

عنوان مقاله:

رویکردی نوین برای دسته بندی اسناد متنی با استفاده از الگوریتم کرم شب تاب

محل انتشار:

دومین کنفرانس بین المللی پژوهش های دانش بنیان در مهندسی کامپیوتر و فناوری اطلاعات (سال: 1396)

تعداد صفحات اصل مقاله: 14

نویسندگان:

ندا یزدان خواه - گروه مهندسی کامپیوتر، واحد مراغه، دانشگاه آزاد اسلامی، مراغه، ایران

فرهاد سلیمانیان قره چیق - گروه مهندسی کامپیوتر، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران

خلاصه مقاله:

دسته بندی متون به دلیل افزایش حجم اطلاعات موجود در اینترنت و همچنین حجم زیاد متون و تولید روزانه هزاران متن متنوع علی الخصوص در فضای مجازی، نقش مهمی را در مدیریت اطلاعات علی الخصوص دسته بندی متون ایفا می کند. برای دسته بندی متون نیازمند شناسایی ویژگی های متون بوده که برای یک متن می توان ویژگیهای زیادی را شناسایی نمود که این یکی از مشکلات پیش رو در این زمینه است. و اغلب ویژگی ها نامربوط و زاید هستند و بر کارایی دسته بندی کننده تاثیر منفی می گذارند. انتخاب ویژگی در حقیقت مهمترین مرحله در دسته بندی متون می باشد که هدف از انتخاب ویژگی، انتخاب بهینه ترین زیر مجموعه ویژگی از کل فضای ویژگی های اصلی مسئله مورد نظر است. بطوریکه ضمن کاهش ابعاد می توان به دقت دسته بندی مطلوب دست یافت. از مزایای انتخاب ویژگی می توان به تسهیل درک و مدل سازی، کاهش نیازهای اندازه گیری، کاهش هزینه های محاسباتی و کاهش پیچیدگی زمانی اشاره کرد. ما در این مقاله، از روش نوینی مبتنی بر الگوریتم کرم شب تاب برای طبقه بندی اسناد متنی استفاده نمودیم. در روش پیشنهادی اسناد متنی در ابتدا پیش پردازش می شوند و کلمات کلیدی موجود در اسناد استخراج می شوند. سپس برای هر یک از کلمات کلیدی، وزن مشخصی بر مبنای تکرار معین می شوند. ارزیابی بر روی مجموعه داده استاندارد Reuters-2557 انجام گرفته است که نتایج حاصل نشان می دهد که مدل پیشنهادی در مقایسه با روش های موجود مانند روش K نزدیک ترین همسایه، KNN-K-Means و سایر روش های موجود برای دسته بندی به نسبت از دقت بیشتری برخوردار می باشد و عملکرد خوبی داشته است.

کلمات کلیدی:

طبقه بندی اسناد متنی، الگوریتم کرم شب تاب، انتخاب ویژگی، بهینه سازی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/696276>

