



中华人民共和国国家标准

GB/T 45923.2—2025

人工智能 知识图谱应用平台 第2部分：性能要求与测试方法

Artificial intelligence—Knowledge graph application platform—
Part 2: Performance requirements and testing method

2025-06-30 发布

2025-10-01 实施

国家市场监督管理总局 发布
国家标准化管理委员会

目次

前言 III

引言 V

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 知识图谱构建性能要求 2

 5.1 知识表示 2

 5.2 知识建模 3

 5.3 知识获取 3

 5.4 知识存储 3

 5.5 知识融合 3

 5.6 知识计算 3

 5.7 知识溯源 4

 5.8 知识演化 4

 5.9 知识备份 4

 5.10 大模型集成 4

6 知识图谱应用性能要求 4

 6.1 响应性 4

 6.2 可移植性 4

 6.3 可用性 4

 6.4 成熟性 5

 6.5 知识应用 5

7 测试方法 5

 7.1 测试环境 5

 7.2 知识图谱构建性能测试方法 5

 7.3 知识图谱应用性能测试方法 13

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。

本文件是 GB/T 45923《人工智能 知识图谱应用平台》的第 2 部分。GB/T 45923 已经发布了以下部分：

——第 2 部分：性能要求与测试方法。

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专利的责任。

本文件由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本文件起草单位：中国电子技术标准化研究院、深圳赛西信息技术有限公司、江苏赛西科技发展有限公司、沈阳东软智能医疗科技研究院有限公司、网智天元科技集团股份有限公司、浪潮软件科技有限公司、中电科大数据研究院有限公司、途普智能科技(北京)有限公司、中国科学院自动化研究所、中国长江三峡集团有限公司、江苏电力信息技术有限公司、中国科学院沈阳自动化研究所、杭州海康威视数字技术股份有限公司、蚂蚁科技集团股份有限公司、中国医学科学院医学信息研究所、中国医学科学院生物医学工程研究所、上海人工智能研究院有限公司、中译语通科技股份有限公司、东方电气集团东方汽轮机有限公司、北京工业大学、上海文镔信息科技有限公司、北京神舟航天软件技术股份有限公司、浪潮通信信息系统有限公司、深圳市硅赫半导体有限公司、山东亿云信息技术有限公司、浙江创邻科技有限公司、厦门市美亚柏科信息安全研究所有限公司、海信集团控股股份有限公司、厦门渊亭信息科技有限公司、北京大学、中国南方电网有限责任公司超高压输电公司、豪尔赛科技集团股份有限公司、星环信息科技(上海)股份有限公司、成都信息工程大学、上海交通大学、广州赛西标准检测研究院有限公司、海义知信息科技(南京)有限公司、国能信息技术有限公司、厦门美亚亿安信息科技有限公司、深圳市优必选科技股份有限公司、国网经济技术研究院有限公司、深圳盼月亮创新技术有限公司、大唐国信滨海海上风力发电有限公司、厦门博视源机器视觉技术有限公司、福建中信网安信息科技有限公司、北京方寸无忧科技发展有限公司、东方电气集团数字科技有限公司、华能煤炭技术研究有限公司、湖北纪元慧智科技有限公司、零犀(北京)科技有限公司、西南交通大学、南方电网科学研究院有限责任公司、北方自动控制技术研究所、中移(杭州)信息技术有限公司、广电运通集团股份有限公司、重庆誉存科技有限公司、福建东飞环境集团有限公司、青岛盘古云枢人工智能科技有限公司、浪潮通用软件有限公司、珠海金山办公软件有限公司、杭州半云科技有限公司、重庆建安仪器有限责任公司、中国联合网络通信有限公司上海市分公司、中国石油天然气股份有限公司勘探开发研究院、曙光信息产业股份有限公司、融科联创(天津)信息技术有限公司、中远海运科技股份有限公司、联通沃悦读科技文化有限公司、暨南大学、济南致业电子有限公司、湖北华中电力科技开发有限责任公司、宁波宁帆信息科技有限公司、深圳数影科技有限公司、国家石油天然气管网集团有限公司、国家石油天然气管网集团有限公司山东分公司、广西机电职业技术学院、天津电子信息职业技术学院、国家管网集团北京管道有限公司、欧冶链金再生资源有限公司、郑州数智技术研究院有限公司。

