

بسمه تعالی

محتوی درس بازیابی پیشرفته اطلاعات

مقدمه:

بازیابی اطلاعات شاخه ای با قدمت زیاد در علوم و مهندسی کامپیوتر است که بصورت سنتی در مورد نحوه ذخیره و بازیابی اطلاعات و مستندات بحث می کند. با پیدایش اینترنت و وب که حجم وسیعی از اطلاعات را در گستره جهانی پوشش می دهد مباحث جدیدی در این رشته ایجاد گردیده است. هدف از درس بازیابی پیشرفته اطلاعات طرح و بحث در مورد این مباحث جدید در بازیابی اطلاعات و معرفی تکنیکها و روشهایی است که در این رابطه معرفی و مورد استفاده قرار می گیرد است.

سرفصل مطالب

- ۱- چهارچوب کلی بازیابی اطلاعات با معرفی اعمال متداول در یک سیستم بازیابی اطلاعات.
- ۲- مدل‌های بازیابی اطلاعات:
 - الف- مدل‌های کلاسیک بازیابی اطلاعات: مدل بولی، مدل برداری و مدل احتمالی
 - ب- مدل‌های گسترش یافته: مدل بولی گسترش یافته، مدل فازی، مدل شبکه های عصبی، مدل LSI، مدل شبکه های بی‌زی
- ۳- نحوه ارزیابی مدل‌های بازیابی اطلاعات. معرفی و بحث در مورد ملاکهای F-Measure, Recall, Precision و غیره
- ۴- زبانهای تعریف پرس و جوها. توضیح در مورد انواع زبانهایی که می توان بوسیله آنها یک پرس و جو را فرموله کرد مانند زبان بولی، براساس کلمات کلیدی، ساختیافته و غیره
- ۵- فرمول سازی مجدد پرس و جوها: معرفی روشهایی برای تصحیح و یا تغییر پرس و جوها برای دریافت جوابهای با دقت بالاتر
- ۶- بررسی خصوصیات متون: قوانینی نظیر قانون Zipf و قانون heap که در نحوه اندیس گذاری مستندات تاثیر دارد مورد بررسی قرار می گیرد.
- ۷- زبانهای متنی: سه دسته از زبانهای غیر ساختاری، نیمه ساختاری و کاملا ساختیافته مورد بررسی قرار می گیرد.
- ۸- پردازش متون: پردازشهای متداول متون نظیر ریشه یابی؛ اندیس گذاری و فشرده سازی مورد بحث قرار می گیرد.
- ۹- بازیابی اطلاعات مستندات غیر متنی: روشهای بازیابی اطلاعات برای مستنداتی نظیر تصاویر مورد بررسی قرار می گیرد.
- ۱۰- بازیابی اطلاعات در محیط های چند زبانی: در نظر گرفتن زبانهای مختلف نظیر فارسی- انگلیسی مورد بحث واقع می شود.
- ۱۱- بازیابی اطلاعات در شبکه های نظیر به نظیر: با معرفی شبکه های نظیر به نظیر مدل‌های جدیدی که در بازیابی اطلاعات در این شبکه پیشنهاد شده مورد مطالعه قرار می گیرد.
- ۱۲- آشنایی با مبحث داده کاوی و ارتباط آن با بازیابی اطلاعات
- ۱۳- آشنایی با وب معنایی و جستجو در آن.

ارزشیابی:

امتحان میان ترم: ۲۰٪

امتحان پایان ترم: ۵۰٪

تکالیف و ۴ پروژه در طول ترم: ۳۰٪

مراجع:

1. Modern Information Retrieval R. Baeza-Yates and B. Ribeiro-Neto; Addison-Wesley, 1999.
2. Information Retrieval Systems: Theory and Implementation G. Kowalski; Kluwer, 1997.
3. Automatic Text Processing G. Salton; Addison-Wesley, 1989.
4. Database System Concepts, 4th Edition A. Silberschatz, H. Korth and S. Sudarshan; McGraw-Hill, 2001.
5. Somen Chakrabatri, Mining the web: Discovering knowledge from hypertext data, Elsevier Science, 2003
6. Various Web resources.