

عنوان مقاله:

استفاده از تکنیک های داده کاوی در دسته بندی خودکار اسناد متنی

محل انتشار:

اولین همایش ملی مهندسی کامپیوتر و فناوری اطلاعات دانشگاه پیام نور (سال: 1393)

تعداد صفحات اصل مقاله: 8

نویسندگان:

محمد بهروزیان نژاد - دانشجوی دکترای کامپیوتر، دانشگاه آزاد اسلامی واحد شوشتر، آموزشکده فنی و حرفه ای سما

ایمان عطارزاده - استادیار رشته کامپیوتر، هیئت علمی، دانشگاه آزاد اسلامی واحد دزفول، گروه کامپیوتر

شادی افتخار - دانشجوی کارشناسی کامپیوتر، دانشگاه آزاد اسلامی واحد شوشتر، گروه کامپیوتر

احمد کاظمی - هیئت علمی، دانشگاه آزاد اسلامی واحد ایذه، گروه کامپیوتر

خلاصه مقاله:

امروزه بخش قابل توجهی از اطلاعات موجود در پایگاه داده های متنی یا اسناد متنی ذخیره می شوند. یکی از مهم ترین مباحثی که مطرح است بحث سازماندهی این اسناد می باشد. یکی از راهکارهای سازماندهی اسناد متنی، دسته بندی آنها می باشد. دسته بندی متون به انتساب اسناد متنی به دسته های واقعی آنها می باشد. دسته بندی اسناد متنی شامل دو مرحله اصلی انتخاب ویژگی و الگوریتم یادگیری می باشد. تاکنون روش های متعددی برای دسته بندی اسناد متنی ارائه شده است. در این مقاله یک روش ترکیبی جهت دسته بندی کارآتر اسناد متنی ارائه شده است. روش پیشنهادی در مرحله انتخاب ویژگی جهت کاهش پیچیدگی از روش فیلتری بهره می برد و با استفاده از دسته بندهای بیز ساده و درخت تصمیم پیاده سازی شده است. نتایج نشان دهنده برتری روش ترکیبی نسبت به دسته بندهای منفرد می باشد.

کلمات کلیدی:

داده کاوی، اسناد متنی، دسته بندی، دسته بندی خودکار متون

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/337433>

