

中华人民共和国国家标准

GB/T XXXXX—XXXX

人工智能 知识图谱应用系统 第2部分：性能要求与测试方法

Artificial intelligence - Knowledge graph application system - Part 2: Performance requirements and test methods

(点击此处添加与国际标准一致性程度的标识)

(征求意见稿)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX - XX - XX 发布

XXXX - XX - XX 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	III
引 言	IV
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 概述	2
6 知识图谱构建性能要求	3
6.1 知识表示	3
6.2 知识建模	3
6.3 知识获取	3
6.4 知识存储	3
6.5 知识融合	4
6.6 知识计算	4
6.7 知识溯源	4
7 知识图谱应用性能要求	4
7.1 响应性	4
7.2 可移植性	4
7.3 可靠性	4
7.3.1 可用性	4
7.3.2 成熟性	5
8 测试环境	5
9 知识图谱构建性能测试方法	5
9.1 知识表示	5
9.1.1 知识表示响应时间	5
9.1.2 知识表示响应速度	5
9.1.3 最大呈现节点数量	6
9.2 知识建模	6
9.2.1 本体模型的明确度	6
9.2.2 本体模型的完整性	6
9.2.3 本体模型的一致度	6
9.2.4 本体模型的最大单调性	7
9.3 知识获取	7
9.3.1 结构化数据	7
9.3.2 半结构化或非结构化数据	7

9.4 知识存储	8
9.4.1 字符集支持度	8
9.4.2 数据加载时间	8
9.4.3 K 跳邻居查询响应时间	8
9.4.4 K 跳邻居查询支持的最大并发数	9
9.4.5 吞吐率	9
9.4.6 QPS (单条指令)	10
9.4.7 QPS (多条指令)	10
9.5 知识融合	10
9.6 知识计算	11
9.7 知识溯源	12
10 知识图谱应用性能测试方法	12
10.1 响应性	12
10.1.1 知识图谱体量	12
10.1.2 知识图谱复杂度	12
10.1.3 效率	13
10.1.4 准确率	13
10.1.5 实时性	13
10.2 可移植性	14
10.3 可靠性	14
10.3.1 可用性	14
10.3.2 成熟性	14
附录 A (资料性) 知识图谱应用系统的测试流程	16
参 考 文 献	17

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由全国信息技术标准化技术委员会(SAC/TC28)提出并归口。

本文件起草单位：

本文件主要起草人：

引 言

目前，知识图谱在金融证券、生物医药、制造业、交通、教育、农业、电信、电商、出版等行业已有非常丰富的应用场景。知识图谱应用系统作为知识图谱在企业或机构落地的具体载体，支撑企业和行业内知识的智能化积累、抽取、挖掘、流通与管理，并提供企业所需的知识服务及系统管理等功能。本系列标准拟由以下部分构成：

- 第1部分：功能要求。目的在于提供知识图谱应用系统的系统架构与功能要求。
- 第2部分，性能要求与测试方法。目的在于提供知识图谱应用系统的性能指标体系及对应测试方法。
- 第3部分，知识服务。目的在于提供基于知识图谱的知识问答、知识检索、知识推荐等知识服务要求。
- 第4部分，知识管理。目的在于提供知识图谱应用系统的知识管理要求。

人工智能 知识图谱应用系统 第2部分：性能要求与测试方法

1 范围

本文件规定了知识图谱应用系统的性能要求、测试环境、测试方法等。
本文件适用于知识图谱应用系统的设计、开发及性能测试。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；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 42131-2022 人工智能 知识图谱技术框架

GB/T AAAAA-AAAA 人工智能 知识图谱应用系统 第1部分：功能要求

3 术语和定义

下列术语和定义适用于本文件。

3.1

知识 knowledge

通过学习、实践或探索所获得的认识、判断或技能。

[来源：GB/T 23703.2-2010, 2.1]

3.2

知识图谱 knowledge graph

以结构化形式描述的知识元素及其联系的集合。

[来源：GB/T 42131-2022, 2.1]

3.3

知识单元 knowledge unit

按照一定关系组织的一组知识元素的集合。

[来源：GB/T 42131-2022, 2.1]