本文件主要起草人：李瑞琪、韩丽、卓兰、程雨航、王伟光、贾承斌、王珂琛、谢真强、周翀、何赛克、黄鹏飞、冯曙明、王挺、陈诚、郭智慧、李姣、蒲江波、胡成林、宋海涛、石嘉豪、陈自岩、杜永萍、仲凯韬、密伟、肖红梅、洪鹏辉、吴士伟、周研、蔡宇锋、洪万福、邹磊、刘常昱、朱斯燕、唐剑飞、李佳、闫理跃、李湘、马立群、张树蕊、杨帆、傅洛伊、蔡惠民、吴萌、姚博、黄胜华、胡芳槐、张天霖、平艳、焦继超、施伟国、

张学琴、陈洪峰、邓徐韬、王耀辉、龚家元、石致远、王华、刘姝、罗红、杨波、阙锦龙、汪诗雨、祝景阳、龙铭刚、蔡巍、梁秉豪、芮子文、张晨、梁乔玲、王资凯、李照川、吴伟、潘风文、翁武焰、李强、方帅、王亚威、王林、侯丽、邵一凡、林叠守、李文杰、彭轶淳、孙晓虎、李方昱、梁磊、何颖、罗锋、刘锐钢、张超、张在萍、刘德彬、陈玮、黄志敏、朱熠锆、曾文佳、姚冬、陈卫强、吴少华、李运玲、王绍函、黄斐然、何熠、曹永军、梁火炼、徐文峰、李君、张岩、孟庆阳、陶晓英、魏立钧、徐文涛、陈漱宇、郎俊奇、张明英、韩宁、李欣、李小波、任勇、许建超、李官政、魏子尧、包之明、杨阳、陈小平、陶炜、李勇、张迎华、田利新、高敏亮、李勇刚、刘祥元、刘威、林正平、杜英俊、张泽民、苏文杰、姚远、胡兵。



引 言

目前,知识图谱在金融证券、生物医疗、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。知识图谱应用平台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。

GB/T 45923《人工智能 知识图谱应用平台》拟由四个部分构成。

- 第1部分:功能要求。目的在于规定知识图谱应用平台的系统架构与功能要求。
- 第2部分:性能要求与测试方法。目的在于规定知识图谱应用平台的性能要求及对应测试方法。
- 第3部分:知识服务。目的在于规定基于知识图谱的知识问答、知识检索和知识推荐等知识服务要求。
- 第4部分:知识管理。目的在于规定知识图谱应用平台的知识管理要求。

人工智能 知识图谱应用平台

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平台的性能要求,描述了相应的测试方法。

本文件适用于知识图谱应用平台的设计、开发应用和性能测试。

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 1988 信息技术 信息交换用七位编码字符集

GB/T 13000 信息技术 通用编码字符集(UCS)

GB 18030 信息技术 中文编码字符集



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, 3.6]

3.3

知识单元 knowledge unit

按一定关系组织的一组知识元素的集合。

[来源:GB/T 42131—2022, 3.7]

3.4

知识图谱应用平台 knowledge graph application platform

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

知识获取的精确率 precision of knowledge acquisition

已获取知识中正确知识的占比。

3.6

知识获取的召回率 **recall of knowledge acquisition**

已获取知识对正确知识覆盖的程度。

3.7

多跳邻居查询响应时间 **response time of K Nearest Neighbour query**

系统完成 1 跳、2 跳、多跳等不同展开相连接层级中关联对象的查询时间。

3.8

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。

3.9

知识图谱体量 **knowledge graph volume**

知识图谱中包含的实体和关系数量。

3.10

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。

3.11

平均故障恢复时间 **mean time to recovery**

系统从故障发生到恢复正常运行所需的平均时间。

3.12

故障排除率 **troubleshooting rate**

在特定时间内故障被有效解决的比例。

3.13

故障频率 **failure frequency**

单位时间内发生故障的次数。

4 缩略语

下列缩略语适用于本文件。

ARM:高级精简指令集机器(Advanced RISC Machine)

