

عنوان مقاله:

استخراج کلمات کلیدی بصورت مستقل از زبان

محل انتشار:

سومین کنگره بین المللی کامپیوتر، برق و مخابرات (سال: 1395)

تعداد صفحات اصل مقاله: 12

نویسندگان:

بهاره هاشم زاده - مربی گروه کامپیوتر و فناوری اطلاعات، دانشگاه تربت حیدریه

علی ماروسی - استادیار گروه کامپیوتر و فناوری اطلاعات، دانشگاه تربت حیدریه

خلاصه مقاله:

بحث اطلاعات، بازیابی و مدیریت آن از منظر اقتصادی، اجتماعی، سیاسی و امنیتی اهمیت بسیار زیادی برای کشورهايمختلف و کمپانی های بزرگ حوزه اطلاعات و فن آوری اطلاعات دارند. اولین گام در این حوزه، شناسایی و استخراج کلمات کلیدی از متون می باشد. یکی از چالش های عمده بر سر راه این امر، وجود زبان های بسیار متنوع برای اطلاعات متنی و وابستگی روش های موجود استخراج کلمات کلیدی به نوع زبان و ساختار کلامی آن زبان خاص می باشد. لذا هدف این پژوهش، طراحی الگوریتمی مستقل از زبان به منظور استخراج کلمات کلیدی می باشد. از این رو با تمرکز روی خاصیت تکرار کلمات کلیدی در هر متن و نسبت تکرار آن در سایر متون و با کمک از الگوریتم TF_IDF این امر انجام شده است و در نهایت میانگین بیشترین تکرار به عنوان کلمه کلیدی انتخاب می گردد. عملکرد الگوریتم پیشنهادی نیز توسط معیار نرخ دقت تشخیص مورد ارزیابی قرار گرفته است. اگرچه کارهای مشابه در این زمینه به صورت مستقل از زبان وجود نداشته ولی مقایسه عملکرد با کارهای نزدیک در حوزه استخراج کلمات کلیدی وابسته به زبان، حکایت از عملکرد قابل قبول الگوریتم پیشنهادی را دارد. لازم به ذکر است که این الگوریتم با الگوریتم مبتنی بر گراف که فقط برای زبان انگلیسی پیاده سازی شده مقایسه گردیده است، که نرخ دقت کلی الگوریتم پیشنهادی 83.34 درصد بوده است.

کلمات کلیدی:

متن کاوی، بازیابی اطلاعات، استخراج کلمه کلیدی، مستقل از زبان، کلمه کلیدی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/576190>

