

信息质量及其评价指标

清华大学经济管理学院 苏强 梁冰

信息质量是信息管理的重点内容,本文对信息质量评价指标的划分进行分析,探讨了数据、信息及信息系统三者质量特性之间的关系,并在综合分析现有文献和观点的基础上,提出了一种新的信息质量指标体系。最后给出信息质量指标的定量分析方法。

引言

数字化生活和网络经济以及电子商务已成为现代生活的主要标志。由于信息对于现代社会的重要性,信息质量已成为信息管理的重点内容。如何来评价信息的质量特性,并给出量化的计算分析方法,正在成为信息技术、管理技术和计算机技术领域所关注的一个焦点问题。

本文重点针对信息质量评价指标的划分,研究了数据、信息及信息系统以及三者质量特性之间的关系,在综合分析现有文献和观点的基础上,提出了一种新的信息质量指标体系,并给出信息质量指标的定量分析方法。

信息质量指标研究综述

为了研究信息的质量特性,首先需要明确信息质量的评价指标。对信息质量评价指标的划分存在着许多观点,我们先对一些较典型的观点进行综述分析。

需要指出的是,许多这方面的文章在论述有关信息质量的概念时使用的均是“数据质量”这一名称,也未提及信息质量和数据质量之间的关系,并认为数据和信息的概念可以同义使用。因此,为了统一起见,在综述中均使用“数据质量”来进行论述。

传统上对数据质量的评价主要关注于数据的准确性,即所记录的数据是否符合实际情况,以及在数据的输入、存储、处理过程中出现的错(误)码率的高低等。实际上对数据质量的评价需要从多方面、多角度进行。Donald, P. Ballou 等人认为数据质量的评价指标有如下四项:

准确性:记录的数据值是否与真实值相符合;

时限性:记录的数据值是否已过时;

完整性:数据的记录是否完整;

一致性:数据值的表达是否在所有出现该数据的地方都是一致的。

此外,在其他的文献中还可以见到诸如数据可认证

性、可追踪性、可信性等多种数据质量评价指标。T.C.Redman甚至提出了20多项数据质量的评价指标,包括数据的准确性、完整性、一致性、循环周期等。

Diane M.Strong等人将15个数据质量评价指标进行了分类,如表1所示。

表1 数据质量分类及评价指标

类别	评价指标
固有数据质量	正确性、客观性、可信性、信誉度
可访问性数据质量	可访问性、访问安全性
关联性数据质量	相关性、价值增量、时限性、完整性、数据量
可表达性数据质量	可翻译性、易懂性、表达简洁性、一致性

Diane等人认为,固有数据质量是指由数据本身的性质决定的质量;可访问性数据质量则是指数据消费者是否既能比较容易的访问到所需的数据,同时又能保证访问过程的安全性;关联性数据质量指数据的收集和处理能否迅速地跟上不断变化的用户需求;可表达性数据质量指数据的表达能力。

综上所述,可以看出这些研究中存在如下一些不足:

(1)以上看法并未对数据质量和信息质量进行明确区分。在这些研究中,认为数据质量和信息质量是同义的。

(2)有些看法中的信息质量评价指标实际上包括了信息系统质量的评价指标在内。如Diane的分类方法中的可访问性数据质量和可表达性数据质量实际上应该是针对信息系统的(如可访问性等)。

信息质量评价指标体系

我们认为:将数据质量和信息质量同义看待,以及在信息质量的评价指标中融入信息系统质量的评价指标并不合理。因此,在这里先明确区分数据质量、信息质量和信息系统质量三者之间的关系;在此基础上,提出信息质量

的评价指标体系。

1. 数据质量、信息质量和信息系统质量

数据本身并不代表任何一类具体的东西,它仅仅是一种抽象的量的概念。而信息则是数据经过一定方式的处理后得到的,信息通过数据形式来表示,是加载在数据之上,对数据具体含义的解释。更重要的是:信息是直接面向用户的,对于不同的信息用户,它具有不同的意义和价值。这一点是数据所不具备的。因此,将数据质量和信息质量同义,容易抹煞掉信息和数据之间的重要区别。而信息系统的设计者们往往更注重数据质量,特别是数据的正确率、完整性、一致性等,因为这些指标直观明了,可以用许多技术手段加以控制,而对信息对于用户所特有的属性则认识不够或考虑不周,造成用户对信息质量的不满。