3.4

知识图谱应用系统 knowledge graph application system

整合知识图谱和业务实现逻辑，并具备知识图谱构建和应用完整信息流与预设功能的系统。

3.5

知识获取的准确率 precision of knowledge acquisition

已获取知识中正确知识的占比。

3.6

GB/T □XXXXX-XXXX

知识获取的召回率 recall of knowledge acquisition

已获得知识对正确知识覆盖的程度。

3.7

K跳邻居查询响应时间 response time of K nearest neighbour query

系统完成1跳、2跳、K跳等不同展开相连接层级中关联对象的查询时间。

3.8

毫秒级数据查询响应率 query response time less than second

查询操作过程中响应在毫秒级以下的次数占比。

3.9

吞吐率 throughput rate

系统在特定并发数下单位时间内处理的请求数量。

3.10

知识图谱体量 knowledge graph volume

知识图谱中包含的实体和关系数量。

3.11

知识图谱复杂度 knowledge graph complexity

知识图谱中包含的实体类型和关系类型数量。

3.12

平均故障时间 mean time to failure

在故障发生时系统处于不可用状态的时长。

3.13

平均失效间隔时间 mean time between failures

系统在运行中失败间隔的时长。

3.14

故障率 failure rate

在规定时间内平均故障数。

4 缩略语

下列缩略语适用于本文件。

HTTP: 超文本传输协议 (Hyper Text Transfer Protocol)

QPS: 每秒查询率 (Queries-per-second)

RPC: 远程过程调用 (Remote Procedure Call)

5 概述

知识图谱是以结构化形式描述的知识元素及其联系的集合。根据GB/T 42131-2022、GB/T AAAAA-AAAA可知，知识图谱应用系统根据功能构成可划分为：知识表示、知识建模、知识获取、知识存储、知识计算等知识图谱构建相关模块；知识应用、知识维护等知识图谱应用相关模块。知识图谱应用系统的知识图谱构建性能主要关注对知识图谱构建过程中各项活动的支持水平；知识图谱应用系统的安全性、可靠性、易用性等特性主要关注知识图谱应用系统在部署和应用过程中的整体性能。本文件仅聚焦知识图谱构建和知识图谱应用相关指标，基础通用软件质量评价指标及其度量方法可参考GB/T 39788-2021、GB/T 25000.20-2021、GB/T 29831.1-2013、GB/T 29831.2-2013等系统与软件工程相关标准中规定。

6 知识图谱构建性能要求

6.1 知识表示

- 6.1.1 知识表示响应时间应不大于 3s。
- 6.1.2 知识表示响应速度应每秒不小于 100 个图形化的节点或关系。
- 6.1.3 系统可呈现的最大节点数量应不小于 200 个。

6.2 知识建模

- 6.2.1 系统中构建的本体模型应对其中包含的术语给出明确和客观语义定义，本体模型的明确度不小于 80%。
- 6.2.2 系统中构建的本体模型应完全表达所描述领域或业务场景内术语的含义，本体模型的完整度不小于 80%。
- 6.2.3 系统中构建的本体模型应正确一致地展示对象和信息，且由术语得出的推理或计算结果不会与术语本身含义产生矛盾，本体模型的一致度不小于 90%。
- 6.2.4 系统中构建的本体模型的术语定义不应存在歧义、重叠或交叉，本体模型的最大单调性不小于 80%。

6.3 知识获取

- 6.3.1 面向结构化数据，知识获取的精确率应不小于 95%；面向半结构化数据和非结构化数据，知识获取的准确率应不小于 75%。
- 6.3.2 面向结构化数据，知识获取的召回率应不小于 95%；面向半结构化数据和非结构化数据，知识获取的召回率应不小于 75%。
- 6.3.3 面向结构化数据，知识获取的 F1 测量值应不小于 95%；面向半结构化数据和非结构化数据，知识获取的 F1 测量值应不小于 75%。

6.4 知识存储

- 6.4.1 知识存储模块支持的字符集类型总数应不小于 10 种。
- 6.4.2 知识存储模块完成基准数据集加载的时间应不大于 10s。
- 6.4.3 知识存储模块完成 1 跳邻居查询的时间应不大于 0.1s；完成 2 跳邻居查询的时间应不大于 0.5s；完成 3 跳邻居查询的时间应不大于 2s。
- 6.4.4 K 跳邻居查询支持的最大并发数应不小于 20 个，用于度量系统能够承受的同时受到的 K 跳查询请求数的极限，超过该请求数，将导致系统效率严重下滑并可能导致系统失效。
- 6.4.5 毫秒级数据查询响应率应不小于 90%。

