

عنوان مقاله:

ارائه یک الگوریتم خوشه بندی تجمعی پایدار جهت بهبود دقت و صحت دسته بندی اسناد متنی

محل انتشار:

کنگره مشترک سیستم های فازی و هوشمند ایران (نوزدهمین کنفرانس سیستم های فازی و هفدهمین کنفرانس سیستم های هوشمند)
(سال: 1399)

تعداد صفحات اصل مقاله: 8

نویسندگان:

زهرا جعفری - واحد توسعه تحقیقات بالینی، بیمارستان قائم، دانشگاه علوم پزشکی مشهد، مشهد، ایران

رضا قائمی - گروه مهندسی کامپیوتر، واحد قوچان، دانشگاه آزاد اسلامی، قوچان، ایران

خلاصه مقاله:

امروزه میزان اطلاعات و مستندات متنی، روز به روز در حال گسترش است. فراهم کردن ابزاری که بتواند به طور موثر و کارا، این اطلاعات گسترده و دانش پنهان درون آن را شناسایی، استخراج و مدیریت کند، امری مهم و ضروری است. یکی از روش های مهم در راستای برآورده کردن این نیاز کاربران، استفاده از تکنیک های متنکاوی همانند روش های دسته بندی خودکار متون است. از جمله مشکلات دسته بندی متون می توان به ابعاد بالای فضای ویژگی، عدم پایداری و کارایی قابل قبول در دسته بندی اشاره نمود. در این مقاله، یک روش سه مرحله ای برای دسته بندی متون پیشنهاد شده است که در آن ابتدا کلمات کلیدی پس از پیش پردازش اسناد، با استفاده از روش (فراوانی - عکس فراوانی سند) استخراج می شوند. درگام بعدی، جهت حل مشکل بالابودن ابعاد فضای ویژگی، از الگوریتم کاهش ویژگی تحلیل مولفه اصلی (PCA) استفاده می شود و در گام آخر، اسناد با استفاده از تکنیک خوشه بندی تجمعی و براساس چهارخوشه بند (Fuzzy C-means + PSO، Complete-linkage، K-) و WOA (Fuzzy C-means+ WOA) دسته بندی می شوند. نتایج به دست آمده از ارزیابی این سیستم بر روی سه پایگاه داده متنی استاندارد Reuters-۲۱۵۷۸، RCV و Classic نشان دهنده این است که روش پیشنهادی، از دقت، پایداری و قابلیت اعتماد نسبتا بالایی برخوردار است.

کلمات کلیدی:

متن کاوی، دسته بندی اسناد، خوشه بندی تجمعی، استخراج کلمات کلیدی، تحلیل مولفه اصلی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1193469>

