

مبانی بازیابی اطلاعات

Principles of Information Retrieval

سرفصل درس

- مقدمه‌ای بر بازیابی اطلاعات
- معرفی سامانه‌های بازیابی اطلاعات بولین و نحوه‌ی شاخص‌گذاری (indexing)
- پیش‌پردازش اسناد: عملیات متنی و نرمال‌سازی واژه‌ها
- تصحیح غلط املایی (spelling correction)
- فشرده‌سازی شاخص
- مدل فضای برداری (Vector Space Model) و بازنمایی tf-idf
- امتیازدهی و رتبه‌بندی اسناد
- ارزیابی سامانه‌های بازیابی اطلاعات و معرفی معیارهای ارزیابی
- مدل‌های بازیابی اطلاعات احتمالاتی
- دسته‌بندی اسناد
 - دسته‌بند Naïve Bayes
 - دسته‌بندهای خطی
- خوشه‌بندی اسناد
 - خوشه‌بندی k-means
- موتورهای جستجوی وب
- خزشگر (crawler)
- تحلیل گراف لینک‌ها و پیدا کردن PageRank

ارزیابی درس

- آزمون میان ترم: ۳۵٪
- آزمون پایانی: ۴۵٪
- تکالیف: ۳۰٪

مراجع

C.D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.