

عنوان مقاله:

خوشه بندی موضوعی اسناد XML در موتورهای جستجو

محل انتشار:

همایش مهندسی کامپیوتر و توسعه پایدار با محوریت شبکه های کامپیوتری، مدلسازی و امنیت سیستم ها (سال: 1392)

تعداد صفحات اصل مقاله: 10

نویسندگان:

حامد زجاجی - عضو هیئت علمی دانشگاه آزاد اسلامی واحد سبزوار

مهدی رواخواه - عضو هیئت علمی دانشگاه آزاد اسلامی واحد فریمان

محمد علیشاهی - عضو هیئت علمی دانشگاه آزاد اسلامی واحد فریمان

خلاصه مقاله:

امروزه موتورهای جستجو نقش مهمی برای دسترسی به اطلاعات در فضای وب دارند. حجم زیاد نتایج باعث سردرگمی کاربران می شود. بنابراین خوشه بندی نتایج جستجو و انتخاب یک برچسب مناسب مبتنی بر محتوای هر خوشه، ضروری است. در این مقاله سعی شده است نتایجی که در جستجوی اسناد XML بدست می آیند با استفاده از داده کاوی به شکلی کارا خوشه بندی و سپس برچسب گذاری شوند. از آنجا که عمل برچسب گذاری گام بلافصل پس از خوشه بندی می باشد اکثر روش های برچسب گذاری، منتج از الگوریتم های خوشه بندی می باشند. در این مقاله نیز با ایجاد تغییراتی در یک الگوریتم خوشه بندی بهینه، قابلیت برچسب گذاری خوشه ها را به آن اضافه کرده ایم. الگوریتم پیشنهادی تحت عنوان GLXCLS قادر است اسناد XML را دریافت و حین انجام خوشه بندی، اطلاعات لازم برای برچسب گذاری را جمع آوری کند و در نهایت پس از برقراری ارتباط با یک پایگاه داده لغوی و تحلیل داده ها، به کمک بیشترین تکرار عناصر و مشابهت ساختاری، مناسب ترین برچسب برای هر خوشه را پیشنهاد دهد. نتایج حاصل از ارزیابی الگوریتم روی مجموعه معتبری از داده ها با استفاده از معیار کلمات پرتکرار و پیشگویانه نشان دهنده برتری الگوریتم پیشنهادی در دقت و نزدیکی برچسب های انتخاب شده نسبت به روش های موجود می باشد.

کلمات کلیدی:

اسناد XML ، خوشه بندی، برچسب گذاری

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/239016>

