

شناسایی و دسته بندی نظام‌مند ویژگی‌های علمی و فنی دانشنامه چندزبانه ویکی‌پدیا به عنوان یک

پایگاه‌دانش غنی در حوزه‌ی داده‌کاوی

امیر حسین جدیدی نژاد^{۵۰}، محمدرضا کیوان‌پور^{۵۱}

چکیده

ویکی‌پدیا بعنوان بزرگ‌ترین دانش‌نامه‌ی چندزبانه‌ی آزاد دنیا که توسط میلیون‌ها نفر در سرتاسر جهان به‌روزرسانی و نگهداری می‌شود، در سال‌های اخیر توجه تعداد زیادی از محققان را در حوزه‌های مختلفی مانند پردازش زبان طبیعی، داده‌کاوی و بازیابی اطلاعات به خود جلب نموده است و به عنوان یکی از مهمترین منابع اطلاعاتی در حوزه‌های پیش گفته مطرح می‌باشد. علت این امر را می‌توان در کیفیت بالای متن مقالات، لینک‌های ارتباطی غنی میان آنها و حجم بسیار زیاد مقالات دانست. با وجود رشد چشمگیر ویکی‌پدیا و محتوای معنایی غنی آن، تاکنون تلاش مؤثری جهت معرفی ویژگی‌های علمی و فنی آن به شکلی نظام‌مند و مطالعه تحلیلی پژوهش‌های انجام شده در این حوزه صورت نگرفته است و پژوهشگران مختلف هر کدام سعی در استفاده از بخشی از محتوا و امکانات ویکی‌پدیا، بسته به علائق کاری و سلايق خود نموده‌اند. در این پژوهش ضمن تحلیل و ارزیابی تحقیقات پیشین صورت گرفته روی منبع داده‌ی ویکی‌پدیا در حوزه‌ی پردازش زبان طبیعی، داده‌کاوی و بازیابی اطلاعات، خصوصیات و ویژگی‌های علمی و فنی ویکی‌پدیا به شکلی نظام‌مند شناسایی و معرفی می‌شود. همچنین در این مقاله نقش و جایگاه هر یک از ویژگی‌های مذکور در حوزه‌های کاربردی مختلف تبیین می‌گردد. نتیجه این تحقیق بسترساز استفاده مؤثر و علمی از ویکی‌پدیا، به عنوان یکی از غنی‌ترین منابع اطلاعاتی، در حوزه داده‌کاوی می‌باشد. بر این اساس زمینه مناسبی جهت مقایسه فنی میان روش‌های مختلف داده‌کاوی، بهسازی روش‌های مختلف داده‌کاوی از طریق شناخت نقاط ضعف و قوت آنها و حتی ارائه روش‌های کامل‌تر در حوزه داده‌کاوی فراهم می‌شود.

کلمات کلیدی

پایگاه دانش ویکی‌پدیا، داده‌کاوی، پردازش زبان طبیعی، بازیابی اطلاعات، دسته‌بندی متون، ارتباط معنایی لغات، شبکه معنا.

⁵⁰ دانشجوی کارشناسی ارشد، گرایش نرم‌افزار، دانشگاه آزاد اسلامی قزوین.

⁵¹ عضو هیات علمی دانشگاه آزاد اسلامی قزوین.