

بخشی از ترجمه مقاله

عنوان فارسی مقاله :

Problems, Methods, and Challenges

In Comprehensive Data Cleansing

عنوان انگلیسی مقاله :

مشکلات ، روش ها و چالش های فرایند پاکسازی کامل داده

Heiko Müller, Johann-Christoph Freytag
Humboldt-Universität zu Berlin zu Berlin,
10099 Berlin, Germany

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 Motivation

The processing and analysis of data is motivated by different objectives. Data are symbolic representations of information, i.e., facts or entities from the world, depicted by symbolic values. Data is collected to form a representation of a certain part of the world, called the mini-world (M) or Universe of Discourse (UoD) [EN00]. For the rest of this paper we consider data items to be tuples, i.e., sets of related discrete values from a finite set of domains. Each tuple represents an entity from the mini-world with each of its values in turn representing one of the entity's properties, e.g., persons with their properties name, date of birth, height, and weight.

The main objective for collecting and processing data is to fulfil different tasks in (i) administration, e.g., to keep track of the employees in a company, the customers of our company, or the sales volume of our companies branches, (ii) supporting business process, e.g., using the customers address list for direct mailing to advertise new products, and (iii) analysis and interpretation to support decision making and to generate new information and knowledge. Decision making in business and government is based on analysis of data to discover trends and keep tab on developments and investments. Scientists analyse data to gain new information and knowledge about the partition of the world they are investigating.



2. انگیزه:

انگیزه های مختلفی انگیزه ای برای پردازش و تحلیل داده به شمار می رود. داده ها نماینده ای نمادین از اطلاعات به عبارتی حقایق یا واحدهای به تصویر کشیده شده توسط مقادیر نمادین به شمار می روند. هدف از گرد آوری داده ها نمایش بخش خاصی از دنیا به نام M یا UoD (DOU) می باشد. در بخش های دیگر این مقاله مجموعه های داده ها در نظر گرفته می شوند. به عبارتی مجموعه مقادیر ناپیوسته مجموعه حوزه های محدود هر مجموعه معرف یکی از واحد های دنیای کوچک بوده و ارقام خصوصیات واحد را نشان می دهد، مثلاً اشخاصی با نام خصوصیات، تاریخ تولد، ارتفاع و وزن.

هدف اصلی گرد آوری و پردازش داده ها انجام وظایف مختلف در زمینه 1 اجرا، مثلاً کنترل کارمندان در کمپانی، مشتریان کمپانی یا حجم فروش شعب کمپانی ها (2 حمایت از فرایندهای تجاری مثلاً استفاده از لیست آدرس مشتریان برای پست مستقیم جهت تبلیغ محصولات جدید (3 تحلیل و تفسیر جهت حمایت از تصمیم گیری و تولید اطلاعات و دانش جدید می باشد. در تجارت و دولت تصمیم گیری بر اساس آنالیز داده ها جهت کشف روندها و حفظ روند و سرمایه گذاری انجام می گیرد. هدف آنالیز داده ها از سوی دانشمندان کسب دانش و اطلاعات جدید در مورد تقسیم بندی دنیایی است که مشغول پژوهش آن هستند. خصوصیت داده هایی با نمایش غلط دنیای کوچک، خطا می باشد. آنها از اندازه گیری های نادرست، عادات غلط و ورودی تپل، حذف در حین جمع آوری و نگهداری داده و غیره نتیجه می گیرند.

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.