CPU:中央处理器(Central Processing Unit)

Graph-RAG:图检索增强生成(Graph-based Retrieval-Augmented Generation)

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 知识图谱构建性能要求

5.1 知识表示

5.1.1 单个实体或关系的响应时间应不大于 0.1 s。

5.1.2 多个实体或关系(不小于 2 个,小于或等于 10 个)的响应时间应不大于 1 s。

5.1.3 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。

5.1.4 系统最大可呈现实体或关系数量应不少于 200 个。

5.1.5 系统可呈现的实体或关系布局方式应不少于 5 种,如:树型、圆型、力导向型、谱系型和双曲线

型等。

5.2 知识建模

5.2.1 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的语义定义,本体模型的明确度不小于 95%。

5.2.2 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%。

5.2.3 系统中构建的本体模型应正确一致地展示实体类型、关系类型和属性,且由术语得出的推理或计算结果不与术语本身含义产生矛盾,本体模型的一致度不小于 90%。

5.2.4 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%。

5.3 知识获取

5.3.1 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,知识获取的精确率应不小于 75%。

5.3.2 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%。

5.3.3 面向结构化数据,知识获取的综合评价指标(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%。

注:综合评价指标(F1)测量值是精确率和召回率的加权调和平均值,兼顾了知识抽取的精确率和召回率。

5.4 知识存储

5.4.1 知识存储模块支持的字符集类型数量应不小于 3 个,并包括:GB 18030 规定的中文编码字符集、GB/T 13000 规定的通用编码字符集、GB/T 1988 规定的信息交换用七位编码字符集。

5.4.2 知识存储模块完成基准数据集加载的时间应不大于 10 s。

5.4.3 知识存储模块完成 1 跳邻居查询的时间应不大于 0.1 s;完成 2 跳邻居查询的时间应不大于 0.5 s;完成 3 跳邻居查询的时间应不大于 2 s。

