

عنوان مقاله:

ارائه روشی برپایه متن کاوی برای استخراج اطلاعات آب و هوا

محل انتشار:

نخستین کنفرانس بین المللی فناوری اطلاعات (سال: 1394)

تعداد صفحات اصل مقاله: 13

نویسندگان:

توحید باقری زنده دل - گروه کامپیوتر، واحد گرمی، دانشگاه آزاد اسلامی، گرمی، ایران

محمدعلی بالاقر - گروه فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز، تبریز، ایران

خلاصه مقاله:

با توجه به ازدیاد بیش از حد داده ها در فضای مجازی و همچنین زیاد شدن تعداد پایگاه های اینترنتی، داده های ساخت نیافته مانند داده های متنی و وب نیز روز به روز در حال افزایش هستند. با اکتشاف در این داده ها می توان الگوهای جذابی را از آن ها استخراج نمود. یکی از روش های استخراج الگو در این نوع داده ها روش های متن کاوی و داده کاوی می باشند که الگوریتم های متعددی را به منظور کشف دانش از داده های متنی موجود در وب محیا کرده اند. از طرف دیگر، هر روز نیاز بیشتری در خصوص استفاده از داده های موجود به منظور استخراج دانش و استفاده از این دانش در زندگی روزمره احساس می شود. بعنوان مثال می توان براساس داده ها و اخبار موجود آب و هوا که در وب سایت های مختلف منتشر می شوند، مدلی به منظور پیش بینی وضعیت آب و هوا بطور خودکار تهیه نمود که وضعیت لحظه ای آب و هوا را براساس استخراج متون و متن کاوی و با استفاده از روش های داده کاوی تحلیل نماید. در این تحقیق هدف پیش بینی آب و هوا براساس متن کاوی اخبار می باشد. به این منظور در حدود 320 اخبار از سایت های مختلف جمع آوری و برچسب گذاری شدند. این دسته بندی دو رده هوای مطلوب و هوای نامطلوب را شامل می شود. سپس با استفاده از تکنیک های متن کاوی مانند قطعه قطعه کردن متن، ویژگی های متن استخراج گردید. در مرحله آماده سازی روش انتخاب ویژگی های تأثیرگذار بر روی داده ها اعمال شد. سپس با استفاده از الگوریتم های مدل سازی شبکه های عصبی مصنوعی، درخت تصمیم، بیز ساده و K نزدیکترین همسایگی مدل سازی به منظور پیش بینی محتوای متن صورت پذیرفت. در نهایت مشخص شد که روش شبکه های عصبی مصنوعی با استفاده از انتخاب ویژگی های تأثیرگذار به روش شاخص سود اطلاعاتی با کسب شاخص دقت 97% و شاخص F برابر با 96% از سایر روش ها عملکرد بهتری در تشخیص متون داشته است.

کلمات کلیدی:

متن کاوی، داده کاوی، شبکه های عصبی مصنوعی، درخت تصمیم، بیز ساده، K نزدیکترین همسایگی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/478117>

