

عنوان مقاله:

دستهبندی اسناد متنی حجیم با استفاده از ترکیب درهمساز کمینه و نزدیکترین همسایه پیشنهادی

محل انتشار:

دومین کنفرانس ملی مهندسی نرم افزار دانشگاه آزاد لاهیجان (سال: 1391)

تعداد صفحات اصل مقاله: 8

نویسندگان:

فاطمه شیخی - گروه مهندسی نرم افزار، دانشگاه آزاد اسلامی واحد شبستر، ایران

حسن نادری - استادیار دانشگاه علم و صنعت، تهران، ایران

خلاصه مقاله:

در دهی گذشته پیشرفتهای زیادی در زمینهی دستهبندی اسناد متنی انجام گرفته و الگوریتمهای زیادی در این زمینه پیشنهاد شده است. در این مقاله به دستهبندی اسناد متنی حجیم پرداخته شده است. روش پیشنهادی به کار بردن درهمساز کمینه 1 به همراه دستهبندی نزدیکترین همسایه 2 میباشد. تابع درهمساز کمینه معمولا برای سرعت بخشیدن در جستجوی جداول، فشردهسازی دادهها و تشخیص رکوردهای تکراری در حجم زیاد داده استفاده میشود [3]؛ اما در این مقاله برای خلاصه کردن اسناد متنی به کار گرفته شده است. استفاده از درهمساز کمینه به همراه یک دستهبند به صورت طبیعی با افت دقت همراه است، اما در عین حال باعث بالا رفتن سرعت عملیات دستهبندی میشود. در این مقاله برای حل مشکل افت دقت حاصل از درهمساز کمینه، یک دستهبندی نزدیکترین همسایه پیشنهاد شده، که در آن به جای در نظر گرفتن شباهت کسینوسی بین اسناد از شباهت جاکارد استفاده شده است. این دستهبند بر روی 500 سند انتخابی از بین 2000 سند موجود در مجموعه داده انتقادات راجع بهفيلم 3، پیادهسازی شده که 400 سند برای آموزش و 100 سند برای آزمایش در نظر گرفته شده است

کلمات کلیدی:

دستهبندی اسناد متنی، درهمساز کمینهk نزدیکترین همسایه و تشابه جاکارد

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/184792>

