

عنوان مقاله:

دسته بندی متون انگلیسی با استفاده از الگوریتم های درخت تصمیم، بگینگ

محل انتشار:

نهمین همایش ملی علوم و مهندسی دفاعی با محوریت فناوری های دانش بنیان دفاعی (سال: 1401)

تعداد صفحات اصل مقاله: 13

نویسندگان:

محسن نوروزی - دکتری کامپیوتر، دانشکده کامپیوتر، شبکه و ارتباطات، دانشگاه امام حسین (ع)، تهران، ایران

علی ارشقی - دکتری برق مخابرات، دانشکده کامپیوتر، شبکه و ارتباطات، دانشگاه امام حسین (ع)، تهران، ایران

جواد اسدی - دانشجوی دکتری سیستم های نرم افزاری، دانشکده کامپیوتر، دانشگاه آزاد اسلامی، قزوین، ایران

خلاصه مقاله:

با توجه به گسترش و توسعه سریع سایت های تجاری و رسانه های اجتماعی حجم زیادی از داده ها تولید می-شود. این حجم زیاد داده های تولید شده نیاز به سازماندهی مناسب دارد. به همین دلیل فرآیند متن کاوی برای استخراج ویژگیهای مرتبط و دانش از اسناد غیر ساخت یافته مورد استفاده قرار می گیرد. متن کاوی درواقع استخراج اطلاعات و دانش مفید از متون غیرساخت یافته است. حجم زیادی از اسناد و منابع در پایگاههای اطلاعاتی الکترونیکی قرار دارند. برخی از این پایگاه های اطلاعاتی شامل اطلاعات با ارزش هستند که دانشانوهی را در خود پنهان کرده اند. روش طبقه بندی اسناد در متن کاوی برای استخراج ویژگی های مرتبط مورد استفاده قرار می گیرد. در این روش با استفاده از داده های آموزشی مدلی ایجاد می شود و سپس داده های جدید به مدل ایجاد شده داده می شود و برچسب آنها پیش بینی می شود. الگوریتم های طبقه بندی زیادی برای زبانانگلیسی و زبان های دیگر وجود دارند. هر کدام از این الگوریتم ها برای بهبود شرایط طبقه بندی با گذشت زمان دست خوش تغییرات و تکاملاتی شده اند. در این مقاله با ترکیب این الگوریتم ها سعی در بهبود طبقه بندی خواهیم داشت و نتایج حاصل را ارائه خواهیم کرد.

کلمات کلیدی:

متن کاوی، طبقه بندی، k نزدیکترین همسایه، درخت تصمیم، بگینگ

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1753334>