6.4.6 在 20 并发数下，一度邻居的吞吐率不低于 10。

6.4.7 QPS（单条指令）用于度量系统在每秒能够响应的单条指令查询次数，应大于等于 100 条。

6.4.8 QPS（多条指令）用于度量系统在每秒能够响应的多条指令查询次数，应大于等于 50 条。

6.4.9 数据加载存储空间应不低于 500GB。

6.5 知识融合

6.5.1 知识融合的精确率用于度量知识融合结果正确知识与总融合知识规模的比例，应不小于 80%。

6.5.2 知识融合的召回率用于度量知识融合结果中已融合知识对正确知识覆盖的程度，应不小于 80%。

6.5.3 知识融合的 F1 测量值用于综合度量知识融合结果的精确性和完整性，应不小于 80%。

6.6 知识计算

6.6.1 知识计算任务的响应时间用于度量知识计算任务启动到结束所耗费的时间，应不大于 10s。

6.6.2 知识计算任务的均方根误差用于度量知识计算任务的准确程度（适用于每个实体都存在计算结果的算法），应不大于 0.1。

6.6.3 知识计算任务的准确率用于度量知识计算结果与预设目标值一致的程度，应不大于 90%。

6.6.4 知识计算模块的逻辑语言丰富度用于度量可支持的逻辑语言的数量，应不小于 3 种。

6.6.5 知识计算模块的规则引擎丰富度用于度量可支持的规则引擎的数量，应不小于 3 个。

6.6.6 知识计算模块的软硬件资源消耗度用于度量计算任务执行过程中耗费的最大软硬件资源，内存应不超过 512GB。

6.7 知识溯源

6.7.1 知识溯源的覆盖率用于度量知识图谱中知识能够溯源的覆盖情况，应不小于 20%。

6.7.2 知识溯源的正确率用于度量知识计算和融合结果的可解释性溯源准确率，应不小于 85%。

7 知识图谱应用性能要求

7.1 响应性

7.1.1 知识图谱体量应不小于 10 万个。

7.1.2 知识图谱复杂度应不小于 10 个。

7.1.3 效率用于度量在规定时间内处理请求的数量，应不小于 5 个/s。

7.1.4 准确率用于度量符合请求的结果占返回结果总数的比例，应不小于 90%。

7.1.5 实时性用于度量知识图谱应用系统在限定的时间内对用户发送的请求的响应比例，应不小于 95%。

7.2 可移植性

7.2.1 平均安装时间应不大于 12h。

7.2.2 安装成功率应不小于 70%。

7.3 可靠性

7.3.1 可用性

7.3.1.1 平均故障时间应不大于 24h。

7.3.1.2 访问权限的控制数量应不小于 20 个。

7.3.2 成熟性

- 7.3.2.1 故障排除率应不小于 99%。
- 7.3.2.2 平均失效间隔时间应不小于 180 天。
- 7.3.2.3 故障率在 3 个月内应不大于 3 次。

8 测试环境

除另有规定外，知识图谱应用系统的测试环境应满足以下要求：

- a) 具备知识图谱应用系统专用的测试工具，且部署了与待测系统相适应的测试场景；
- a) 测试场所的环境温度：15℃~35℃；
- b) 测试场所的相对湿度：25%~75%；
- c) 操作系统：基于 x86 架构的操作系统，如：Ubuntu 16.04 系统、CentOS 7.8 系统、麒麟系统等；
- d) 测试服务器 CPU：大于等于 32 核；
- e) 测试服务器内存：大于等于 256 GB；
- f) 测试服务器硬盘：大于等于 1 TB。

知识图谱应用系统的测试流程参见附录A。

9 知识图谱构建性能测试方法

9.1 知识表示

9.1.1 知识表示响应时间

知识表示响应时间的测试步骤包括：

- a) 打开知识表示相关模块；
- b) 设置拟呈现的实体或关系的数量；
- c) 在用户触发渲染操作的瞬间开始计时，统计发出请求的时间 t_a ；
- d) 等待渲染过程完成，定义明确的渲染结束标准，并统计结束时间 t_r ；
- e) 按公式（1）计算知识表示响应速度 T_p ：

$$T_p = t_r - t_a \quad (1)$$

- f) 判断 T_p 是否符合 6.1.1 的要求。

9.1.2 知识表示响应速度

知识表示响应速度的测试步骤包括：

- a) 打开知识表示相关模块；
- b) 设置拟呈现的实体或关系的数量 N_p ；
- c) 在用户触发渲染操作的瞬间开始计时，统计发出请求的时间；
- d) 等待渲染过程完成，定义明确的渲染结束标准，并统计结束时间；
- e) 使用结束时间减去开始时间计算知识表示响应时间 T_p ；
- f) 按公式（2）计算知识表示响应速度 X_p ：

$$X_p = \frac{N_p}{T_p} \quad (2)$$

g) 判断 X_p 是否符合 6.1.2 的要求。

9.1.3 最大呈现节点数量

最大呈现节点数量的测试步骤包括：

- a) 打开知识表示相关模块；
- b) 设置拟呈现的实体或关系的数量；
- c) 统计呈现结束时的节点数量；
- d) 递增拟呈现的实体或关系的数量；
- e) 按照 b) 继续执行，直至呈现结束时节点数量无法继续增加；
- f) 统计最终呈现的节点数量；
- g) 判断终呈现的节点数量是否符合 6.1.3 的要求。

