

عنوان مقاله:

بررسی روش های پالایش متنی صفحات وب

محل انتشار:

نخستین همایش داخلی مهندسی کامپیوتر برق و فناوری اطلاعات (سال: 1392)

تعداد صفحات اصل مقاله: 8

نویسندگان:

محمد رضا فیضی درخشی - استادیار، گروه مهندسی کامپیوتر، دانشگاه تبریز

شیرا رشیدی - کارشناسی ارشد، گروه علوم کامپیوتر، دانشگاه تبریز

خلاصه مقاله:

توسعه سریع شبکه جهانی وب، تنوع دیدگاه ها و توانایی ارسال هر نوع اطلاعاتی روی اینترنت، موجب شده است تا دسترسی به محتوای موهن، مضر و غیر قانونی بیش از پیش آسان گردد. برای کنترل محتوای متنی اینترنت شناسایی صفحات وب با محتوای مضر حائز اهمیت است. با طبقه بندی صفحات وب می توان چنین صفحاتی را از صفحات سالم تشخیص داده و از دسترسی آسان به اینگونه مطالب جلوگیری کرد. با توجه به اینکه از واقعیت پالایش مطالب اینترنت گریزی نیست، شناخت و بررسی روش های پالایش متنی به فهم صحیح از پالایش، نقاط قوت و ضعف آن ها کمک شایانی می نماید. به همین دلیل در این مقاله مختصراً به پالایش و روش های آن و برخی روش های رایج طبقه بندی در امر پالایش پرداخته شده است. اگرچه بیان همه روش ها ممکن نیست، اما این مقاله می تواند آگاهی مناسبی از پالایش متنی و روش های آن، برای خواننده فراهم کرده و فرد را به منابع مناسب هدایت کند.

کلمات کلیدی:

پالایش، طبقه بندی صفحات وب، فراوانی کلمات، انطباق کلید واژه، تشابه اسناد

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/267662>