5.4.4 多跳邻居查询具备的最大并发数应不少于 20 个。

5.4.5 在 10 万条知识规模的条件下,毫秒级 1 跳的数据查询响应率应不小于 90%。

5.4.6 在 20 并发数下,一度邻居的吞吐率不低于 10。

5.4.7 QPS(单条指令)用于度量系统在每秒能响应的单条指令查询次数,应不小于 100 条。

5.4.8 QPS(多条指令)用于度量系统在每秒能响应的多条指令查询次数,应不小于 50 条。

5.4.9 数据加载存储空间容量应不低于 500 GB。

5.4.10 知识存储周期应不低于 3 个月。

5.5 知识融合

5.5.1 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%。

5.5.2 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%。

5.5.3 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%。

5.6 知识计算

5.6.1 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用的时间,应不大于 10 s。

5.6.2 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。

5.6.3 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%。

5.6.4 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 512 GB。

5.6.5 知识计算模块具备的算法类型应不少于三种。

5.6.6 知识计算模块中图数据库具备的算法类型应不少于三种。

5.7 知识溯源

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

5.7.2 知识溯源的准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%。

5.8 知识演化

5.8.1 平台应具备支持三个以上版本的能力。

5.8.2 平台的版本更新周期应支持设置为每三个月至少一次。

5.9 知识备份

5.9.1 平台的本地知识备份周期应支持设置为每三个月至少一次。

5.9.2 平台的异地知识备份周期应支持设置为每三个月至少一次。

5.10 大模型集成

5.10.1 宜至少具备一种服务接口协议。

5.10.2 宜具备至少二种大模型或 2 种版本以上的大模型。

5.10.3 宜具备至少一种 Graph-RAG 架构。

6 知识图谱应用性能要求

6.1 响应性

6.1.1 知识图谱体量应不小于 10 万个。

6.1.2 知识图谱复杂度应不小于 10 个。

6.1.3 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。

6.1.4 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%。

6.1.5 实时性用于度量知识图谱应用平台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%。

6.2 可移植性

6.2.1 平均安装时间应不大于 12 h。

6.2.2 安装成功率应不小于 80%。



6.3 可用性

6.3.1 平均故障恢复时间应不大于 3 h。

6.3.2 访问权限的控制数量应不少于 20 个。

6.4 成熟性

6.4.1 故障排除率应不小于 99%，故障类型包括：平台失去响应、非正常退出、功能失效或造成操作系统崩溃等。

6.4.2 故障频率在 1 个月内应不大于 1 次。

6.5 知识应用

6.5.1 应具备至少一种通用知识应用，如：知识问答、综合知识检索、知识推荐、辅助决策和内容生成等。

6.5.2 宜具备至少一种细分领域的知识应用，如：辅助诊断、设备运维、智能客服和反欺诈等。

7 测试方法

7.1 测试环境

除另有规定外，知识图谱应用平台的测试环境应符合以下要求。

- a) 具备知识图谱应用平台专用的测试工具，且部署了与待测系统相适应的测试场景。
- b) 环境温度：15℃～35℃。
- c) 相对湿度：25%～75%。
- d) 大气压力：86 kPa～106 kPa。
- e) 操作系统：至少具备一种基于 x86 架构或基于 ARM 的操作系统。
- f) 测试服务器 CPU：不小于 32 核。
- g) 测试服务器内存：不小于 256 GB。
- h) 测试服务器外存：不小于 1 TB。

7.2 知识图谱构建性能测试方法

7.2.1 知识表示

7.2.1.1 知识表示响应时间

知识表示响应时间的测试步骤如下：

- 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)计算知识表示响应时间 T_p 。

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

7.2.1.2 知识表示响应速度

知识表示响应速度的测试步骤如下：

- 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 。

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

7.2.1.3 最大可呈现实体或关系数量

最大可呈现实体或关系数量的测试步骤如下:

- 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) 统计呈现结束时的实体或关系数量;
- d) 递增拟呈现的实体或关系的数量;
- e) 按 b) 继续执行,直至呈现结束时实体或关系数量无法继续增加;
- f) 统计最终呈现的实体或关系数量。

7.2.1.4 实体或关系布局方式数量

实体或关系布局方式数量的测试步骤如下:

- a) 打开知识表示相关模块;
- b) 设置拟呈现实体或关系的布局方式;
- c) 运行设置的布局方式;
- d) 按 b) 设置并运行下一项布局方式,直至遍历所有的布局方式;
- e) 统计实体或关系布局方式数量。

7.2.2 知识建模

7.2.2.1 本体模型的明确度

该特性的测试步骤如下:

- 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 $X_{O.C}$ 。

$$X_{O.C} = \frac{A_{O.C}}{B_{O.C}} \dots\dots\dots (3)$$

7.2.2.2 本体模型的完整度

该特性的测试步骤如下:

- 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) 筛选出对应的原始数据;
- 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 $X_{O.I}$ 。

$$X_{O.I} = \frac{A_{O.I}}{B_{O.I}} \dots\dots\dots (4)$$

7.2.2.3 本体模型的一致度

该特性的测试步骤如下：

- 打开知识模块相关模块；
- 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；
- 筛选出对应的原始数据；
- 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} ；
- 按公式(5)计算本体模型的一致度 X_{O_U} 。

$$X_{O_U} = \frac{A_{O_U}}{B_{O_U}} \dots\dots\dots (5)$$

7.2.2.4 本体模型的最大单调性

该特性的测试步骤如下：

- 打开知识模块相关模块；
- 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；
- 筛选出对应的原始数据；
- 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} ；
- 按公式(6)计算本体模型的最大单调性 X_{O_M} 。



$$X_{O_M} = \frac{A_{O_M}}{B_{O_M}} \dots\dots\dots (6)$$

7.2.3 知识获取

7.2.3.1 结构化数据

知识获取相关模块性能的测试步骤如下。

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

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

- 按公式(8)计算知识获取的召回率 $Recall_A$ 。

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

- 按公式(9)计算知识获取的 F1 测量值 $F1\text{-score}_A$ 。

$$F1\text{-score}_A = 2 \times \frac{Precision_A \times Recall_A}{Precision_A + Recall_A} \dots\dots\dots (9)$$