9.2 知识建模

9.2.1 本体模型的明确度

该特性的测试步骤包括：

- a) 打开知识模块相关模块；
- b) 随机抽取 B_{O_C} 个构建后本体模型中的实体类型、关系或属性；
- c) 筛选出对应的原始数据；
- d) 确定被抽取结果中定义清晰的实体类型、关系或属性数量 A_{O_C} ；
- e) 按公式 (3) 计算本体模型的明确度：

$$X_{O_C} = \frac{A_{O_C}}{B_{O_C}} \quad (3)$$

- f) 判断本体模型的明确度是否符合 6.2.1 的要求。

9.2.2 本体模型的完整度

该特性的测试步骤包括：

- a) 打开知识模块相关模块；
- b) 随机抽取 B_{O_I} 个构建后本体模型中的实体类型、关系或属性；
- c) c) 出对应的原始数据；
- d) 确定被抽取结果中完全表达所描述领域内术语含义的实体类型、关系或属性数量 A_{O_I} ；
- e) 按公式 (4) 计算本体模型的完整度：

$$X_{O_I} = \frac{A_{O_I}}{B_{O_I}} \quad (4)$$

- f) 判断本体模型的完整度是否符合 6.2.2 的要求。

9.2.3 本体模型的一致度

该特性的测试步骤包括：

- a) 打开知识模块相关模块；
- b) 随机抽取 B_{O_U} 个构建后本体模型中的实体类型、关系或属性；
- c) 筛选出对应的原始数据；
- d) 确定被抽取结果中能正确一致展示对象和信息的实体类型、关系或属性数量 A_{O_U} ；
- e) 按公式 (5) 计算本体模型的一致度：

$$X_{O,U} = \frac{A_{O,U}}{B_{O,U}} \quad (5)$$

f) 判断本体模型的一致度是否符合 6.2.3 的要求。

9.2.4 本体模型的最大单调性

该特性的测试步骤包括：

- a) 打开知识模块相关模块；
- b) 随机抽取 $B_{O,M}$ 个构建后本体模型中的实体类型、关系或属性；
- c) 筛选出对应的原始数据；
- d) 确定被抽取结果中存在定义交叉或重叠的实体类型、关系或属性数量 $A_{O,M}$ ；
- e) 按公式（6）计算本体模型的最大单调性：

$$X_{O,M} = \frac{A_{O,M}}{B_{O,M}} \quad (6)$$

f) 判断本体模型的最大单调性是否符合 6.2.4 的要求。

9.3 知识获取

9.3.1 结构化数据

知识获取相关模块性能的测试步骤包括：

- a) 打开知识获取相关模块；
- b) 设置知识获取的参数；
- c) 统计发出指令的开始时间；
- d) 统计发出指令的结束时间；
- e) 统计获取并与真实相符的实体、关系或属性数量 TP_A ；
- f) 统计获取但与真实不符的实体、关系或属性数量 FP_A ；
- g) 统计被标注为真实但未识别的实体、关系或属性数量 FN_A ；
- h) 按公式（7）计算知识获取的准确率：

$$Precision_A = \frac{TP_A}{TP_A + FP_A} \quad (7)$$

i) 按公式（8）计算知识获取的召回率：

$$Recall_A = \frac{TP_A}{TP_A + FN_A} \quad (8)$$

j) 按公式（9）计算知识获取的 F1 测量值：

$$F1-score_A = 2 \times \frac{Precision_A \times Recall_A}{Precision_A + Recall_A} \quad (9)$$

k) 判断面向结构化数据的知识获取准确率、召回率和 F1 测量值是否符合 6.3 的要求。

9.3.2 半结构化或非结构化数据

知识获取相关模块性能的测试步骤包括：

- a) 打开知识获取相关模块；
- b) 设置知识获取的参数；
- c) 统计发出指令的开始时间；
- d) 统计发出指令的结束时间；

- e) 筛选出对应的原始半结构化数据或非结构化数据;
- f) 统计被识别并与真实相符的实体、关系或属性数量 TP_A ;
- g) 统计被识别但与真实不符的实体、关系或属性数量 FP_A ;
- h) 统计被标注为真实但未识别的实体、关系或属性数量 FN_A ;
- i) 统计面向半结构化数据或非结构化数据的精确率、召回率和 F1 测量值;
- j) 判断面向半结构化数据或非结构化数据的知识获取准确率、召回率和 F1 测量值是否符合 6.3 的要求。

