

迭代三测试文档

版本	作者	完成时间	内容
v1.0	蒋祚竑	2021.6.18	问答系统功能测试

迭代三工作内容概述：

迭代三中我们以智能问答系统为核心目标，构建了一个三国志主题的人物、事件、地点的智能问答系统。由于该问答系统为整个迭代三唯一的新增功能，因此迭代三的测试是针对该问答系统的功能测试。

迭代三测试说明：

我们的项目同时实现了单跳问答和基于特定模板的多跳问答。单跳是指单实体单属性问题，例如“某某的爸爸是谁”，多跳指的是多实体-多属性问题，例如“某地某年发生了什么事”（包含地点、年份、事件实体，地点、年份作为限制条件），“某人的亲属关系”（亲属关系在我们的系统中包含目标实体的儿子、女儿、父亲、母亲四个属性）。因此我们的功能测试分为单跳问答测试和多跳问答测试两部分。

单跳问答测试：

我们在知识图谱中存储了事件、人物、地点、时间四种类型的节点。其中时间节点和地点节点仅用于辅助多跳问答系统实现时间相关的查询，因此单跳问答功能测试只针对人物、事件两种类型节点。

人物：

针对数据库中存储的人物节点的属性，我们设计了以下测试问题模板：

- 1.XX为何而死
- 2.XX怎么死的
- 3.怎么死的XX
- 4.XX去世原因
- 5.XX什么时候走的
- 6.XX何时死的
- 7.XX何时与世长辞
- 8.XX的死去原因怎么死的
- 9.XX荐举了谁
- 10.XX的荐主是谁
- 11.XX被谁荐举的
- 12.XX支持谁

- 13.XX的支持者有谁
- 14.XX的别名
- 15.XX的老家在哪呀
- 16.XX是哪里人阿
- 17.XX是哪国的
- 18.XX是哪个朝代的
- 19.XX字是什么
- 20.讲一讲XX的生平经历
- 21.XX被封为什么爵位
- 22.XX哪年出生的
- 23.XX什么时候出生的
- 24.XX当过什么官
- 25.XX当过哪些官
- 26.XX的儿子
- 27.XX的女儿
- 28.XX的母亲
- 29.XX的父亲
- 30.XX的配偶

对于每个问题模板，我们都至少使用两个人物进行两次以上的查询测试。

正确率：在数据库中存储有对应信息的情况下，均能给出正确答案。

响应时间：对于问题中直接包含词汇表中属性名的情况，例如XX的别名，XX的儿子/女儿等，平均响应时间为500ms，对于无法匹配到词汇表属性名的情况，例如"XX被封为什么爵位"，"XX哪年出生的"，"XX当过什么官"，"XX何时与世长辞"等问题，平均响应时间为8.3s。

事件：

针对数据库中存储的人物节点的属性，我们设计了以下测试问题模板：

- 1.XX什么时候发生的
- 2.XX是什么时候的事情
- 3.XX哪年发生的
- 4.XX在哪发生的
- 5.XX在什么地方发生的
- 6.和XX有关的事件
- 7.和XX有关的事情
- 8.XX的相关事件
- 9.XX的影响
- 10.XX的意义

- 11.XX的过程
- 12.XX的经过
- 13.XX有哪些人参与
- 14.XX的相关人物
- 15.XX的起因
- 16.XX发生的原因
- 17.为什么会有XX

对于每个问题模板，我们均套用至少两个事件进行两次以上的查询测试。

正确率：问题模板3和14回答错误，其他模板在数据库中存储有对应信息的情况下，均能给出正确答案。问题3的错误原因在于我们的相似度计算模型在该问题中为“地点”这一答案打了很高的分数，我们猜测是训练集数据导致的过拟合。问题14的错误原因是“相关”一词直接匹配到了事件节点的属性名，便不再进行后续的答案选择。对于这两种错误情况，我们均使用了词汇表暴力匹配的方式解决。

响应时间：对于问题中直接包含词汇表中属性名的情况，例如XX发生的原因，XX的影响等，平均响应时间为450ms，对于无法匹配到词汇表属性名的情况，例如“XX在哪发生的”，“和XX有关的事件”，“XX有哪些人参与”等问题，平均响应时间为2.5s。

多跳问答测试：

目前我们共实现了六个多跳问题模板：

- 1.某地某年发生了什么事
- 2.某地某年有谁出生
- 3.某人的亲属关系
- 4.某人的社会关系
- 5.介绍某人
- 6.介绍某地

对于每个问题模板我们套用了五组以上的问题进行测试。在数据库存储有相应问题答案的情况下，以上六个模板均能达到100%正确率，并且响应时间不超过0.5s。