CSE6240: Web Search and Text Mining, Spring 2013

- **Lecture**: TR 12:05pm - 01:25pm
- **Office hours**: T 1:30pm - 2:30pm
- **Instructor**: Hongyuan Zha, Office: 1314 KACB, zha@cc.gatech.edu, Phone: 404-385-149
- TAs: Liangda Li, Office: 1316 KACB, xiaoxiangqingyu@gmail.com; Hua Ouyang, Office: 1316 KACB, houyang@gatech.edu

---

Course Description

This course will cover the data analytic aspects of three closely related topics: Web search, recommendation systems and social media and network analysis. The emphasis is on methodologies and applications. We will not spend much time on implementations and system issues. Major themes addressed include probabilistic and statistical methods, dynamic user behavior modeling in social media and social networks. Ideally you should have formal exposure to data mining and machine learning at the level of CSE6740 (some of the open access ML courses will be helful) and be comfortable with using a script and/or high-level language.

List of Topics

- Introduction to IR: inverted indices, query processing, tf-idf weighting, scoring, anchor texts, precision and recall, DCG (Week 1: 1/6)
- Link analysis: PageRank and HITS algorithms (Week 2: 1/13)
- Learning to rank methods (Week 3: 1/20)
- Implicit relevance feedback using user click and behavior data (Week 4: 1/27)
- Latent dirichlet allocation and extensions (Week 5: 1/3)
- Introduction to RS: Content-based recommendations, collaborative recommendations, user and item-based methods, matrix factorization, evaluation (Week 6: 2/10)
- Matrix factorization and item/user features (Week 7: 2/17)
- Cold start problems in collaborative filtering (Week 8: 2/24)
- Incorporating multiple data sources in collaborative filtering (Week 9: 3/3)
- Strong and weak ties in social networks (Week 10: 3/10)
- Week 11: Spring Break
- Diffusion in networks (Week 12: 3/24)
- Epidemic/Hawkes models (Week 13: 3/31)
- Structure discovery from user behaviors (Week 14: 4/7)

- Project presentation (Weeks 15 & 16)

Grading

1. Homework assignments: 50%
2. Project: 50%