

## عنوان مقاله:

ارائه مدلی جهت تشخیص سوالات تکراری در پایگاه داده StackOverflow

## محل انتشار:

دومین کنفرانس بین المللی افق های نو در علوم مهندسی (سال: 1397)

تعداد صفحات اصل مقاله: 15

## نویسندگان:

ابوالفضل عرب مختاری - مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه شیراز

سید مصطفی فخراحمند - استادیار بخش مهندسی و علوم کامپیوتر و فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر، دانشگاه شیراز

## خلاصه مقاله:

به منظور اشتراک گذاری دانش ها، افزایش سرعت در حل مشکلات نرم افزاری و استفاده از راه حل های آنلاین برای مشکلات برنامه نویسی، از وب سایت هایی تحت عنوان انجمن های پرسش و پاسخ استفاده می شود. وب سایت StackOverflow محبوب ترین و پرکاربردترین وب سایت در زمینه مشکلات مربوط به برنامه نویسی است. در سایت StackOverflow ممکن است دو یا چند سوال در مورد مسئله یکسانی پرسیده شده باشند و چون زمان برای کسی که سوال را پرسیده بسیار مهم است می توان با ارائه سوالات مشابه که قبلا پاسخ داده شده اند کمک به سزایی به آن ها کرد و از این رو کاربر لازم نیست مدت زیادی را برای یافتن پاسخ خود منتظر بماند. پیدا کردن و مشخص کردن سوالات تکراری در StackOverflow در حال حاضر توسط مدیران و کاربران با امتیاز بالا به صورت دستی انجام می شود. از آنجاکه روزانه تعداد سوالات زیادی مطرح می شود مشخص کردن سوالات تکراری به صورت دستی بسیار زمان بر و پرهزینه است. از این رو ایجاد یک روش خودکار برای پیدا کردن سوالات تکراری یا مشابه می توانند کمک بسیار زیادی به توسعه دهندگان نرم افزار و مدیران سایت کند. مدل پیشنهادی در این مقاله، ضمن کشف و استخراج موضوع از متن سوالات، با بهره گیری از الگوریتم وزن دهی اسناد در سیستم های بازیابی اطلاعات و استفاده از مدل ماشین بردار پشتیبان (SVM) یک رویکرد جدید در زمینه تشخیص سوالات تکراری موجود در وب سایت StackOverflow ارائه می دهد. این مطالعه در جهت بهبود کارهای گذشته و افزایش دقت بازیابی سوالات تکراری در این حوزه است. نتایج به دست آمده نشان دهنده کارایی بالاتر و موثرتر مدل پیشنهادی در مقایسه با کارهای انجام شده در گذشته می باشد.

## کلمات کلیدی:

سوالات تکراری، StackOverflow، معیار شباهت، مدل فضای برداری (VSM)

## لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/875738>

