

عنوان مقاله:

پیاده سازی و بررسی نقش کلمات کم تکرار و موازی سازی در کشف اسناد تکراری

محل انتشار:

کنفرانس بین المللی علوم و مهندسی (سال: 1394)

تعداد صفحات اصل مقاله: 7

نویسندگان:

جواد گلی - دانشجوی کارشناسی ارشد

حسن نادری - استاد راهنما

خلاصه مقاله:

با توجه به گستردگی بحث اسناد تکراری تعاریف مختلف بسیاری وجود دارد که دیگران استفاده کرده اند، و هر یک بسته به مورد استفاده خاص بوده و دارای انگیزه های متفاوتی هستند. به طور کلی، اسناد مشابه اسنادی هستند که فقط کمی در محتوا متفاوت می باشند. با توجه به این تعریف جامع می توان انواع تعاریف را برای اسناد مشابه در نظر گرفت. بعنوان نمونه سند هایی که فقط در ساختار ظاهری با هم تفاوت دارند و در محتوای اصلیشان یکسان می باشند. یا در حالت دیگر، اسنادی که به هم ریخته شده اند و اصطلاحاً با ویرایش متون در نگاه اول سند تکراری به نظر نمی آید. یا در حالت دیگر، گاهی اوقات با تغییر عنوان های یک متن از نسخه اول کپی برداری صورت میگیرد. یا در حالت دیگر جزئیات کم اهمیت سند حذف شده ولی محتوای اصلی سند همچنان وجود دارد. بخش عمده ای از کلمات در هر سند کلمات اضافی و اغلب کم معنی و کم ارتباط با محتوای سند هستند از این رو روش پیشنهادی ما عملکرد بهتری در بحث حافظه و سرعت خواهد داشت. عملیاتی که پس از خواندن اطلاعات صورت می گیرد شامل شمارش کل کلمات یک سند، تعداد کلمات کم تکرار آن و همچنین تعداد لغات منحصر به فرد یا همان یکتا می باشد سپس عملیات مقایسه بصورت موازی صورت می گیرد

کلمات کلیدی:

اسناد مشابه، سرقت ادبی، نسخه های تکراری، عملیات موازی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/424446>

