

عنوان مقاله:

مروری بر سیستم های استخراج اطلاعات و روش های استخراج رابطه

محل انتشار:

چهارمین کنفرانس ملی دستاوردهای نوین در برق و کامپیوتر و صنایع (سال: 1397)

تعداد صفحات اصل مقاله: 11

نویسندگان:

افشین شهرکی - عضو هیئت علمی، گروه فناوری اطلاعات، دانشگاه غیرانتفاعی و غیردولتی هاتف، زاهدان

میثم محمودی - دانشجوی کارشناسی ارشد، دانشگاه غیرانتفاعی و غیردولتی هاتف، زاهدان

خلاصه مقاله:

داده های غیرساخت یافته، که اکثرا به شکل داده های متنی هستند، حدود 80 درصد از اطلاعات دیجیتالی موجود را تشکیل می دهند. اما یافتن، دسترسی، تحلیل یا استفاده از این داده های غیرساخت یافته به راحتی امکان پذیر نیست. بنابراین به ابزارها و روش هایی نیازمند هستیم که اطلاعات را پیدا کرده و آنها را به پایگاه های دانشی تبدیل کند که قابل استفاده توسط الگوریتم ها باشد. در این نوشتار، ابتدا به معرفی سیستم های استخراج اطلاعات می پردازیم و معماری و اجزای مختلف سیستم های استخراج اطلاعات و همچنین task های اصلی سیستم های استخراج اطلاعات شامل تشخیص موجودیت های نام دار، تشخیص اشاره های مختلف به موجودیت های نام دار، استخراج رابطه و استخراج رخدادها را معرفی می کنیم. و در ادامه روش های مختلف استخراج رابطه (که یکی از Task های مهم استخراج اطلاعات است)، از جمله، روش های استخراج رابطه بوسیله ی الگوهای نوشته شده با دست، روش های با ناظر برای استخراج رابطه، روش های Bootstrapping و مبتنی بر یادگیری الگو و روش های نظارت از راه دور برای استخراج رابطه، را معرفی می کنیم. در نهایت با توجه به محدودیت هایی که در روش های استخراج اطلاعات کلاسیک وجود دارند به معرفی روش های استخراج اطلاعات باز می پردازیم که با کمترین پیش فرض ها تنها مجموعه ای از اسناد ورودی را به عنوان ورودی می پذیرند و تمام روابط موجود در این اسناد غیرساخت یافته را استخراج می کنند.

کلمات کلیدی:

استخراج اطلاعات، استخراج رابطه، اطلاعات دیجیتالی، پایگاه داده

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/851866>