7.2.3.2 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如下：

- 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。

7.2.4 知识存储

7.2.4.1 字符集类型数

字符集类型数的测试步骤如下：

- 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) 确认支持的字符集类型是否包含 GB 18030 规定的中文编码字符集、GB/T 13000 规定的通用编码字符集、GB/T 1988 规定的信息交换用七位编码字符集。

7.2.4.2 数据集加载时间

数据集加载时间的测试步骤如下。

- a) 打开知识存储相关模块。
- b) 准备拟加载的数据集，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。

$$X_{li} = \text{Rec}_{li} - \text{Send}_{li} \quad \dots\dots\dots (10)$$

式中：

- X_{li} ——第 i 次加载的响应时间；
- Rec_{li} ——第 i 次加载收到响应结果的时间；
- Send_{li} ——第 i 次发出加载数据请求的时间。

- f) 执行 n 次步骤 c)~e)，并计算平均值。

7.2.4.3 多跳邻居查询响应时间

多跳邻居查询响应时间的测试步骤如下。

- a) 打开知识存储相关模块。
- b) 设置多跳查询参数和多跳查询的超时时间阈值。
- c) 统计发出指令的时间 Send_{si} 。
- d) 统计查询结束的时间 Rec_{si} 。
- e) 按公式(11)和公式(12)计算多跳邻居查询响应时间 X_{si} 和多跳邻居查询的超时时间 Y_i 。

$$X_{si} = \text{Rec}_{si} - \text{Send}_{si} \quad \dots\dots\dots (11)$$

$$Y_i = X_{si} - X_T \quad \dots\dots\dots (12)$$

式中：

- X_{si} ——第 i 次多跳查询的响应时间；
 Rec_{si} ——第 i 次多跳查询收到响应结果的时间；
 $Send_{si}$ ——第 i 次发出多跳查询请求的时间；
 Y_i ——第 i 次多跳查询的超时时间；
 X_T ——多跳查询的超时时间阈值。

- f) 执行 n 次步骤 c)~e), 并计算平均值。

7.2.4.4 多跳邻居查询具备的最大并发数

多跳邻居查询具备的最大并发数的测试步骤如下。

- 打开知识存储相关模块。
- 设置并发请求数序列。
- 针对每个需要尝试的并发请求数执行并发测试。
- 针对每个请求, 测试并发请求的响应时间或吞吐率。
- 监控测试结果, 如果出现以下情况, 可停止测试:
 - 测试过程出现异常, 测试无法正常完成;
 - 响应时间变得不可测试。
- 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。
- 统计在测试停止时, 多跳邻居查询具备的最大并发数。

7.2.4.5 数据查询响应率

数据查询响应率的测试步骤如下。

- 确定点样本和边样本。
- 打开知识存储相关模块。
- 设置并发请求线程数和总操作数。
- 设置并发操作序列。
- 每个线程循环执行操作序列。
- 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。
- 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。
- 按公式(13)计算毫秒级数据查询响应率 X_R 。

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

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。

7.2.4.6 吞吐率

吞吐率测试步骤如下。

- 确定点样本和边样本。
- 打开知识存储相关模块。
- 设置并发请求线程数和总操作数。
- 设置并发操作序列。
- 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(14)计算吞吐率 RPS。

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

式中:

CR_g —— 总请求数;

T_g —— 处理这些请求的总完成时间。

- i) 还可依次执行如下压力测试,并统计吞吐率:
 - 1) 执行查询点和查询边压力测试;
 - 2) 执行增加点和增加边压力测试;
 - 3) 执行修改点和修改边压力测试;
 - 4) 执行删除点和删除边压力测试。

7.2.4.7 QPS(单条指令)

QPS(单条指令)的测试步骤如下:

- a) 打开知识存储相关模块;
- b) 设置并发请求线程数和持续时间;
- 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,当测试时间等于步骤 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。

7.2.4.8 QPS(多条指令)

QPS(多条指令)的测试步骤如下:

- a) 打开知识存储相关模块;
- b) 设置并发请求线程数和持续时间;
- 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,当测试时间等于步骤 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。

7.2.4.9 数据加载存储空间容量

数据加载存储空间的测试步骤如下:

- a) 打开知识存储相关模块;
- b) 选择待加载的数据存储位置;
- c) 统计可用的数据加载存储空间容量。

7.2.4.10 知识存储周期

知识存储周期的测试步骤如下:

- a) 打开知识存储相关模块;
- b) 选择已归档的知识图谱;
- c) 统计该知识图谱中知识的创建时间;

- d) 统计该知识图谱中知识的归档时间；
- e) 根据归档时间与创建时间差值,计算知识存储周期。

7.2.5 知识融合

知识融合相关模块性能的测试步骤如下。

- a) 打开知识表示相关模块。
- b) 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) 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)、测试时长和并发用户数(如适用)等,确保测试参数的设置能准确反映系统在实际应用中的使用情况。
- d) 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) 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) 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。

$$\text{Precision}_F = \frac{TP_F}{TP_F + FP_F} \dots\dots\dots (15)$$

- k) 按公式(16)计算知识融合的召回率。

$$\text{Recall}_F = \frac{TP_F}{TP_F + FN_F} \dots\dots\dots (16)$$

- l) 按公式(17)计算知识融合的 F1 测量值 $F1\text{-score}_F$ 。

$$F1\text{-score}_F = 2 \times \frac{\text{Precision}_F \times \text{Recall}_F}{\text{Precision}_F + \text{Recall}_F} \dots\dots\dots (17)$$

7.2.6 知识计算

知识计算相关模块性能的测试步骤如下。

- a) 按说明书运行待测的系统。
- b) 选定待计算的知识图谱。
- c) 选定知识计算任务,配置参数,并执行计算任务。
- d) 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。
- e) 按公式(18)计算知识计算任务的响应时间 T 。

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

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准确率。

$$\text{Precision} = \frac{TP_{KC}}{TP_{KC} + FP_{KC}} \dots\dots\dots (19)$$

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。

g) 执行 n 次步骤 c)~f)，按公式(20)计算知识计算任务的均方根误差。

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \quad \dots\dots\dots (20)$$

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。

h) 统计知识计算模块可具备的算法模型的数量。

i) 统计计算任务执行过程中耗费的最大软硬件资源。

7.2.7 知识溯源

知识溯源相关模块性能的测试步骤如下。

- a) 按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 。

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

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)计算知识溯源的准确率 A 。

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

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。

7.2.8 知识演化

知识演化相关模块性能的测试步骤如下：

- a) 按说明书运行待测的系统；
- b) 选定待运行的知识图谱；
- c) 对该知识图谱进行版本名称定义，并生成新的版本；
- d) 重复执行三次步骤 c)，计算系统内成功存储的知识图谱版本数量；
- e) 打开知识图谱版本更新周期设置模块；
- f) 查询知识图谱版本更新周期可选设置列表。

7.2.9 知识备份

知识备份相关模块性能的测试步骤如下：

- 按说明书运行待测的系统；
- 选定待运行的知识图谱及其备份；
- 打开系统内本地知识备份周期设置模块；
- 查询系统内本地知识备份可选设置列表；
- 打开系统内异地知识备份周期设置模块；
- 查询系统内异地知识备份可选设置列表。

7.2.10 大模型集成

大模型集成相关模块性能的测试步骤如下：

- 按说明书运行待测的系统；
- 选定系统集成的大模型；
- 统计系统支持的大模型服务接口协议数量；
- 统计系统集成的大模型数量及其版本数量；
- 统计系统具备的 Graph-RAG 架构数量。

7.3 知识图谱应用性能测试方法

7.3.1 响应性

7.3.1.1 知识图谱体量

知识图谱体量的测试步骤如下：

- 运行系统,并选定待测的知识图谱；
- 统计知识图谱中的实体数量 V_e ；
- 统计知识图谱中的关系数量 V_r ；
- 按公式(23)计算实体数量和关系数量的总和 V_g 。

$$V_g = V_e + V_r \quad \dots\dots\dots (23)$$

7.3.1.2 知识图谱复杂度

知识图谱复杂度的测试步骤如下：

- 运行系统,并选定待测的知识图谱；
- 统计知识图谱中的实体类型数量 F_e ；
- 统计知识图谱中的关系类型数量 F_r ；
- 按公式(24)计算实体类型数量和关系类型数量的总和 F_g 。

$$F_g = F_e + F_r \quad \dots\dots\dots (24)$$

7.3.1.3 效率

效率的测试步骤如下：

- 运行系统,并选定待测的知识图谱；
- 设置查询参数,如多跳数等；
- 设置并发参数；
- 启动查询操作；



- e) 设定 1 min,统计成功处理的请求总数;
- f) 按公式(25)计算系统的效率 E 。

$$E = \frac{A_R}{T} \dots\dots\dots (25)$$

式中:

A_R ——处理请求总数;

T ——处理请求所用的时间。

7.3.1.4 准确率

准确率的测试步骤如下:

- a) 运行系统,并选定待测的知识图谱;
- b) 设置查询参数,如多跳数等;
- c) 设置并发参数;
- d) 启动查询操作;
- e) 设定 1 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(26)计算准确率。

$$\text{Precision}_R = \frac{TP_R}{TP_R + FP_R} \dots\dots\dots (26)$$

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。

7.3.1.5 实时性

实时性的测试步骤如下:

- a) 运行系统,并选定待测的知识图谱;
- b) 设置查询参数,如多跳数等;
- c) 设置并发参数;
- d) 启动查询操作;
- e) 设定 1 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(27)计算系统的实时性 X_A 。

$$X_A = \frac{A_R}{A} \dots\dots\dots (27)$$

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。

7.3.2 可移植性

可移植性的测试步骤如下:

- a) 运行系统,并选定待测的知识图谱;
- b) 在目标硬件环境下安装和运行系统,确定系统可进行安装并能正常启动,功能可正常运行;

- c) 在目标操作系统下安装和运行系统,确定系统可进行安装并能正常启动,功能可正常运行;
- d) 检查系统是否具备常见的接口调用,如:HTTP 和 RPC 等;
- e) 检查系统是否具备数据的导入导出;
- f) 统计在目标硬件环境和操作系统中完成安装的时间;
- g) 统计在目标硬件环境和操作系统中安装 5 次的成功率。

7.3.3 可用性

7.3.3.1 平均故障恢复时间

平均故障恢复时间的测试步骤如下:

- a) 按说明书运行待测的知识图谱应用平台;
- 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,计算系统的平均故障恢复时间;
- e) 统计给定运行时间内通过异常输入出现故障的次数;
- f) 计算平均故障恢复时间。

7.3.3.2 访问权限控制数量

访问权限控制数量的测试步骤如下:

- a) 按说明书运行待测的知识图谱应用平台;
- b) 打开平台访问权限控制模块;
- c) 统计平台上针对用户和角色设置的访问权限控制条目数量;
- d) 统计平台上针对知识展示、知识编辑、知识下载等方面设置的访问权限条目数量;
- e) 根据步骤 c)和步骤 d)的总条目数量,计算访问权限控制数量。

7.3.4 成熟性



成熟性的测试步骤如下。

- a) 按说明书运行待测的知识图谱应用平台。
- 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) 设置平台运行时间为 1 个月。
- f) 在给定运行时间内,记录平台运行日志。
- g) 检查日志中平台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。
- h) 检查平台是否提供终止和退出的功能。
- i) 按公式(28)计算在给定运行时间内的故障排除率 X_c 。

$$X_c = \frac{N_c}{N_F} \dots\dots\dots (28)$$

式中:

N_c ——被纠正故障数量;

N_F ——测试中出现的故障数量。

- j) 按公式(29)计算在给定运行时间内的故障频率 N_{FA} 。

$$N_{FA} = \frac{N_F}{T_R}$$
..... (29)

式中：
N_F —— 系统发生故障数量；
T_R —— 系统运行时间。

7.3.5 知识应用

- 知识应用的测试步骤如下：
- a) 按说明书运行待测的知识图谱应用平台；
 - b) 计算平台上的通用知识应用数量；
 - c) 计算平台上的细分领域知识应用数量。



参 考 文 献

- [1] GB/T 23703.2—2010 知识管理 第2部分:术语
 - [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(SQure) 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
 - [5] GB/T 39788—2021 系统与软件工程 性能测试方法
 - [6] GB/T 42131—2022 人工智能 知识图谱技术框架
-