9.4 知识存储

9.4.1 字符集支持度

字符集支持度的测试步骤包括:

- a) 打开知识存储相关模块;
- b) 统计在数据库管理系统中被描述的导入数据的字符集类型;
- c) 判断字符集支持度是否符合 6.4.1 的要求。

9.4.2 数据加载时间

数据加载时间的测试步骤包括:

- a) 打开知识存储相关模块;
- b) 准备拟加载的数据,并计算数据量;
- c) 统计发出指令的时间;
- d) 统计加载结束的时间;
- e) 按公式(10)计算数据加载时间:

$$X_i = Rec_i - Send_i(i \sim n) \tag{10}$$

式中:

X_i ——第 i 次加载的响应时间;

Rec_i ——第 i 次加载收到响应结果的时间;

$Send_i$ ——第 i 次发出加载数据请求的时间;

n ——测试的总数。

- f) 执行 n 次 c) 至 e), 并计算平均值;
- g) 判断数据加载时间是否符合 6.4.2 的要求。

9.4.3 K 跳邻居查询响应时间

K 跳邻居查询响应时间的测试步骤包括:

- a) 打开知识存储相关模块;
- b) 设置 K 跳查询参数;
- c) 统计发出指令的时间 $Send_i(i \sim n)$;
- d) 统计查询结束的时间 Rec_i ;
- e) 计算查询响应时间 X_i ;
- f) 执行 N 次 c) 至 e), 并按公式(11)、公式(12)计算平均值。

$$X_i = Rec_i - Send_i(i \sim n) \tag{11}$$

$$Y_i = (X_i - X_T) (i \sim n, X_i > X_T) \dots\dots\dots \tag{12}$$

式中：

X_i ——第 i 次 K 跳查询的响应时间；

Rec_i ——第 i 次 K 跳查询收到响应结果的时间；

$Send_i$ ——第 i 次发出 K 跳查询请求的时间；

Y_i ——第 i 次 K 跳查询的超时时间；

X_T ——K 跳查询的超时时间阈值；

n ——测试的总数。

g) 判断数据加载时间是否符合 6.4.3 的要求。

9.4.4 K 跳邻居查询支持的最大并发数

K 跳邻居查询支持的最大并发数的测试步骤包括：

- a) 打开知识存储相关模块；
- b) 设置并发请求数序列；
- c) 针对每个需要尝试的并发请求数执行并发测试；
- d) 针对每个请求，测试并发请求的响应时间或吞吐率；
- e) 监控测试结果，如果出现以下情况，可停止测试：
 - 1) 测试过程出现异常，测试无法正常完成；
 - 2) 响应时间变得不可测试。
- f) 如果测试完仍然未得到最大并发请求数，增加尝试次数或调整尝试范围与递增量；
- g) 统计在测试停止时，K 跳邻居查询支持的最大并发数；
- h) 判断 K 跳邻居查询支持的最大并发数是否符合 6.4.4 的要求。

9.4.5 吞吐率

测试步骤包括：

- a) 确定点样本、边样本；
- b) 打开知识存储相关模块；
- c) 设置并发请求线程数、总操作数；
- d) 设置并发操作序列；
- e) 每个线程循环执行操作序列；
- f) 监控测试结果，当累计操作数等于设置的总操作数则停止测试；
- g) 针对每个线程的每个请求，验证结果的正确性和记录请求的响应时长；
- h) 按公式 (13) 计算毫秒级数据查询响应率：

$$X_R = \frac{A_R}{B_R} \quad (13)$$

式中：

A_R ——查询响应时间在毫秒级别以下的次数；

B_R ——图谱查询的总次数。

i) 按公式 (14) 计算吞吐率：

$$RPS = \frac{CR_g}{T_g} \quad (14)$$

式中：

CR_g ——总请求数；

T_g ——处理这些请求的总完成时间。

j) 还可依次执行如下压力测试，并统计吞吐率：

- 1) 执行查询点、查询边压力测试；
- 2) 执行增加点、增加边压力测试；
- 3) 执行修改点、修改边压力测试；
- 4) 执行删除点、删除边压力测试。

k) 判断毫秒级数据查询响应率、吞吐率是否符合 6.4.5、6.4.6 的要求。

9.4.6 QPS（单条指令）

QPS（单条指令）的测试步骤包括：

- a) 打开知识存储相关模块；
- b) 设置并发请求线程数、持续时间；
- c) 设置单条指令查询参数；
- d) 每个线程循环执行设置的单条指令；
- e) 监控测试时间，当测试时间等于 b) 设置的持续时间则停止测试；
- f) 针对每个线程的每个请求，验证结果的正确性和记录请求的响应时长；
- g) 统计 QPS，并判断是否符合 6.4.7 的要求。

