

بسمه تعالی

فرم شرح درس

نام درس: متن کاوی و وب کاوی تعداد واحد: ۳ نوع واحد: نظری پیش نیاز: --

زمان برگزاری کلاس: روز: سه شنبه و چهارشنبه ساعت: ۱۰ تا ۱۲، ۸ تا ۱۰ مکان برگزاری: کلاس ۱۱ ساختمان جغرافیا

تعداد دانشجویان: ۵

هدف درس: این درس بر آشنایی با روش های استخراج الگو از داده های متنی و صفحات وب تمرکز دارد. برای این منظور، ابتدا پیش زمینه های مورد نیاز از جبری خطی و یادگیری ماشین آماری را معرفی می کند و سپس به معرفی مسائل مطرح در زمینه های متن کاوی و وب کاوی و الگوریتم های مربوط می پردازد تا توانایی لازم برای پژوهش در این حوزه ها کسب شود.

فعالیت های آموزشی:

معرفی درس جدید همراه با طرح سوؤال برای سیال سازی ذهن، پرسش و پاسخ درباره ی مسائل مطرح شده، جمع بندی مطالب ارائه شده در کلاس، ارائه تکالیف و پروژه های برنامه نویسی برای انجام در خارج از کلاس، معرفی مقاله برای مطالعه در خارج کلاس.

منابع اصلی درس:

Sholom M. Weiss, Nitin Indurkha, Tong Zhang, Fundamentals of Predictive Text Mining, Springer, 2010.

Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents and Usage Data, Springer, 2nd ed., 2011.

جدول زمان بندی ارائه مطالب

جلسه مدت زمان (ساعت)	سرفصل درس	شیوه تدریس	نحوه ارزیابی	منابع مکمل درس
جلسه های هفته اول، ۲ ساعت و ۵۰ دقیقه	معرفی متن کاوی و وب کاوی و اهداف درس	سخنرانی	پرسش و پاسخ در کلاس	در اختیار گذاشتن مقالاتی در زمینه مفاهیم درس و جنبه های جدید پژوهشی مرتبط با درس
جلسه های هفته دوم، ۲ ساعت و ۵۰ دقیقه	مدل سازی فضای برداری (VSM)، توابع شباهت: همبستگی، شباهت کسینوسی، فایل معکوس و فشرده سازی آن،	سخنرانی	پرسش و پاسخ در کلاس	
جلسه های هفته سوم، ۲ ساعت و ۵۰ دقیقه	تجزیه SVD و روش های شاخص گذاری معنایی LSA و pLSA	سخنرانی	پروژه ۱ برای پیاپی سازی الگوریتم pLSA	در اختیار گذاشتن مقالات جدید در زمینه شاخص گذاری معنایی برای مطالعه
جلسه های هفته چهارم، ۲ ساعت و ۵۰ دقیقه	مدل سازی موضوعی (Topic Modeling) و تخصیص پنهان دیریکله (LDA) و گونه برخط آن (OLDA)	سخنرانی	پرسش و پاسخ در کلاس	

بسمه تعالی

فرم شرح درس

ادامه جدول زمانبندی ارائه مطالب

جلسه مدت زمان (ساعت)	سرفصل درس	شیوه تدریس	نحوه ارزیابی	منابع مکمل درس
جلسه‌های هفته پنجم، ۲ ساعت و ۵۰ دقیقه	روش‌های یادگیری نیمه نظارت شده در دسته‌بندی اسناد متنی: روش LU، روش هم-آموزی (Co-training)	سخنرانی	پروژه ۲ برای پیاده‌سازی الگوریتم‌های LU و هم-آموزی	
جلسه‌های هفته ششم، ۲ ساعت و ۵۰ دقیقه	ادامه روش‌های یادگیری نیمه نظارت شده در دسته‌بندی اسناد متنی: روش Transductive، روش PU	سخنرانی	پرسش و پاسخ در کلاس	
جلسه‌های هفته هفتم، ۲ ساعت و ۵۰ دقیقه	نظرکاوی و تحلیل احساسات (sentiment)، هرزه‌نگاری در وب	سخنرانی	پروژه ۳ برای پیاده‌سازی الگوریتم‌های Transductive PU و	در اختیار گذاشتن مقالات جدید در زمینه یادگیری نیمه‌نظارت شده تصمیم برای مطالعه
جلسه‌های هفته هشتم، ۲ ساعت و ۵۰ دقیقه	رتبه‌دهی	سخنرانی	پرسش و پاسخ در کلاس	
جلسه‌های هفته نهم، ۲ ساعت و ۵۰ دقیقه	فرایند مارکف، معرفی الگوریتم HITS	سخنرانی	پرسش و پاسخ در کلاس	
جلسه‌های هفته دهم، ۲ ساعت و ۵۰ دقیقه	معرفی الگوریتم PageRank، دسته‌بندی صفحات وب به کمک پیوندها، هرزه‌نگاری پیوند،	سخنرانی	پرسش و پاسخ در کلاس	در اختیار گذاشتن مقالات جدید در زمینه جستجوی وب برای مطالعه
جلسه‌های هفته یازدهم، ۲ ساعت و ۵۰ دقیقه	معرفی شبکه‌های اجتماعی، اثر دنیای کوچک، مدل‌سازی شبکه،	سخنرانی	پروژه ۴ برای پیاده‌سازی الگوریتم‌های HITS و PageRank	

جلسه‌های هفته دوازدهم، ۲ ساعت و ۵۰ دقیقه	معرفی معیارهای: گره‌های مرکزی، درجه، نزدیکی، بینابینی، پیوند قوی، مدل‌سازی تاثیر در شبکه‌های اجتماعی و بیشینه‌سازی آن،	سخنرانی	پرسش و پاسخ در کلاس	
جلسه‌های هفته سیزدهم، ۲ ساعت و ۵۰ دقیقه	معرفی جوامع و روش‌های آشکارسازی جوامع، آشکارسازی جوامع در شبکه‌های ناهمگون	سخنرانی	پرسش و پاسخ در کلاس	
جلسه‌های هفته چهاردهم، ۲ ساعت و ۵۰ دقیقه	الگوریتم‌های خوشه‌بندی و دسته‌بندی در شبکه‌های اجتماعی	سخنرانی	پرسش و پاسخ در کلاس	در اختیار گذاشتن مقالات جدید در زمینه شبکه‌های اجتماعی برای مطالعه
جلسه‌های هفته پانزدهم، ۲ ساعت و ۵۰ دقیقه	سیستم‌های پیشنهادده: روش پالایش همکارانه، روش محتوا پایه،	سخنرانی	پروژه ۵ برای پیاپی‌سازی روش پالایش همکارانه	
جلسه‌های هفته شانزدهم، ۲ ساعت و ۵۰ دقیقه	ادامه سیستم‌های پیشنهادده: روش دانش پایه، روش ترکیبی، جمع‌بندی درس	سخنرانی	پرسش و پاسخ در کلاس	