



# 中华人民共和国国家标准

GB/T 32630—2016

---

## 非结构化数据管理系统技术要求

Technical requirements for unstructured data management system

2016-04-25 发布

2016-11-01 实施

---

中华人民共和国国家质量监督检验检疫总局  
中国国家标准化管理委员会 发布

# 目 次

前言 .....	I
1 范围 .....	1
2 符合性 .....	1
3 规范性引用文件 .....	1
4 术语和定义 .....	1
5 缩略语 .....	1
6 功能性要求 .....	2
6.1 总体要求 .....	2
6.2 存储与计算设施 .....	2
6.3 存储管理 .....	2
6.4 特征抽取 .....	3
6.5 索引管理 .....	4
6.6 查询处理 .....	4
6.7 访问接口 .....	4
6.8 管理工具 .....	5
6.9 转换加载 .....	5
6.10 分析挖掘 .....	5
6.11 可视展现 .....	5
6.12 统计信息 .....	6
7 质量要求 .....	6
7.1 信息安全性 .....	6
7.2 易用性 .....	7
7.3 维护性 .....	7
7.4 兼容性 .....	7

## 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：浙江大学、中国电子技术标准化研究院、清华大学、中国人民大学、北京航空航天大学。

本标准主要起草人：庄越挺、陈岭、邵健、孙建伶、高林、吴新松、刘贤刚、李莹、卫凤林、王建民、王朝坤、陈俊、杜小勇、陈晋川、马殿富、郎波。

# 非结构化数据管理系统技术要求

## 1 范围

本标准规定了非结构化数据管理系统的功能性要求和质量要求。  
本标准适用于非结构化数据管理系统产品的研制、开发和测试。

## 2 符合性

对于非结构化数据管理系统是否符合本标准的规定如下：

- a) 非结构化数据管理系统若满足本标准基本要求中的所有要求,则称其满足本标准的基本要求;
- b) 非结构化数据管理系统在满足所有基本要求的前提下,若满足某部分扩展要求,则称其满足本标准的基本要求和该部分扩展要求;
- c) 非结构化数据管理系统若满足本标准基本要求和扩展要求中的所有要求,则称其满足本标准的所有要求。

## 3 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 18030—2005 信息技术 中文编码字符集

GB/T 32908—2016 非结构化数据访问接口规范

## 4 术语和定义

下列术语和定义适用于本文件。

### 4.1

**非结构化数据 unstructured data**

没有明确结构约束的数据,如文本、图像、音频、视频等。

### 4.2

**非结构化数据管理系统 unstructured data management system**

对非结构化数据进行管理、操作的大型基础软件,提供非结构化数据存储、特征抽取、索引、查询等管理功能。

## 5 缩略语

下列缩略语适用于本文件。

IDF:逆向文件频率(Inverse Document Frequency)

MFCC:梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient)

PB:千万亿字节(Peta Byte)

SIFT:尺度不变特征转换(Scale-invariant Feature Transform)

TF:词频(Term Frequency)

## 6 功能性要求

### 6.1 总体要求

非结构化数据管理系统的总体要求如下:

- a) 应包括存储与计算设施、存储管理、特征抽取、索引管理、查询处理、访问接口、管理工具七个基本组成部分;
- b) 宜包括转换加载、分析挖掘、可视展现三个扩展组成部分。

### 6.2 存储与计算设施

#### 6.2.1 基本要求

存储与计算设施基本要求如下:

- a) 应支持磁盘、磁盘阵列、内存存储、键值存储、关系型存储、分布式文件系统等一种或多种存储设施;
- b) 应支持单机、并行计算集群、分布式计算集群等一种或多种计算设施。

#### 6.2.2 扩展要求

无。

### 6.3 存储管理

#### 6.3.1 基本要求

存储管理基本要求如下:

