

عنوان مقاله:

خوشه بندی اسناد ایکس ام ال با رویکردی مبتنی بر سطح و هستی شناسی کلمات در شبکه های معنایی

محل انتشار:

دومین کنفرانس بین المللی بازشناسی الگو و تحلیل تصویر (سال: 1393)

تعداد صفحات اصل مقاله: 7

نویسندگان:

سارا اسدی حق - دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد قزوین، قزوین، ایران

فریبرز محمودی - استادیار، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد قزوین، قزوین، ایران

خلاصه مقاله:

منابع متنی موجود در وب که دارای ساختار مبتنی بر ایکس ام ال می باشند به سرعت در حال افزایش هستند و یکی از روش هایمداول برای سازماندهی آنها استفاده از شیوه های خوشه بندی می باشد. ارزیابی شباهت به عنوان یک مسئله مهم در خوشه بندی مطرح است و الگوریتم های زیادی در این زمینه بر اساس ساختار و محتوا ارائه شده اند. در خوشه بندی با چالش هایی مانند ابعاد بالا و دقت مواجه هستیم، در ارزیابی شباهت محتوا انتخاب کلمات کلیدی می تواند در رفع این چالش ها بسیار موثر باشد و استفاده از اطلاعات معنایی میتواند منجر به انتخاب مناسب این کلمات شود، که این اطلاعات به صورت مجموعه های هستی شناسی مانند وردنت موجود می باشند. از این رو در این مقاله تمرکز بر استفاده از شبکه های معنایی به منظور کاهش ابعاد و افزایش دقت بوده و رویکرد پیشنهادی با تعیین سطوح کلمات در شبکه معنایی و مفاهیم نظریه اطلاعات و تشکیل زنجیره های معنایی کلمات و انتخاب کلمات کلیدی مناسب ارائه می شود، نتایج به دست آمده در مقایسه با کارهای گذشته نشان دهنده افزایش میانگین دقت و شاخص خوشه بندی به میزان 0.24 و 0.28 می باشد.

کلمات کلیدی:

ایکس ام ال، خوشه بندی، زنجیره کلمات، شبکه های معنایی، هستی شناسی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/373464>

