

عنوان مقاله:

بهبود بازیابی اطلاعات بر اساس تشابه معنایی کلمات کلیدی با استفاده از رتبه دهی مبتنی بر گراف

محل انتشار:

چهارمین کنفرانس ملی فناوری اطلاعات، کامپیوتر و مخابرات (سال: 1396)

تعداد صفحات اصل مقاله: 11

نویسندگان:

سیدمحمد جوادی مقدم - عضو هیئت علمی، گروه کامپیوتر، دانشگاه بزرگمهر، قاینات

مجید عبدالرزاق نژاد - عضو هیئت علمی، گروه کامپیوتر، دانشگاه بزرگمهر، قاینات

مهناز قادری فریز - دانشکده مهندسی کامپیوتر، گروه نرم افزار، دانشگاه آزاد اسلامی، بیرجند

خلاصه مقاله:

کلمات کلیدی در اسناد متنی، کلماتی از متن اسناد هستند که بیشترین بار مفهومی متن را به همراه داشته و نیز یک نسخه فشرده متن محسوب می شود در نتیجه نیاز به روش های استخراج خودکار کلمات کلیدی را به شدت افزایش داده اخیرا روش های رتبه بندی مبتنی بر گراف کاربرد موفق در حوزه وب داشته یک مشکل عمده اکثر این روش ها تاکید بیش از حد بر پارامتر هم جواری کلمات در ایجاد و وزندهی یال های گراف متنی و صرف نظر از شاخص های آماری شده است. در این پژوهش برانیم شباهت معنایی کلمات کلیدی را به صورت فرمت پیچیده تری از متغیر TF-IDF (روش وزندهی کلاسیک) به عنوان شاخص آماری بیان کنیم. با تعریف متغیر جدید که بیانگر ترتیب کاهنده از احتمال ارتباطشان با پرس وجوی کاربر است و یک روش مشخص به عنوان رتبه بندی احتمال؛ الگوریتم معروف BM25 است، در این پژوهش اطلاعات آماری روش رتبه بندی احتمال ارتباط کلمات کلیدی، از جمله تعداد اسناد مشابه و اسناد کل مجموعه در وزندهی گراف استفاده شده است. هدف این مقاله این است که شباهت معنایی اسناد مختلف با سند مورد نظر بررسی کنیم با رتبه بندی کلمات کلیدی مجموعه اسناد مرجع، اسنادی که دارای کلمات کلیدی با بالاترین اولویت اند، شبیه ترین اسناد به سند مورد بررسی است. مقایسه نتایج روش جدید با روش های قبلی افزایش دقت 93% در اسناد استخراج شده مشابه سند مورد بررسی را نشان می دهد.

کلمات کلیدی:

اطلاعات آماری، رتبه دهی مبتنی بر گراف، کلمه کلیدی، کلمات کلیدی استخراجی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/668872>

