

عنوان مقاله:

تشخیص سرقت علمی در متون بازنویسی شده براساس شاخص گذاری با استفاده از شبکه واژگان فارسی

محل انتشار:

چهارمین کنفرانس بین المللی تحقیقات بین رشته ای در مهندسی برق، کامپیوتر، مکانیک و مکترونیک در ایران و جهان اسلام (سال: 1399)

تعداد صفحات اصل مقاله: 13

نویسندگان:

میلاذ سرانجام - دانشکده مهندسی کامپیوتر و فناوری اطلاعات، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران

وحید رستمی - دانشکده مهندسی کامپیوتر و فناوری اطلاعات، واحد قزوین، دانشگاه آزاد اسلامی، قزوین، ایران

خلاصه مقاله:

گسترش روز افزون اینترنت، دسترسی به ایده ها، مقالات، مستندات و دست نوشته های دیگران آسان تر گردیده این امر موجب تبادل سریع اطلاعات و در مقابل، سبب آسان تر شدن سرقت ادبی یا Plagiarism شده است. عموماً روش های اندازه گیری تشابه متن مبتنی بر بردار تکرار واژه، روابط معنایی بین کلمات را نادیده می گیرند، که در واقع مانعی برای محاسبه شباهت متن، در ابعاد بالا و پراکندگی بیشتر بردار سند می باشد. برای غلبه بر این مشکلات، از تجزیه مقادیر منفرد (SVD) برای کاهش ابعاد و حذف نویزهای مدل ارائه متن استفاده می شود. مقدار بهینه مقادیر منفرد محاسبه می شود و ارتباط معنایی بین کلمات را می توان در فضای معنایی ساخته شده محاسبه کرد. در این پژوهش یک الگوریتم ساخت شاخص معکوس (TF-IDF) و تعاریف مشابه بینبردارها برای محاسبه شباهت بین دو سند در سطح معنایی پیشنهاد شده است. جهت ارزیابی کارایی الگوریتم، از مجموعه داده مسابقات plagdet2016 مرکز تحقیقات مخابرات ایران استفاده شده است. در ادامه، نتایج تجربیدر مورد مجموعه ی متنی مورد ارزیابی شده نشان می دهد که الگوریتم پیشنهادی معیارهای ارزیابی کارایی مانند را افزایش می دهد.

کلمات کلیدی:

سرقت متن، ارتباط معنایی، تجزیه مقادیر منفرد، اندازه گیری تشابه متن، LSA

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1184161>