- a) 应提供涵盖原始数据、基本属性、底层特征、语义特征的概念层存储建模功能;
- b) 应提供逻辑层的存储建模功能;
- c) 支持整型、浮点型、布尔型、字符串、日期、日期时间、二进制块等基本数据类型;
- d) 支持向量、矩阵、关联等数据类型;
- e) 应支持根据建好的逻辑层存储模型创建存储实例;
- f) 应支持在创建好的存储实例上插入、修改、删除非结构化数据;
- g) 应支持删除存储实例;
- h) 应支持非结构化数据操作的原子性。

#### 6.3.2 扩展要求

存储管理扩展要求如下:

- a) 应支持全局事务的定义并保证事务的原子性、一致性、隔离性和持久性;
- b) 应支持数据类型的多值结构和层次结构;
- c) 应支持在不同的存储设施上创建存储实例并实现自动映射;
- d) 应支持 PB 级数据存储。

## 6.4 特征抽取

### 6.4.1 基本要求

#### 6.4.1.1 特征抽取总体要求

特征抽取基本要求的总体要求如下：

- a) 应支持从非结构化数据中抽取名称、类型、创建者和创建时间等基本属性；
- b) 应支持文本、图像、音频、视频中至少一种非结构化数据特征的抽取。

#### 6.4.1.2 文本特征抽取

文本特征抽取基本要求如下：

- a) 应支持从文本中抽取词；
- b) 应支持从文本中抽取停用词；
- c) 应支持从文本中抽取 TF-IDF 特征；
- d) 应支持从文本中抽取关键词。

#### 6.4.1.3 图像特征抽取

图像特征抽取基本要求如下：

- a) 应支持从图像中抽取颜色空间、主导颜色、色彩内容、颜色结构、颜色布局等颜色特征；
- b) 应支持从图像中抽取均匀纹理、边缘直方图等纹理特征；
- c) 应支持从图像中抽取区域形状、轮廓形状等形状特征。

#### 6.4.1.4 音频特征抽取

音频特征抽取基本要求如下：

- a) 应支持从音频中按静音抽取片段；
- b) 应支持从音频中抽取音频波谱特征；
- c) 应支持从音频中抽取 MFCC 特征。

#### 6.4.1.5 视频特征抽取

视频特征抽取基本要求如下：

- a) 应支持从视频中按镜头抽取片段；
- b) 应支持从视频中抽取关键帧。

### 6.4.2 扩展要求

#### 6.4.2.1 特征抽取总体要求

应支持文本、图像、音频、视频中两种及以上非结构化数据特征的抽取。

#### 6.4.2.2 文本特征抽取

文本特征抽取扩展要求如下：

- a) 应支持从文本中抽取人名、地名、机构名、时间等内容；
- b) 应支持从文本中抽取主题词；
- c) 应支持从文本中抽取摘要；

- d) 应支持从文本中抽取语义特征。

#### 6.4.2.3 图像特征抽取

图像特征抽取扩展要求如下：

- a) 应支持从图像中抽取 SIFT 局部区域描述特征；
- b) 应支持从图像中抽取语义特征。

#### 6.4.2.4 音频特征抽取

音频特征抽取扩展要求如下：

- a) 应支持从音频中抽取音频指纹特征；
- b) 应支持从音频中抽取旋律特征；
- c) 应支持从音频中抽取音色和响度特征；
- d) 应支持从音频中抽取语义特征。

#### 6.4.2.5 视频特征抽取

视频特征抽取扩展要求如下：

- a) 应支持从视频中抽取运动特征；
- b) 应支持从视频中抽取语义特征。

### 6.5 索引管理

#### 6.5.1 基本要求

应提供索引建立、维护和删除功能。

#### 6.5.2 扩展要求

索引管理扩展要求如下：

- a) 应支持 B+树索引、倒排索引、全文索引和高维索引多种索引类型；
- b) 应支持 PB 级数据索引。

### 6.6 查询处理

#### 6.6.1 基本要求

查询处理基本要求如下：

- a) 应支持范围查询、全文查询、样例查询、语义查询和跨媒体查询多种查询方式；
- b) 应提供查询结果排序和批量返回功能。

#### 6.6.2 扩展要求

查询处理扩展要求如下：

- a) 应提供查询相关性反馈功能；
- b) 应支持 PB 级数据查询。

### 6.7 访问接口

#### 6.7.1 基本要求

访问接口基本要求如下：

- a) 应依从 GB/T 32908—2016 中第 4 章规定的查询语言访问接口要求；
- b) 应依从 GB/T 32908—2016 中第 5 章规定的应用程序访问接口要求。