9.4.7 QPS（多条指令）

QPS（多条指令）的测试步骤包括：

- a) 打开知识存储相关模块；
- b) 设置并发请求线程数、持续时间；
- c) 设置多条指令查询参数序列；
- d) 每个线程循环执行多条指令序列；
- e) 监控测试时间，当测试时间等于 b) 设置的持续时间则停止测试；
- f) 针对每个线程的每个请求，验证结果的正确性和记录请求的响应时长；
- g) 统计 QPS，并判断是否符合 6.4.8 的要求。

9.5 知识融合

知识融合相关模块性能的测试步骤包括：

- a) 打开知识表示相关模块。
- b) 准备测试数据：根据测试需求，准备一系列具有代表性、覆盖多种类型和复杂度的知识数据作为测试输入。确保测试数据足以评估知识融合模块在不同负载下的处理速度。
- c) 设置测试参数：在测试界面中，配置测试参数，包括数据输入速率（每秒发送的数据量）、测试时长、并发用户数（如适用）等。确保测试参数的设置能够准确反映系统在实际应用中的使用情况。
- d) 启动测试并记录时间：启动测试程序，开始发送测试数据到知识融合模块进行处理。同时，记录测试开始的精确时间作为基准时间点。
- e) 统计处理结束时间：当达到预设的测试时长或系统处理完所有输入数据时，记录此时的精确时间作为测试结束时间。

- f) 计算处理速度：根据测试期间系统处理的数据总量以及测试时长，计算出平均处理速度（如每秒处理的数据量）。
- g) 统计融合结果中与真实相符的实体、关系或属性数量 TP_F 。
- h) 统计融合结果中已融合但与真实不符的实体、关系或属性数量 FP_F 。
- i) 统计融合结果中待融合但未融合的实体、关系或属性数量 FN_F 。
- j) 按公式（15）计算知识融合的精确率：

$$Precision_F = \frac{TP_F}{TP_F + FP_F} \quad (15)$$

- k) 按公式（16）计算知识融合的召回率：

$$Recall_F = \frac{TP_F}{TP_F + FN_F} \quad (16)$$

- l) 按公式（17）计算知识融合的 F1 测量值：

$$F1-score_F = 2 \times \frac{Precision_F \times Recall_F}{Precision_F + Recall_F} \quad (17)$$

- m) 判断知识融合的精确率、召回率、F1 测量值是否符合 6.5 的要求。

9.6 知识计算

知识计算相关模块性能的测试步骤包括：

- 按照说明书运行待测的系统；
- 选定待计算的知识图谱；
- 选定知识计算任务，配置参数，执行计算任务；
- 完成计算任务并记录知识计算任务的启动时间和结束时间；
- 按公式（18）计算知识计算任务的响应时间：

$$T = t_{end} - t_{start} \quad (18)$$

式中：

t_{start} ——知识计算任务启动时间；

t_{end} ——知识计算任务结束时间。

- f) 按公式（19）计算知识计算任务的均方根误差：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \quad (19)$$

式中：

x_i ——第 i 次知识计算结果；

\hat{x}_i ——第 i 次知识计算结果的准确值。

- g) 按公式（20）计算知识计算任务的准确率：

$$Precision = \frac{TP_{KC}}{TP_{KC} + FP_{KC}} \quad (20)$$

式中：

TP_{KC} ——真阳性：知识计算结果中与预设目标值一致的数量；

FP_{KC} ——假阳性：知识计算结果中与预设目标值不一致的数量。

- h) 统计知识计算模块可支持的逻辑语言的数量；

- i) 统计知识计算模块可支持的规则引擎的数量;
- j) 统计计算任务执行过程中耗费的最大软硬件资源;
- k) 判断知识计算任务的响应时间是否符合 6.6 的要求。

9.7 知识溯源

知识溯源相关模块性能的测试步骤包括:

