

ICS 35.240.70
L 70



中华人民共和国国家标准

GB/T 38667—2020

信息技术 大数据 数据分类指南

Information technology—Big data—Guide for data classification

2020-04-28 发布

2020-11-01 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

前言	I
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 分类过程	2
5.1 概述	2
5.2 分类规划	3
5.3 分类准备	3
5.4 分类实施	4
5.5 结果评估	5
5.6 维护改进	5
6 分类视角	6
6.1 概述	6
6.2 技术选型视角	6
6.3 业务应用视角	6
6.4 安全隐私保护视角	6
7 分类维度	6
7.1 概述	6
7.2 技术选型维度	7
7.3 业务应用维度	9
7.4 安全隐私保护维度	12
8 分类方法	12
8.1 线分类法	12
8.2 面分类法	13
8.3 混合分类法	13
附录 A (资料性附录) 大数据分类示例	14

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本标准起草单位:中国科学院信息工程研究所(信息安全国家重点实验室)、国家信息中心、浪潮软件集团有限公司、智慧神州(北京)科技有限公司、方正国际软件(北京)有限公司、国网安徽省电力有限公司(电力科学研究院)、中国铁道科学研究院集团有限公司、中国电子技术标准化研究院、上海二零卫士信息安全有限公司、联通大数据有限公司、中国保险信息技术管理有限责任公司、九次方大数据信息集团有限公司、中电长城网际系统应用有限公司、广东电网有限责任公司信息中心、中电科大数据研究院有限公司、北京大学、山东省计算中心(国家超级计算济南中心)。

本标准主要起草人:陈驰、马红霞、马书南、田雪、高亚楠、黄先芝、单震、张慧敏、张煜、顾广宇、吴艳华、郑金子、尹卓、叶林、干露、关泰璐、李燕超、郎佩佩、闵京华、魏理豪、禄凯、张吉才、冯念慈、赵俊峰、史丛丛、孙嘉阳。

信息技术 大数据 数据分类指南

1 范围

本标准提供了大数据分类过程及其分类视角、分类维度和分类方法等方面的建议和指导。
本标准适用于指导大数据分类。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 4754—2017 国民经济行业分类

GB/T 35295—2017 信息技术 大数据 术语

3 术语和定义

GB/T 35295—2017 界定的以及下列术语和定义适用于本文件。为了便于使用,以下重复列出了GB/T 35295—2017 中的某些术语和定义。

3.1

大数据 big data

具有体量巨大、来源多样、生成极快、且多变等特征,并且难以用传统数据体系结构有效处理的包含大量数据集的数据。

注:国际上,大数据的4个特征普遍不加修饰地直接用 volume、variety、velocity 和 variability 予以表述,并分别赋予了它们在大数据语境下的定义:

- a) 体量 volume:构成大数据的数据集的规模。
- b) 多样性 variety:数据可能来自多个数据仓库、数据领域或多种数据类型。
- c) 速度 velocity:单位时间的数据流量。
- d) 多变性 variability:大数据其他特征,即体量、速度和多样性等特征都处于多变状态。

[GB/T 35295—2017,定义 2.1.1]

3.2

数据集 data set

数据记录汇聚的数据形式。

注:它可以具有大数据的体量、速度、多样性和易变性特征。数据集的特征表征的是数据本身或静态数据,而数据的特征,当其在网络上传输时或暂时驻留于计算机存储器中以备读出或更新时,表征的是动态数据。

[GB/T 35295—2017,定义 2.1.46]

3.3

大数据分类 big data classification

根据大数据的属性或特征,将其按一定的原则和方法进行区分和归类,并建立起一定的分类体系和排列顺序的过程。