

عنوان مقاله:

خلاصه سازی چند سندی با استفاده از متن کاوی و راهکار گراف های رویداد

محل انتشار:

هشتمین کنفرانس بین المللی فناوری اطلاعات و دانش (سال: 1395)

تعداد صفحات اصل مقاله: 7

نویسندگان:

مریم داستانی داکره - دانشجوی کارشناسی ارشد مهندسی کامپیوتر- نرم افزار - دانشگاه آزاد رشت

فاطمه احمدی آبکناری - استادیار- گروه مهندسی کامپیوتر و فناوری اطلاعات- دانشگاه پیام نور رشت

خلاصه مقاله:

با توجه به افزایش حجم اسناد متنی برای پاسخگویی به نیازهای اطلاعاتی کاربران، تکنیک های بازیابی اطلاعات موجود به تنهایی از کارایی مناسب برخوردار نیستند. مطالعه حجم زیاد متن ها برای کاربران بسیار سخت و زمان گیر است و در اختیار داشتن خلاصه ای از مطالب مهم می تواند بسیار مفید باشد. باتوجه به اینکه تعداد اسناد خبری مرتبط با رویدادهای جهان واقعی و اطلاعات رویدادگرا، بهسرعت درحال رشد است بازیابی کارآمد و ارائه اطلاعات مختصر مرتبط با رویداد از اهمیت بسیاری در جلب نظر مخاطبان برخوردار است. در این مقاله راهکاری جدید مبتنی بر گراف رویداد به منظور بازیابی اطلاعات و خلاصه سازی چند سندی معرفی میگردد. در ابتدا از روش فضای بردار وزنی برای تشخیص عبارات تکراری استفاده شده و سپس میزان شباهت نمونه های خبری در قالب اسنپیت های خبری از پیکره متن اصلی و پرس جو با استفاده از ضریب تشابه دوبردار محاسبه میگردد. سپس با استفاده از گراف رویداد، یک مدل نمایش سند مبتنی بر رویداد برای معنانشناسی رویدادهای سطح جمله محاسبه می گردد که بر اساس آن اطلاعات مرتبط با رویدادهای توصیف شده در متن فیلتر شده و بازسازی میشود. در این روش با استفاده از هسته گراف ضرب تنسور و کونرمال، شباهت بین پرسوجوها و سندها اندازه گیری می شود. با توجه به کامل نبودن مدلهای موجود، راهکار معرفی شده در این مقاله با تکیه بر گراف رویداد شباهت بین پرسوجوها و سندها با تفکیک هم رخدادی رؤوس غیرمتناظر با استفاده از روش استخراج روابط معنایی موجود در متن و تکنیک های برچسب زنی معنایی لغات، اندازه گیری میشود و همچنین روابط زمانی بین آنها نیز تعیین می گردد. در گام بعد اسناد بر اساس نمرات شباهت بدست آمده رتبه بندی شده. نتایج ارزیابی چهار روش فوق دلالت بر افزایش چشمگیر صحت و دقت مدل پیشنهادی این مقاله در مقایسه با مدلهای فضای بردار وزنی، گراف ضرب کونرمال و گراف ضرب تنسور بر روی مجموعه های آزمایشی رویدادگرای خبری دارد.

کلمات کلیدی:

استخراج رویداد، بازیابی اطلاعات، خلاصه سازی چندسندی، گراف رویداد، متن کاوی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/548713>

