

بخشی از ترجمه مقاله



عنوان فارسی مقاله :

پاکسازی داده : مسائل و رویکردهای فعلی

عنوان انگلیسی مقاله :

Data Cleaning: Problems and Current Approaches

Erhard Rahm* Hong Hai Do

University of Leipzig, Germany

<http://dbs.uni-leipzig.de>

توجه !



این فایل تنها قسمتی از ترجمه میباشد.

برای تهیه مقاله ترجمه شده کامل با فرمت ورد (قابل ویرایش) همراه با نسخه انگلیسی

مقاله، [اینجا](#) کلیک نمایید.

بخشی از ترجمه مقاله

2 Data cleaning problems

This section classifies the major data quality problems to be solved by data cleaning and data transformation. As we will see, these problems are closely related and should thus be treated in a uniform way. Data transformations [26] are needed to support any changes in the structure, representation or content of data. These transformations become necessary in many situations, e.g., to deal with schema evolution, migrating a legacy system to a new information system, or when multiple data sources are to be integrated.

As shown in Fig. 2 we roughly distinguish between single-source and multi-source problems and between schema- and instance-related problems. Schema-level problems of course are also reflected in the instances; they can be addressed at the schema level by an improved schema design (schema evolution), schema translation and schema integration. Instance-level problems, on the other hand, refer to errors and inconsistencies in the actual data contents which are not visible at the schema level. They are the primary focus of data cleaning. Fig. 2 also indicates some typical problems for the various cases. While not shown in Fig. 2, the single-source problems occur (with increased likelihood) in the multi-source case, too, besides specific multi-source problems.

2. مسائل پاکسازی داده

در این بخش تعداد کثیری از مسائل مربوط به کیفیت داده طبقه بندی و به روش پاکسازی داده و تبدیل داده حل می شوند. بین این مسائل رابطه نزدیکی برقرار بوده و به همین خاطر باید به طور یکنواخت با آنها رفتار شود. برای حمایت از تغییرات صورت گرفته در ساختار، طرز نمایش یا محتوای داده، لازم است از روش تبدیل داده استفاده شود. در بسیاری از مواقع، از جمله برای رسیدگی به تکامل طرح، تبدیل سیستم کهنه به سیستم اطلاعاتی جدید و تلفیق منابع متعدد داده، تبدیل داده یکی از عوامل الزامی می باشد. همان گونه که در شکل 2 نشان داده شده است، بین مسائل یک و چند منبع و بین مسائل وابسته به طرح و نمونه تمایز قایل می شویم. مسائل وابسته به طرح دوره در نمونه ها نیز منعکس شده اند. در سطح طرح می توان آنها را با ارتقاء طرح (تکامل طرح)، ترجمه و تلفیق طرح بررسی نمود. از طرف دیگر، مسائل وابسته به نمونه، به خطاها و ناهمسانی های موجود در محتوای داده های حقیقی اشاره می کنند که در سطح طرح قابل مشاهده نیستند. در فرایند پاکسازی داده، نکته اصلی همین می باشد. شکل 2 پاره ای از مسائل مختلف را نیز نشان می دهد. این نکته در شکل 2 نشان داده نشده است که مسائل تک منبعی (با احتمال زیاد) به غیر از مسائل چند منبعی خاص، در مسئله چند منبعی نیز رخ می دهند.

توجه!

این فایل تنها قسمتی از ترجمه میباشد.



برای تهیه مقاله ترجمه شده کامل با فرمت ورد (قابل ویرایش) همراه با نسخه انگلیسی مقاله، [اینجا](#) کلیک نمایید.

همچنین برای مشاهده سایر مقالات این رشته [اینجا](#) کلیک نمایید.