

کاوش قوانین انجمنی در اسناد XML با روش

شاخص جدولی بهینه شده

علیرضا زارع فرخادی^{۲۲۰}، امیر اسمعیل زاده^{۲۲۱}، افشین مهرابی^{۲۲۲}، فریبرز محمودی^{۲۲۳}

چکیده

استفاده از اسناد XML بعنوان یک استاندارد جهانی، در پایگاه داده‌ها روزبه‌روز در حال افزایش است. از اینرو روابط موجود بین داده‌های XML بعنوان یک چالش عمده مطرح و نیاز به استخراج دانش از داده‌های خام جهت افزایش هوش سازمانی کاملاً محسوس است. روش شاخص جدولی بعنوان یک روش در این راستا مطرح است. در این مقاله راهکاری جهت بهبود این روش ارائه شده است که هم از نظر فرآیند کاوش و هم از نظر زمان استخراج اطلاعات، روش نام برده را بهبود بخشیده است. روش شاخص جدولی قوانین را از محتویات XML بدون اطلاع از روابط موردنظر کاربر استخراج کرده و به کاوش قوانین ناشناخته، بروی اسناد با ساختارهای مشابه می‌پردازد. درحالیکه در روش ارائه شده در این مقاله بدون اینکه اطلاعی از ساختار اسناد در دسترس باشد، می‌توان به کشف قوانین ناشناخته پرداخت. برای این منظور، در ابتدا داده‌های خام را از اسناد XML استخراج و بر اساس ساختار درختی آن، جدول رابطه‌ای مشتمل بر تراکنشها و آیتمها لیست شده و سپس به روش کاوش قوانین انجمنی، قوانین معتبر بدست آمده و دانش مورد نیاز از روی آن قوانین بدست می‌آید.

کلمات کلیدی

پایگاه داده XML، داده کاوی، کاوش قوانین انجمنی.

²²⁰ کارشناسی ارشد مهندسی کامپیوتر- نرم‌افزار، دانشگاه آزاد اسلامی- قزوین، a.farkhady@gmail.com

²²¹ کارشناسی ارشد مهندسی کامپیوتر- نرم‌افزار، دانشگاه آزاد اسلامی- قزوین، amir.esmaeilzadeh@gmail.com

²²² کارشناسی ارشد مهندسی کامپیوتر- نرم‌افزار، دانشگاه آزاد اسلامی- قزوین، afshin.mehrabi@gmail.com

²²³ استادیار گروه کامپیوتر، دانشکده برق، کامپیوتر و فناوری اطلاعات، دانشگاه آزاد اسلامی- قزوین، mahmoudi@itrc.ac.ir