

عنوان مقاله:

ایجاد یک طبقه بند خودکار متون زبان فارسی به منظور استفاده در کتابخانه های دیجیتال شهر الکترونیک

محل انتشار:

همایش ملی شهر الکترونیک (سال: 1390)

تعداد صفحات اصل مقاله: 5

نویسندگان:

سامان باشباغی - دانشجوی کارشناسی ارشد هوش مصنوعی دانشگاه بوعلی سینا

مصطفی پرچی - دانشجوی کارشناسی ارشد هوش مصنوعی دانشگاه بوعلی سینا

حسن ختنلو - استادیار دانشگاه، عضو هیئت علمی دانشگاه بوعلی سینا

حسن بشیری - عضو هیئت علمی دانشگاه صنعتی همدان

خلاصه مقاله:

دسته بندی یا طبقه بندی متون، برچسب زدن متون زبان طبیعی با توجه به طبقه بندی های موضوعی حاصل از مجموعه های از قبل مشخص شده می باشد. دسته بندی خودکار متون و انتساب اسناد به دسته های از پیش تعیین شده، در ده سال اخیر توجهات زیادی را به سوی خود جلب کرده است. این مسأله به خاطر رشد سریع متون الکترونیکی و مسأله نیاز به دسترسی اسناد، مدارک و مقالات علمی، صفحات وب، نامه های الکترونیکی و کتابخانه های دیجیتال بوده است. در این تحقیق برای پیاده سازی دسته بندی خودکار برای متون فارسی، از الگوریتم K-NN که از روشهای یادگیری با ناظر است به منظور دسته بندی متون و اسناد شهر الکترونیک استفاده شده است. به منظور ارزیابی نتایج حاصل از دسته بندی، مجموعه اسناد فارسی پیکره همشهری به عنوان مجموعه های آموزشی و آزمایشی بکار رفته است. نتایج آزمایشات که با استفاده از توابع مشابهت مختلف به دست آمده است، نشان از دقت بسیار بالای روش دارد.

کلمات کلیدی:

دسته بندی متن، شاخص گذاری متن، شهر الکترونیک و الگوریتم K نزدیکترین همسایه

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/152722>