同样,在确定信息质量的评价指标时不应把明显属于信息系统的评价指标也归纳入在内。因为这样做一方面混淆了信息和信息系统在属性上的区别,另一方面会造成信息质量的评价指标过多过杂,分散了信息设计人员的注意力,也不利于信息系统的设计和实施。

图1和图2分别显示了数据、信息和信息系统之间的关系以及三者质量特性之间的关系。

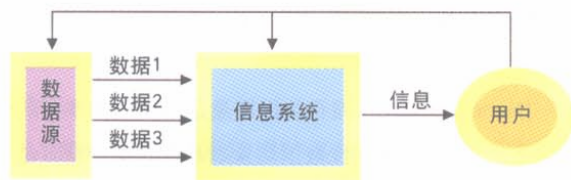


图1 数据、信息和信息系统之间的关系

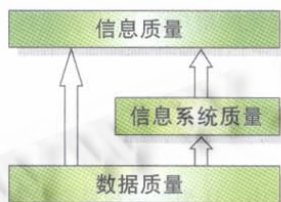


图2 数据质量、信息质量和信息系统质量之间的关系

由图1可见,数据是信息系统的“加工”源材料,信息是信息系统的产品。用户是信息产品的最终使用者,信息系统的最终目标是用户满意。因此,系统将会根据用户的要求不断调整信息系统和数据源,使信息更加具有价值。

由图2可见,数据质量是信息质量的基础,一部分数据直接影响信息质量(这部分数据在信息系统中只经过简单的传递,并不进行处理行变换),另一部分数据(在信息系统中进行一定的处理和变换的数据)的质量则是通过信息系统间接影响信息质量。因此,数据质量和信息系统的

质量两者相互作用共同决定了信息的质量。

2. 信息质量评价指标体系

信息质量评价宗旨是“使用户感到满意”。对信息质量的评价应该从多个角度进行,但主要应从用户的视角出发。这里,在综合前人研究中所共有的和比较一致的指标内容基础上,设计出信息质量的评价指标体系。

信息质量的评价指标应该包括如下方面:

(1)时限性: 时限性是指信息有明显的时间限制,超出这一时间限制的信息将失去价值。信息的时限性与数据的时间有效性是有区别的。数据只要在一定的时间范围内便是有效的,而信息不但要有效,还要及时。信息系统对数据的处理速度可以称为及时性。信息的时限性应受到及时性和数据的时间有效性的综合影响,即在数据有效时间内越及时产生的信息的时限性越好。

(2)有用性: 这个指标的概念很直观,即只有用户认为有用的信息才是好的信息。信息的有用性体现了设计的信息内容与用户期望的信息内容之间的差距。为了确保信息的有用性,就必须在信息设计阶段与用户进行密切联系,尽量使信息内容最大限度地满足用户需求。

(3)充分性: 信息不但要对用户有用,其内容还需要有适当的充分程度。充分程度过低会使用户得不到所需的详细信息;而信息过于充分,一方面会产生大量的信息冗余,浪费信息系统的处理资源,另一方面也迫使用户在大量的信息内容中寻找对自己有用的部分,造成用户信息获取上的困难。

(4)可信性: 应该使用户觉得其所获得的信息是可信的。信息的可信性来源于两个方面: 一方面是用于处理的数据来源是可信的,如由专家提供的数据要比一般人提供的数据更有权威性;另一方面是信息系统对数据的处理方式和处理手段应该是可靠的,如在进行数据计算或统计时所用到的公式或方法应该使用户觉得是有道理的。

(5)易用性: 信息应易于用户的使用。这在信息系统的设计上自然要求为用户提供强大而方便的信息查询手段。但就信息本身来说,信息应该是表达清晰、简洁明了、易于理解。这里并未提及信息的准确性、完整性等常见的信息质量评价指标,因为我们认为信息准确性、完整性是依靠数据质量的准确性和完整性来保证的。因此,应当将其划分为数据质量的评价指标之中。

