

## عنوان مقاله:

استفاده از الگوریتم رتبه بندی BM25 و بهبود آن در متون غیرساخت یافته با استفاده از تطبیق مفاهیم

## محل انتشار:

کنفرانس بین المللی یافته های نوین پژوهشی در مهندسی برق و علوم کامپیوتر (سال: 1394)

تعداد صفحات اصل مقاله: 8

## نویسندگان:

مسعود عگبی - گروه کامپیوتر، واحد اهواز، دانشگاه آزاد اسلامی اهواز، ایران

منصور امینی لاری - گروه کامپیوتر، واحد مرودشت، دانشگاه آزاد اسلامی مرودشت، ایران

## خلاصه مقاله:

یکی از مهمترین موضوعات در پردازش زبان طبیعی و بازیابی اطلاعات، یافتن ریشه کلمات است. یکی از روشهای افزایش کارایی سیستمهای بازیابی اطلاعات، استفاده از ریشهیابی کلمات است. زیرا اشتقاقیات مختلف یک کلمه به ریشه آنکلمه تبدیل میشوند. در نتیجه جستجو بر اساس ریشه کلمه انجام خواهد شد و اندازه ساختار ایندکس کاهش مییابد. حال با توجه به پیچیدگی علوم در صورتیکه در حجم عظیمی از دادهها تمایل به کارهای تحقیقاتی به صورت دستی داشته باشیم این کار زمانگیر و گاهی غیر ممکن میباشد. لذا استفاده از روشی که به صورت اتوماتیک این کارهای تحقیقاتی را انجام دهد میتواند مفید باشد. در این مقاله الگوریتمی برای رتبه بندی BM25 در متون غیرساخت یافته با استفاده از تطبیق مفاهیم در زبان فارسی ارائه شده است و سپس نتیجه آن در بازیابی اطلاعات با الگوریتمهای BM25 استاندارد، مورد ارزیابی قرار گرفته است. استفاده از این الگوریتم در متون فارسی و همچنین متونی که شامل ترکیبی از زبانهای فارسی و لاتین باشد گاهی به دلیل برخی مسائل باعث کاهش دقت، صحت و بازخوانی این الگوریتم میگردد. الگوریتم ارائه شده با استفاده از قواعد ساختواری زبان فارسی و استفاده از مجموعه لغات برای جلوگیری از ایجاد ریشههای نادرست، بهریشهیابی کلمات میپردازد

## کلمات کلیدی:

رتبه بندی، داده کاوی، متن کاوی، تطبیق مفاهیم

## لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/404867>

