

چکیده

هدف اصلی داده‌های پیوندی، تحقق وب معنایی و استخراج دانش از طریق پیوند دادن داده‌های موجود روی وب می‌باشد. یکی از موانع دستیابی به این هدف، وجود مشکلات و خطاها در داده‌های منتشر شده است که باعث ایجاد پیوندهای نادرست و در نتیجه استنتاج‌های نامعتبر می‌گردد. مطالعه کارهای انجام شده در این حوزه نشان می‌دهد که بیشتر محققان بر ارزیابی کیفیت داده پس از انتشار تمرکز داشته و ارزیابی داده‌ها پیش از انتشار را برعهده مالک یا منتشرکننده داده گذاشته‌اند. با توجه به اینکه کیفیت داده‌ها تأثیر مستقیم بر موفقیت پروژه داده‌های پیوندی و تحقق وب معنایی دارد، بهتر است تا کیفیت هر یک از مجموعه‌های داده در مراحل اولیه انتشار و قبل از اضافه شدن به ابر داده‌های پیوندی ارزیابی شود.

در این رساله، یک رویکرد مبتنی بر سنجه برای ارزیابی پیش از انتشار مجموعه داده‌های پیوندی ارائه می‌شود. برای این منظور، با مطالعه مدل‌ها و چارچوب‌های کیفیت داده، مدل ISO-25012 به‌عنوان مدل مبنا انتخاب و با نگاشت ابعاد کیفیت مدل مبنا به ابعاد کیفیت داده‌های پیوندی، شش بُعد کیفی کامل‌بودن، دقت معنایی، دقت نحوی، سازگاری، یکتایی و پیوندپذیری به‌عنوان ابعاد کیفیت ذاتی داده‌های پیوندی شناسایی شده‌اند. سپس، با استفاده از رویکرد هدف، پرسش، سنجه برای هر یک از ابعاد کیفی شش-گانه، سنجه‌های مناسب تعریف و به‌صورت رسمی بیان شده است. استراتژی ارزیابی مدل پیشنهادی، شامل سه مرحله اعتبارسنجی تئوری، ارزیابی تجربی به روش آزمایشات و نظرسنجی از خبرگان بوده که فرایند اجرای هر یک، به تفصیل در رساله مورد بررسی قرار گرفته است. در پایان، با بهره‌گیری از روش‌های یادگیری، یک مدل پیش‌بینی کیفیت بر مبنای سنجه‌های پیشنهادی ارائه شده است. نتایج ارزیابی‌ها نشان داد که سنجه‌های پیشنهادی معتبر، کاربردی و مقیاس‌پذیر هستند و همچنین قابلیت ارزیابی کیفیت ذاتی هر مجموعه داده را به‌صورت کاملاً خودکار دارند. بنابراین، منتشرکنندگان داده‌ها قادرند تا با استفاده از مدل پیشنهادی، کیفیت ذاتی مجموعه داده خود را قبل از انتشار ارزیابی کنند.