

عنوان مقاله:

استخراج کلمات کلیدی اسناد فارسی

محل انتشار:

سیزدهمین کنفرانس سالانه انجمن کامپیوتر ایران (سال: 1386)

تعداد صفحات اصل مقاله: 6

نویسندگان:

مسلم محمدی جنقرا - عضو هیئت علمی دانشگاه آزاد اسلامی واحد ملکان

مرتضی آنالویی - استادیار دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران

خلاصه مقاله:

این مقاله، یک روش آماری ترکیبی، برای استخراج کلمات کلیدی ۱ اسناد فارسی، پیشنهاد کرده است. روش پیشنهادی مبتنی بر پیکره متنی ۲ میباشد. ابتدا عمل ریشه یابی و حذف کلمات عم می ۳ انجام میگردد. سپس ویژگی های آماری برای کلمات مختلف محاسبه شده و با استفاده از فازی سازی و اعمال قواعد فازی، کلمات کلیدی محتمل، انتخاب می شوند. گام بعدی محاسبه رخداد ه مزمان ۴ پیشین و پسین کلمات کلیدی محتمل، با کلمات تکرار شونده، ۵ در جملات سند است. با اعمال یک آستانه وفقی ۶ روی رخداد همزمان کلمات، کلمات کلیدی دو کلمه ای را مشخص می- کنیم. بر خلاف اکثر روش های آماری که فقط کلمات کلیدی یک کلمه ای را استخراج می کنند، استفاده از این روش کلمات کلیدی دو کلمه ای نیز استخراج می شوند. استفاده از ترکیب روش فازی و رخداد همزمان کلمات بهبود خوبی را نشان می دهد و کلمات کلیدی بامعنیتتری پیشنهاد می کند

کلمات کلیدی:

استخراج کلمات کلیدی، اسناد فارسی، رخداد همزمان، فازی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/41661>