## 6.7.2 扩展要求

应依从 GB/T 32908—2016 中第 6 章规定的 Web 服务访问接口要求。

## 6.8 管理工具

### 6.8.1 基本要求

管理工具基本要求如下：

- a) 应提供安装部署工具；
- b) 应提供逻辑层模型定义工具；
- c) 应提供查询分析工具。

### 6.8.2 扩展要求

应提供任务管理工具。

## 6.9 转换加载

### 6.9.1 基本要求

转换加载基本要求如下：

- a) 应支持常用格式的文本数据转换加载；
- b) 应支持常用格式的图像数据转换加载；
- c) 应支持常用格式的音频数据转换加载；
- d) 应支持常用格式的视频数据转换加载。

### 6.9.2 扩展要求

应提供面向医疗、工业控制、金融等特定应用领域的非结构化数据转换加载工具集。

## 6.10 分析挖掘

### 6.10.1 基本要求

分析挖掘基本要求如下：

- a) 应支持分析模型定义；
- b) 应支持分析模型执行；
- c) 应支持挖掘模型定义；
- d) 应支持挖掘模型执行。

### 6.10.2 扩展要求

应提供数据挖掘工具集。

## 6.11 可视展现

### 6.11.1 基本要求

可视展现基本要求如下：



- a) 应支持文本数据展现；
- b) 应支持图像数据展现；
- c) 应支持音频数据展现；
- d) 应支持视频数据展现；
- e) 应提供柱状图可视化方法；
- f) 应提供折线图可视化方法；
- g) 应提供饼图可视化方法。

#### 6.11.2 扩展要求

可视展现扩展要求如下：

- a) 应提供标签云可视化方法；
- b) 应提供关系图可视化方法；
- c) 应提供基于地图的可视化方法。

### 6.12 统计信息

#### 6.12.1 时间统计信息

##### 6.12.1.1 基本要求

应提供系统平均数据插入时间的统计信息。

##### 6.12.1.2 扩展要求

应提供系统平均查询响应时间的统计信息。

#### 6.12.2 容量统计信息

##### 6.12.2.1 基本要求

应提供系统可支持的数据量的统计信息。

##### 6.12.2.2 扩展要求

应提供系统可支持的并发用户数的统计信息。

## 7 质量要求

### 7.1 信息安全性

#### 7.1.1 基本要求

信息安全性基本要求如下：

- a) 应支持创建、删除用户；
- b) 应支持用户设置密码；
- c) 应支持创建、删除角色；
- d) 应支持用户角色的授予、收回、查看；
- e) 应提供权限体系；
- f) 应支持用户和角色权限的授予、收回、查看；
- g) 应支持数据加密。

### 7.1.2 扩展要求

应支持用户审计。

## 7.2 易用性

### 7.2.1 基本要求

易用性基本要求如下：

- a) 应提供完整的用户手册；
- b) 应提供联机帮助；
- c) 应提供图形化管理界面；
- d) 应提供模型定义和数据操作的交互工具。

### 7.2.2 扩展要求

应提供向导工具。

## 7.3 维护性

### 7.3.1 基本要求

维护性基本要求如下：

- a) 应支持非结构化数据的导入与导出；
- b) 应支持数据多副本；
- c) 应支持日志机制；
- d) 应支持存储实例的备份与恢复；
- e) 应提供故障恢复工具。

### 7.3.2 扩展要求

应支持系统模块的热插拔。

## 7.4 兼容性

### 7.4.1 基本要求

兼容性基本要求如下：

- a) 应支持多种操作系统运行环境；
- b) 应支持 GB 18030—2005 的强制部分；
- c) 应支持 C++ 或 Java 主流编程语言。

### 7.4.2 扩展要求

无。

---