- a) 按照说明书运行待测的系统;
- b) 选定待溯源的知识图谱中的实体或关系;
- c) 选定知识溯源任务, 配置参数, 执行溯源任务;
- d) 完成溯源任务并记录计算时间;
- e) 统计溯源结果中正确的实体或关系等知识数量;
- f) 统计知识图谱中知识计算和融合结果的总数量;
- g) 按公式 (21) 计算知识溯源的覆盖率:

$$C = \frac{N_a}{N} \quad (21)$$

式中:

N_a ——能准确找到溯源的知识数目;

N ——知识总数目。

注: 如覆盖率较低, 说明知识图谱可信度可能较低。

- h) 按公式 (22) 计算知识溯源的正确率:

$$A = \frac{N_{az}}{N_z} \quad (22)$$

式中:

N_{az} ——知识计算和融合结果可解释性溯源正确的个数;

N_z ——知识计算和融合结果总数目。

- i) 判断知识计算任务的响应时间是否符合 6.7 的要求。

10 知识图谱应用性能测试方法

10.1 响应性

10.1.1 知识图谱体量

知识图谱体量的测试步骤包括:

- a) 运行系统, 并选定待测的知识图谱;
- b) 统计知识图谱中的实体数量 V_e ;
- c) 统计知识图谱中的关系数量 V_r ;
- d) 计算实体数量和关系数量的总和;
- e) 判断知识图谱的体量是否符合 7.1.1 的要求。

10.1.2 知识图谱复杂度

知识图谱体量的测试步骤包括:

- a) 运行系统, 并选定待测的知识图谱;

- b) 统计知识图谱中的实体类型数量 F_e ;
- c) 统计知识图谱中的关系类型数量 F_r ;
- d) 计算实体类型数量和关系类型数量的总和;
- e) 判断知识图谱的体量是否符合 7.1.2 的要求。

10.1.3 效率

效率的测试步骤包括:

- a) 运行系统,并选定待测的知识图谱;
- b) 设置查询参数,如 K 跳数等;
- c) 设置并发参数;
- d) 启动查询操作;
- e) 设定 l_{min} ,统计成功处理的请求总数;
- f) 按公式(23)计算系统的效率:

$$E = \frac{A_R}{T} \quad (23)$$

式中:

A_R ——处理的请求总数;

T ——处理请求所用的时间。

- g) 判断系统的效率是否符合 7.1.3 的要求。

10.1.4 准确率

准确率的测试步骤包括:

- a) 运行系统,并选定待测的知识图谱;
- b) 设置查询参数,如 K 跳数等;
- c) 设置并发参数;
- d) 启动查询操作;
- e) 设定 l_{min} ,统计成功处理的请求总数;
- f) 统计被识别并与真实相符的响应数量;
- g) 统计被识别但与真实不符的响应数量
- h) 按公式(24)计算准确率:

$$Precision_R = \frac{TP_R}{TP_R + FP_R} \quad (24)$$

式中:

TP_R ——真阳性:被识别并与真实相符的响应数量;

FP_R ——假阳性:被识别但与真实不符的响应数量。

- i) 判断系统的效率是否符合 7.1.4 的要求。

10.1.5 实时性

实时性的测试步骤包括:

- a) 运行系统,并选定待测的知识图谱;
- b) 设置查询参数,如 K 跳数等;
- c) 设置并发参数;

- d) 启动查询操作;
- e) 设定 t_{min} , 统计在该时间内获得响应的请求总数;
- f) 统计发送的请求总数;
- g) 按公式 (25) 计算系统的实时性:

$$X_A = \frac{A_R}{A} \quad (25)$$

式中:

A_R ——发出请求后在限定时间内响应的次数;

A ——发出请求的总次数。

- h) 判断系统的效率是否符合 7.1.5 的要求。

10.2 可移植性

该特性的测试步骤包括:

- a) 运行系统, 并选定待测的知识图谱;
- b) 在新的目标硬件环境下安装、运行系统, 确定系统可进行安装并能正常启动, 功能可正常运行;
- c) 在新的目标操作系统下安装、运行系统, 确定系统可进行安装并能正常启动, 功能可正常运行;
- d) 检查系统是否支持常见的接口调用, 如: HTTP、RPC 等;
- e) 检查系统是否支持数据的导入导出;
- f) 统计在目标硬件环境、操作系统中完成安装的时间;
- g) 统计在目标硬件环境、操作系统中安装 3 次的成功率;
- h) 判断系统的可移植性是否符合 7.2 的要求。

10.3 可靠性

10.3.1 可用性

该特性的测试步骤包括:

