

عنوان مقاله:

دسته بندی افزایشی کلان داده های اخبار وب با مدل سازی و سلسله مراتب موضوع

محل انتشار:

کنفرانس ملی پیشرفت های اخیر در مهندسی و علوم نوین (سال: 1397)

تعداد صفحات اصل مقاله: 9

نویسندگان:

سینا دامی - استادیار گروه کامپیوتر، دانشکده فنی و مهندسی، واحد تهران غرب، دانشگاه آزاد اسلامی، تهران، ایران

لیلا کاشی ها - دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، واحد تهران غرب، دانشگاه آزاد اسلامی، تهران، ایران

خلاصه مقاله:

رسانه های آنلاین یکی از کانال های مهم کاربران اینترنت برای دسترسی به اخبار است. وب سایت های خبری روزانه انبوهی از مقالات خبری را منتشر می کنند. هر روز حجم عظیمی از اخبار آنلاین تولید و به روزرسانی می شود. پردازش و تحلیل این حجم عظیم داده ها یک چالش مهم است. این چالش باید با استفاده از تکنیک های کلان داده ها که حجم عظیمی از داده ها را ظرف مدت زمان کوتاهی پردازش می کنند مرتفع شود. همچنین، از آنجا که ما به سمت انفجار اطلاعات رسانه های اجتماعی می رویم، تکنیک هایی مانند متن کاوی یا تجزیه و تحلیل شبکه های اجتماعی باید جدی گرفته شود. این مقاله بر روی یکی از رایج ترین کارهای روزمره یعنی خواندن اخبار وب، تمرکز می کند. وب سایت های خبری هزاران مقاله تولید می کنند که بازه وسیعی از موضوعات و دسته بندی ها را شامل می شود. اینمساله می تواند بعنوان یک مساله کلان داده در نظر گرفته شود. برای استخراج اطلاعات مفید، اینمقاله های خبری باید توسط تکنیک های کلان داده ها پردازش شوند. در این زمینه، یک رویکرد برای گروه بندی حجم عظیمی از مقالات خبری مختلف در دسته ها و موضوعات مختلف بر اساس محتوای آنها ارائه دادیم. چون این گروه بندی ها توسط مقالات جدید به روزرسانی می شوند، رویکرد پیشنهادی ما مبتنی بر یک مدل سازی موضوعی و با استفاده از یک الگوریتم یادگیری افزایشی و سلسله مراتبی موضوع است. روش پیشنهادی می تواند مدل توصیف کننده یک دسته را بر اساس تغییرات محتوای مقالات متناظر در زمان بلادرنگ به روزرسانی کند. این رویکرد با موفقیت با استفاده از اخبار آنلاین مورد ارزیابی قرار گرفته است.

کلمات کلیدی:

کلان داده ها، اخبار وب، متن کاوی، مدل سازی موضوعی، یادگیری افزایشی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/824397>

