

عنوان مقاله:

بررسی تأثیر روشهای طبقه بندی بر میزان تشخیص صفحات وب اسپم

محل انتشار:

سومین کنفرانس بین المللی پژوهشهای کاربردی در مهندسی کامپیوتر و فن آوری اطلاعات (سال: 1394)

تعداد صفحات اصل مقاله: 14

نویسندگان:

فائزه اصدقی - دانشجوی دکتری هوش مصنوعی، دانشکده کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شاهرود، سمنان، ایران

علی سلیمانی - دانشیار گروه هوش مصنوعی، دانشکده کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شاهرود، سمنان، ایران -

خلاصه مقاله:

امروزه با توجه به رشد اطلاعات در وب، موتورهای جستجو به عنوان یک ابزار برای ورود به دنیای وب مورد توجه قرار گرفته اند. آنها فهرستی از نتایج مرتبط با پرسش کاربر را در اختیار او قرار می دهند. از آنجا که اکثر کاربران تنها نتایج صفحه نخست و از آن میان فقط روی سه یا پنج پیوند اولیه را مورد بازدید قرار می دهند، حضور یک صفحه در نتایج بالای موتورهای جستجو به معنای بازدیدکننده بیشتر و نیز درآمد بیشتر است. در این میان وب اسپم یک روش غیرقانونی و غیراخلاقی به منظور افزایش رتبه صفحات اینترنتی توسط فریب الگوریتم های موتورهای جستجو می باشد. از آنجا که هرز ارائه شده کیفیت نتایج برای موتورهای جستجو اهمیت بسیاری دارد، روش های مختلفی برای تشخیص صفحات وب است. در این مقاله قصد داریم تأثیر روش های منتخب طبقه بندی را بر میزان تشخیص این صفحات با در نظر گرفتن چگونگی انتخاب ویژگی ها، بررسی نماییم. بدین منظور از مجموعه داده UK-WEBSPPAM- 2007 استفاده کرده و 21 روش مختلف طبقه بندی را برای تشخیص صفحات وب اسپم از دیگر صفحات بر روی حالات مختلف انتخاب ویژگی اعمال کردیم. بهترین نتیجه از اعمال الگوریتم های طبقه بندی بر ترکیب ویژگی های مبتنی بر محتوا و ویژگی های مبتنی بر پیوند بدست آمد

کلمات کلیدی:

وب اسپم، طبقه بندی، ویژگی های مبتنی بر محتوا، ویژگی های مبتنی بر پیوند، ویژگی های مبتنی بر پیوند تبدیل یافته

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/467133>