- a) 按照说明书运行待测的知识图谱应用系统;
- b) 选定待测的知识图谱;
- c) 选择概率最大的前 25%的可造成系统异常输入元素构成系统运行剖面;
- d) 统计在系统发生故障情况下, 系统处于不可用状态的时长;
- e) 统计给定运行时间内通过异常输入出现故障的次数;
- f) 计算平均故障时间;
- g) 统计访问权限的控制数量;
- h) 判断系统的可用性是否符合 7.3.1 的要求。

10.3.2 成熟性

该特性的测试步骤包括:

- a) 按照说明书运行待测的知识图谱应用系统;
- b) 选定待测的知识图谱;
- c) 选择概率最大的前 25%的可造成系统异常输入元素构成系统运行剖面;
- d) 检查系统是否有针对该错误的反馈信息;

- e) 检查系统运行出现异常时是否出现失去响应、非正常退出、功能失效或造成操作系统崩溃等异常现象；
- f) 检查系统是否提供终止、退出的功能；
- g) 按照公式（26）计算在给定运行时间内的故障排除率：

$$X_C = \frac{N_C}{N_F} \quad (26)$$

式中：

N_C ——被纠正故障数量；

N_F ——测试中出现的故障数量。

- h) 按照公式（27）计算在给定运行时间内的平均失效间隔时间：

$$T_F = \frac{T_R}{N_F} \quad (27)$$

式中：

T_R ——系统运行时间；

N_F ——系统发生故障数量。

- i) 按照公式（28）计算在给定运行时间内的故障率：

$$N_{FA} = \frac{N_F}{T_R} \quad (28)$$

式中：

N_F ——系统发生故障数量；

T_R ——系统运行时间。

- j) 判断系统的成熟性是否符合 7.3.2 的要求。

附录 A (资料性) 知识图谱应用系统的测试流程

知识图谱应用系统的测试流程分为需求收集、需求分析、测试计划、测试设计、测试执行、测试总结六个环节，见图A.1。

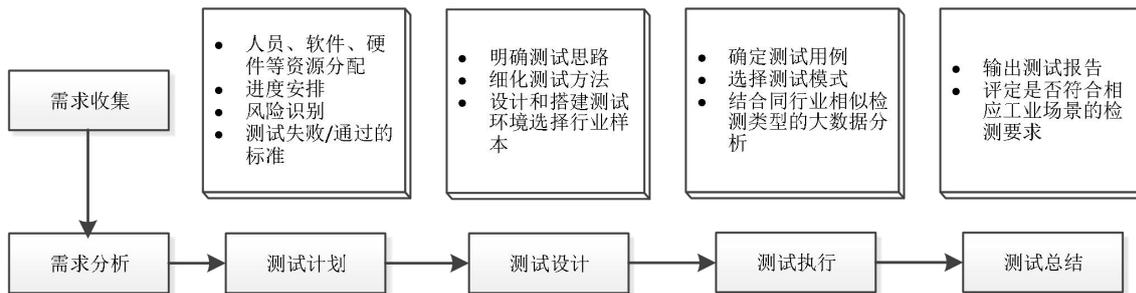


图 A.1 知识图谱应用系统测试流程

- a) 需求收集：结合待测知识图谱应用系统的具体应用场景，调研梳理待测系统在该应用场景下的具体需求及条件，形成需求文档。
- b) 需求分析：根据需求文档，结合具体应用场景设置相关参数和指标要求，形成参数化的测试需求文档，并根据上述文件及相关国家标准、行业标准，确定测试依据、测试周期。
- c) 测试计划：根据测试需求文档及相关标准，制定测试计划。测试计划应明确人员、软件、硬件、样本数据环境等资源分配，测试进度安排，测试过程风险识别以及测试通过和失败的标准。
- d) 测试设计：根据需求文档和测试计划，明确测试思路，确定待测系统的测试方法，设计检测流程，搭建测试环境，随后选择具体应用场景下的行业样本。
- e) 测试执行：根据确定的测试范围、目标和检测流程，选择单元级测试或系统级测试等模式，确定测试用例，对待测系统进行测试。
- f) 测试总结：根据测试结果出具测试报告，评定是否符合相应场景的检测要求。

参 考 文 献

- [1] GB/T 39788-2021 系统与软件工程 性能测试方法
 - [2] GB/T 25000.20-2021 系统与软件工程 系统与软件质量要求和评价（SQuaRE）第 20 部分：质量测量框架
 - [3] GB/T 29831.1-2013 系统与软件功能性 第1部分：指标体系
 - [4] GB/T 29831.2-2013 系统与软件功能性 第2部分：度量方法
-