



(کاربرگ طرح درس)

نسخه اولیه: ۱۳۹۷/۱۰/۲۹

تاریخ به‌روز رسانی: ۱۳۹۸/۰۱/۱۷

نیمسال دوم سال تحصیلی ۹۸-۱۳۹۷

فارسی: مبانی بازیابی اطلاعات و جستجوی وب		تعداد واحد: نظری ۳		مقطع: کارشناسی <input checked="" type="checkbox"/> کارشناسی ارشد <input type="checkbox"/> دکتری <input type="checkbox"/>		
نام درس		لاتین: Foundations of Information Retrieval and Web Search				
مدرس: سمانه امامی		شماره تلفن دفتر کار (اتاق ۳۷۳): ۰۲۳-۳۱۵۳۳۷۵۵				
پست الکترونیکی: s_emami@semnan.ac.ir		منزلگاه اینترنتی: http://semami.profile.semnan.ac.ir				
برنامه تدریس در هفته: دوشنبه‌ها (ساعت ۹/۳۰ تا ۱۰/۳۰ - کلاس ۱۱۲) و سه‌شنبه‌ها (ساعت ۱۰/۳۰ تا ۱۲/۳۰ - کلاس ۱۲۱)						
اهداف درس: درس مقدمه‌ای بر بازیابی اطلاعات در دو بخش ارائه می‌گردد. در بخش اول آشنایی با ساختار فایل‌های مختلف و نحوه ذخیره و بازیابی اطلاعات در آن‌ها مورد بررسی قرار می‌گیرد و در بخش دوم چگونگی بازیابی اطلاعات وب و ساخت شاخص برای داده‌های ناساختمند بیان می‌شود.						
زمان امتحان: میان‌ترم (۲۷ فروردین ۱۳۹۸ - ساعت ۱۰:۳۰ تا ۱۲:۳۰) و پایان‌ترم (۲۹ خرداد ۱۳۹۸ - ساعت ۱۰:۳۰ تا ۱۲:۳۰)						
نحوه ارزشیابی		امتحان میان‌ترم (M)	امتحان پایان‌ترم (F)	تمرین‌های اصلی (E)	تمرین‌های اختیاری (O)	سوالات امتیازی (Q)
درصد نمره		۹	۹	۲	۱+	۱+
فرمول محاسبه نمره		$G = M + F + E (+O + Q)$				
قوانین درس		۱- تمام تمرین‌های درس به صورت دستی تحویل داده خواهند شد. ۲- برای هر تمرین به ازای هر روز تأخیر ۱۰٪ از نمره اخذ شده‌ی آن کسر خواهد شد. قانون محاسبه‌ی تأخیر تنها تا زمان حل تمرین‌ها در کلاس حل تمرین برقرار خواهد بود و در صورت تحویل بعد از حل تمرین نمره‌ای در نظر گرفته نمی‌شود.				
منابع و مآخذ درس		مرجع اصلی: Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, " An Introduction to Information Retrieval ," Cambridge University Press, 2008. مراجع دیگر: به صفحه Piazza درس مراجعه شود.				
صفحه گروه درس		در Piazza: https://piazza.com/semnan_university/spring2019/cs112/home				

بودجه‌بندی درس

مبحث	شماره هفته آموزشی
معرفی درس، بیان مقدمات، شناخت چند حافظه ذخیره‌سازی و چگونگی عملکرد آن‌ها	۱
چگونگی ذخیره‌سازی داده بر روی رسانه‌های ذخیره‌سازی از دید فیزیکی و منطقی	۲
بررسی پارامترهای مختلف در کارایی سیستم ذخیره‌سازی و بازیابی اطلاعات	۳
چگونگی ذخیره‌سازی داده‌های ساختمند در سیستم‌های مبتنی بر فایل و پایگاه داده‌ها	۴
معرفی ساختار فایل‌های مختلف (ساختار پایل و مستقیم)	۵
ادامه معرفی ساختار فایل‌های مختلف (ساختار شاخص‌دار ترتیبی)	۶
ادامه معرفی ساختار فایل‌های مختلف (ساختار چندشاخصی)	۷
ادامه معرفی ساختار فایل‌های مختلف (ساختار مستقیم و چند حلقه‌ای)	۸
بازیابی بولین و معرفی شاخص وارونه	۹
چگونگی استخراج واژگان از اسناد و تعیین مجموعه واژگان	۱۰
ساخت لیست‌پست‌ها، پست‌های موقعیتی و پرس‌وجوهای اصطلاح	۱۱
جستجو در لغت‌نامه، تکنیک‌های پرس‌وجوهای جایگزین، تصحیح املائی و آوایی	۱۲
ساخت شاخص (شاخص‌گذاری بلوکی مبتنی بر مرتب‌سازی، درون حافظه‌ای تک‌گذره، توزیع‌شده، پویا و...)	۱۳
فشرده‌سازی شاخص (قانون Zipf، Heap، فشرده‌سازی لغت‌نامه، فشرده‌سازی لیست‌پست‌ها و...)	۱۴
وزن‌دهی عبارات و مدل فضای بردار	۱۵
محاسبه نمره‌ها و وزن‌ها در یک سیستم کامل جستجو	۱۶