

عنوان مقاله:

ارائه یک روش ترکیبی جهت بهبود افزایش دقت دسته بندی اسناد علمی با استفاده از الگوریتم KNN و تکنیک شاخص گذاری معنایی پنهان

محل انتشار:

اولین کنفرانس ملی ایده های نو در مهندسی کامپیوتر (سال: 1394)

تعداد صفحات اصل مقاله: 6

نویسندگان:

محمد کافیان - دانشجوی کارشناسی ارشد، دانشگاه آزاد اسلامی واحد نجف آباد

مهدی باطنی - استادیار گروه مهندسی کامپیوتر، دانشگاه شیخ بهایی اصفهان

خلاصه مقاله:

در سال های اخیر الگوریتم های زیادی برای مساله دسته بندی اسناد، پیشنهاد شده است که از استراتژی های گوناگونی که اکثراً بر مبنای روش های بانظر می باشند، الگو گرفته اند. در واقع روش های زیادی جهت دسته بندی کردن اسناد ارائه شده اند که بدون توجه به مفاهیم نهفته درون متن، عملیات دسته بندی را انجام می دهند. تکنیک پیشنهادی ما در این مقاله این است ابتدا کلمات مهم و پرتکرار از متن اسناد علمی با استفاده از تکنیک TF-IDF بدست آمده و سپس با مقایسه با مفاهیم و کلمات کلیدی درخت سلسله مراتبی از رشته کامپیوتر، بتوان حوزه های کاری اسناد را بدست آورد و در نهایت با استفاده از یکی از تکنیک های پر کاربرد به نام LSI، از بین حوزه های تشخیص داده شده، بتوان دسته بندی را به طور تخصیص برای هر کدام از اسناد علمی انجام داد. در نهایت با استفاده از الگوریتم KNN، کار دسته بندی را انجام داد.

کلمات کلیدی:

دسته بندی متون، متن کاوی، پردازش زبان طبیعی، الگوریتم KNN

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/418133>