信息质量的评价分析方法

明确了信息质量的评价指标后,就可以依据这些指

标对信息质量进行分析、评价并设法提高。对于不同的指标应分别采取定性或定量的方法进行评价。

信息系统对数据进行处理、产生信息的过程类似于产品的制造过程。因此也可以把信息系统称为信息制造系统。基于这种观点,可以把信息系统输出的信息称为信息产品。与一般产品设计过程类似,在信息产品的设计阶段,就可以明确信息产品的质量要求。所不同的是,信息产品的许多质量指标的评价需要在信息产品设计时进行,如信息的有效性、可信性、易用性等。对这些指标的评价需要在信息内容和格式大体确定后,由用户对其进行满意程度的评价,结合用户对这些质量指标的重视程度,得出综合评价。

可以将用户对信息的有效性、可信性、易用性等指标的重视程度划分为几个等级,用户为每个指标确定一个重视等级。并将用户对信息产品的该指标满意程度也划分为几个等级,由用户确定对在信息产品中落实该指标的满意程度。重视程度和满意程度的等级可以用定量的数字与之相对应,从而能够对综合满意程度进行量化计算。计算可采用如下公式:

$$S = \sum R_i W_i \quad (1)$$

其中 S 表示综合满意程度, W_i 表示用户对第 i 项指标的重视程度, R_i 表示用户对第 i 项指标的满意程度。

由于在信息产品设计过程中,对信息产品质量的评价并未涉及信息的时限性及数据质量等需要在数据处理过程中进行跟踪评价的指标,因此 S 并不代表信息对于用户的实际价值,可以把它看作是信息的内在价值 V_i , 因为信息的设计反映了信息的内在价值。

对于需要在数据处理过程中对数据的属性变化进行跟踪,从而做出质量评价的指标,如信息的时限性,以及数据质量等,则可以通过对信息系统的建模来进行评价。前面提到可以把信息系统看作是信息制造系统,于是可以依据制造系统建模的思路,来建立信息制造系统处理过程模型。

文献提出了一种信息制造系统的建模方法。该方法中包含的主要功能模块有(参见图3):

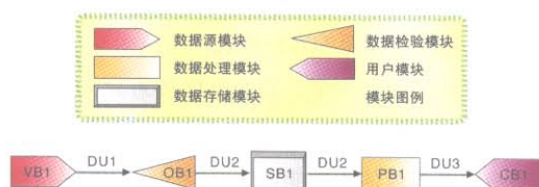


图3 一个简单的信息制造系统模型

· 数据源模块: 代表各种原始输入数据的来源。

· 数据处理模块: 代表信息制造过程中的各种数据处理过程。

· 数据存储模块: 代表用于存储数据的各种介质。

· 数据检验模块: 代表对数据质量进行检验和修补的过程。数据检验模块和数据处理模块的区别在于: 数据检验模块要对数据的准确性、完整性、一致性等进行检验; 而处理模块只根据预先设定好的操作程序对数据进行处理, 如合并输入数据等, 但它不会处理数据中存在的缺陷。

· 用户模块: 代表使用信息产品的用户。

图3给出了一个简单的信息制造系统模型。其中, 用带箭头的直线表示各模块之间的数据流入流出。在各模块中可以合理地设置计算公式, 以反映输入模块的数据经过处理后的质量变化情况。例如输出的数据质量 DQ_{out} 应受输入数据质量 DQ_{in} 和处理过程的正确性 (也可以称为处理效率, 用 PE 表示) 的影响, 它们之间有如下函数关系:

$$DQ_{out} = f(DQ_{in}, PE) \quad (2)$$

具体的函数表达方式则可根据实际情况确定。

通过对这些质量评价指标的跟踪, 便可最后得到信息产品的质量评价指标的值, 将这些指标与在信息设计阶段确定的、信息的内在价值 V_i 按某种方式进行合并计算, 便可以得到信息产品的实际价值 V_A 。

例如, 通过建模计算得到了信息产品的数据质量值 DQ 和时限性值 T , 可采用如下公式计算信息产品的实际价值 V_A :

$$V_A = V_i (W \times (DQ)^\alpha + (1-W) \times T^\beta) \quad (3)$$

公式(3)表示信息产品的实际价值 V_A 是信息产品的内在价值 V_i 与经过加权后的时限性 T 和数据质量 DQ 的乘积。权值 W 用于区分用户对于数据质量 DQ 和时限性 T 的重视程度, 而指数 α 、 β 则分别表示用户对于数据质量和时限性变化的敏感程度。

小结

本文对数据、信息和信息系统以及三者质量特性之间的关系进行了分析, 提出信息质量评价指标体系, 并设计出信息质量的定量分析方法。■