

عنوان مقاله:

خوشه بندی اسناد وب فارسی با استفاده از ترکیب روشهای داده کاوی و یک الگوریتم تکاملی

محل انتشار:

ششمین کنفرانس بین المللی یافته های نوین علوم و تکنولوژی با محوریت علم در خدمت توسعه (سال: 1398)

تعداد صفحات اصل مقاله: 8

نویسندگان:

پریسا چابکی بناب - دانشجوی مقطع کارشناسی ارشد رشته ی نرم افزار، واحد بناب، دانشگاه آزاد اسلامی، بناب، ایران

علی رضا حاجی اسکندر - عضو هیئت علمی گروه کامپیوتر، واحد بناب، دانشگاه آزاد اسلامی، بناب، ایران

سیدعلی شریفی - عضو هیئت علمی گروه کامپیوتر، واحد بناب، دانشگاه آزاد اسلامی، بناب، ایران

خلاصه مقاله:

افزایش حجم اطلاعات روی وب و جستجو در بین هزاران صفحه وب باعث سردرگمی میشود. به منظور دستیابی سریع به اطلاعات مورد نیاز در وب، لازم است روشی خودکار برای خوشه بندی صفحات وب ارائه شود. در مساله خوشه بندی صفحات وب، هدف دسترسی سریع و دقیق به اطلاعات وب میباشد. تحقیقاتی که در گذشته انجام گرفته هر یک با روشهای متفاوت سعی در بهبود نتایج خوشه بندی اسناد وب داشته اند. این تحقیقات بیشتر بر روی اسناد وب انگلیسی بوده است؛ در این مقاله خوشه بندی بر روی صفحات وب فارسی اعمال شده است. در این مقاله، خوشه بندی با استفاده از 100 صفحه وب در 5 حوزه ورزشی، خبری، پزشکی، آموزشی و دانشگاهی انجام شده است. وبسایت هایی با موضوع های مشابه به هم، در یک خوشه قرار میگیرند تا نتیجه جستجو بین هزاران وب موجود، دقیق و سریع باشد. این کار با استفاده از ترکیب الگوریتم بهینه سازی ازدحام ذرات و K-Means انجام گرفته است. نتایج به دست آمده حاکی از آن است که روش پیشنهادی تمامی گروه ها را با دقت بالای 80% و دو گروه ورزشی و خبری را با دقت 95% خوشه بندی میکند.

کلمات کلیدی:

خوشه بندی، الگوریتم بهینه سازی ازدحام ذرات، الگوریتم تکاملی، تگ های HTML، الگوریتم K-Means

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/920218>

