

عنوان مقاله:

استخراج اطلاعات: روش ها و کاربردها

محل انتشار:

اولین کنفرانس بین المللی بازیابی تعاملی اطلاعات (سال: 1395)

تعداد صفحات اصل مقاله: 7

نویسندگان:

داود حاصلی - دانشجوی دکترای علم اطلاعات و دانش شناسی دانشگاه تهران

ملوک السادات حسینی بهشتی - استادیار پژوهشگاه علوم و فناوری اطلاعات ایران

سمیه پاک نهاد - دانشجوی دکترای علم اطلاعات و دانش شناسی دانشگاه شیراز

خلاصه مقاله:

استخراج اطلاعات اشاره به استخراج خودکار اطلاعات ساختار یافته همچون موجودیت ها، روابط بین موجودیت ها، و موجودیت های توصیف ویژگی از منابع غیرساختاریافته دارد. این امر قالب های پرس و جو برای منابع غیر ساختاریافته را نسبت به جستجوهای کلید واژه های مجرد، بسیار غنی تر می سازد. هنگامی که داده های ساختاریافته و غیرساختاریافته با هم موجود هستند، استخراج اطلاعات، یکپارچه سازی این دو نوع منبع و اعمال جستجو بین آنها را امکان پذیر می سازد. استخراج خودکار اطلاعات از منابع غیر ساختار یافته راه های جدیدی را برای جستجو، سازمان دهی و تجزیه و تحلیل داده ها با بهره گیری از علم معناشناسی از پایگاه داده های ساختار یافته و داده های غیرساختاریافته گشوده است. امروزه، تمایل زیادی به تبدیل دسکتاپ شخصی به پایگاه داده های ساختاریافته، تبدیل دانش انتشارات علمی به پیشینه های ساختاریافته، و مهار اینترنت برای پرس و جوهای کشف حقایق ساختاریافته وجود دارد. در نتیجه، جوامع مختلف زیادی از پژوهشگران وجود دارند که درصدد استفاده از فنون یادگیری ماشین، بازیابی اطلاعات، و زبان شناسی رایانشی برای جنبه های مختلف مشکل استخراج اطلاعات هستند. استخراج اطلاعات شامل دو نوع استخراج هدفمند اطلاعات و استخراج آزاد اطلاعات است. استخراج هدفمند اطلاعات نوع خاصی از اطلاعات را که احتمالا به شکل واضحی در متن بیان شده اند، برای اهداف مشخصی استخراج می کند. سه روش عمده برای استخراج هدفمند اطلاعات عبارتند از: استخراج اطلاعات مبتنی بر قاعده، مدل های گرافی و تعریف تابع هسته در استخراج آزاد اطلاعات با تعدادی نمونه آغازین مربوط به هر رابطه و بدون هیچ داده ای برای آموزش، اطلاعات را استخراج می کنند. در استخراج آزاد اطلاعات از سامانه های بی ناظر، نیمه نظارتی و خود ناظر استفاده می شود. کاربردهای استخراج هدفمند اطلاعات را در چهار گروه کاربردهای تجاری (شامل پیگیری اخبار، مراقبت از مشتری، تمایز داده ها و تبلیغات طبقه بندی شده)، کاربردهای شخصی، کاربردهای علمی، و کاربردهای مبتنی بر وب (شامل پایگاه های استنادی، پایگاه داده های نظریه ای، وب سایت های اجتماع ها، خرید مقایسه ای، گنجاندن تبلیغ در صفحات وب، و جستجوهای وب ساختار یافته) دسته بندی می شود. کاربردهای استخراج آزاد اطلاعات شامل استخراج دانش و استنتاج از متن زبان طبیعی است. چالش های استخراج آزاد اطلاعات نیز شامل عدم دخالت انسان، ناهمگونی در پیکره، و کارایی است. در مقاله حاضر هریک از موارد گفته شده به تفصیل معرفی و بحث شده اند.

کلمات کلیدی:

استخراج اطلاعات، استخراج هدفمند اطلاعات، استخراج آزاد اطلاعات

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/572863>



