



سید حسین خواسته

دانشگاه صنعتی خواجه نصیرالدین طوسی

سید حسین خواسته استادیار دانشگاه صنعتی خواجه نصیرالدین طوسی است، وی مدارک کارشناسی مهندسی برق، کارشناسی ارشد و دکترا مهندسی کامپیوتر را به ترتیب در سالهای 1383، 1385 و 1391 از دانشگاه صنعتی شریف اخذ نموده است. ایشان مجری پروژه های متعددی در حوزه های هوش مصنوعی، تحلیل شبکه های اجتماعی، مدیریت شبکه های مخابراتی و پروژه های ترکیبی سخت افزار/ نرم افزار بوده است. دکتر خواسته از سال 1394 درس تحلیلها و سیستم های دادگان حجیم (کلان داده) را در مقطع تحصیلات تکمیلی دانشگاه خواجه نصیر تدریس می نماید. ایشان عضو هیات مدیره انجمن کامپیوتر ایران، بنیاد ملی نخبگان، کارگروه تشخیص مصادیق شرکتهای دانش بنیان و کارگروه نظارت بر نحوه هزینه کرد وامهای صندوق نوآوری و شکوفایی هستند. ایشان نویسنده (و همکار) بیش از 50 مقاله پژوهشی، مروری و آموزشی در مجلات و کنفرانسهای معتبر بوده اند.

ایشان به عنوان دبیر کمیته دانشجویی کنفرانس مهندسی برق 1396، دبیر کمیته انتشارات کنفرانس مهندسی کامپیوتر 1397 و به عنوان داور کنفرانسها و مجلات متعدد داخلی و خارجی خدمت کرده است.

کلان داده، مفاهیم و کاربردها

کلان داده چیست و چه مشخصاتی دارد؟ شاید یک پاسخ این باشد: کلان داده اصطلاحی است که برای توصیف حجم زیادی از داده های پیچیده و متغیر که با سرعت بالایی تولید می شوند، به کار می رود. برای جمع آوری، ذخیره سازی، توزیع، تجزیه و تحلیل و همچنین مدیریت چنین داده هایی، روشها و فناوری های پیشرفته ای مورد نیاز هستند.

کلان داده یک واقعیت در جهان کنونی است و مسئله‌ای است که سیستم‌های واقعی باید با آن دست‌وپنجه نرم کنند. جوامع اطلاعاتی نوین به واسطه‌ی مخازن گسترده‌ای از داده تعریف می‌شوند؛ این داده‌ها می‌توانند عمومی یا خصوصی باشند. بنابراین، با افزایش مجموعه داده‌های مورد علاقه، تمامی برنامه‌های کاربردی عملی باید قادر به افزایش مقیاس باشند. در این سخنرانی مفاهیم اصلی مطرح در زمینه کلان داده مانند: مشخصه‌های 3V و 4V و 5V در مورد کلان داده، چارچوب‌های اصلی مدیریت و کار با کلان داده مانند: Hadoop، اکوسیستم هادوپ و نرم افزارهای موجود در این اکوسیستم، چارچوب map-reduce، روش‌های نوین پردازش داده‌های جریان و ابزارهای Spark و Storm و برخی از الگوریتم‌های نوین مورد استفاده برای پردازش جریان داده‌ها و نکات مهم این الگوریتم‌ها مورد بحث و بررسی قرار می‌گیرند.

همچنین چند مثال عملیاتی و کاربردی از سیستم‌های عملیاتی مبتنی بر کلان داده‌ها، نحوه فیلتر کردن ایمیل‌های spam، تحلیل داده‌های فروش فروشگاه‌های بزرگ (مانند Amazon) و تحلیل اطلاعات کانال‌های فارسی تلگرام، مورد بررسی قرار می‌گیرند و برخی از الگوریتم‌ها و تکنولوژی‌های به کار رفته به این منظور بررسی می‌شوند.