenance and Evolution (ICSME), 2014 IEEE International Conference on (pp. 291-300).* IEEE.
- [4] Dredze, M., Wallach, H. M., Puller, D., & Pereira, F. (2008, January). Generating summary keywords for emails using topics. *In Proceedings of the 13th international conference on Intelligent user interfaces (pp. 199-206)*. ACM.