目录

一 引言

* 1. 编写目的
  2. 背景
  3. 参考资料

二 用途

2.1功能

2.2 性能

2.3 安全性

三 运行环境

四 使用说明

一 引言

1.1编写目的

本说明书为指导平台使用人员使用“百科知识生产平台”而编写，希望该手册使他们在使用系统过程中能起到一定的作用。本手册介绍了百科知识生产平台的预处理、知识建模、知识获取、知识融合、知识管理、知识共享和重用等功能。

1.2背景

该系统由清华大学知识工程实验室开发。使用者是知识图谱研究人员、工程师、学生等人员。

1.3参考资料

二 用途

2.1功能



如图中所示为百科类知识图谱构建流程，其中实线部分为必须部分，而虚线部分为可选项。本系统考虑有四大类百科类知识图谱构建流程：第一种，不带自动更新和不带知识融合的知识图谱构建流程，则图中的新词获取和知识融合是不选择的，整体流程则为预处理、知识建模、知识获取、知识管理和知识共享于重用几个部分；第二种，带自动更新但是不带知识融合的知识图谱构建流程，则图中的新词获取保留，但是知识融合不保留，整体流程则为预处理、知识建模、知识获取、知识管理和知识共享于重用几个部分，其中完成知识获取后，如果有新的词汇获取到，然后进行相关百科页面获取，再然后进行预处理、知识建模、知识获取、知识管理和知识共享于重用部分；第三种，不带自动更新但是带知识融合的知识图谱构建流程，则图中的新词获取不选择，但是选择知识融合，整体的流程则为预处理、知识建模、知识获取、知识融合、知识管理和知识共享与重用几个部分，其中每个知识图谱在构建的流程中当完成知识获取步骤后，和其他知识库共同完成知识融合步骤，最后再进行知识管理和知识共享与重用步骤；第四种，带有自动更新和知识融合的知识图谱构建过程，则图中的新词获取和知识融合选择，整体的流程则为预处理、知识建模、知识获取、知识融合、知识管理和知识共享与重用几个部分，其中任意一个知识库完成知识获取后，如果有新的词汇获取到，然后进行相关百科页面获取，再然后进行预处理、知识建模、知识获取，然后和其他知识库共同完成知识融合步骤，最后再进行知识管理和知识共享与重用步骤。



1预处理：

功能介绍： 预处理是指在知识图谱构建前对获取到的数据进行的一些处理。本平台预处理是对百科原始数据进行截取，预处理结果为结构化数据。

上传数据格式要求：百度百科上传文件为jason格式的html集合， 中文维基百科和英文维基百科上传文件为dump 。

2.1概念抽取

功能介绍：概念抽取主要是从资源数据中抽取得到类别数据。本平台概念抽取是对进行过预处理的百科原始数据进行抽取得到其概念类别，概念抽取结果为三元组。

上传数据格式要求：百度百科、中文维基百科和英文维基百科上传文件为预处理后格式化的数据，其中百度百科推荐使用本平台预处理概念截取结果，维基百科推荐使用本平台预处理信息框截取结果。

2.2 概念上下位生成

功能介绍：概念抽取主要是从资源数据中得到类别上下位关系。本平台概念上下位生成是对进行过预处理的百科原始数据进行处理的得到概念上下位，概念上下位生成结果为三元组。

上传数据格式要求：百度百科、中文维基百科和英文维基百科上传文件为预处理后格式化的数据，其中百度百科推荐使用预处理概念截取结果，维基百科使用预处理概念截取结果、信息框截取结果。

3.1实例抽取

功能介绍：实例抽取主要是从资源数据中抽取得到实例数据。本平台实例抽取是对进行过预处理的百科原始数据进行抽取得到其实例，实例抽取结果为三元组。

上传数据格式要求：百度百科、中文维基百科和英文维基百科上传文件为预处理后格式化的数据，其中百度百科推荐使用本平台预处理摘要截取、简介截取、歧义截取和信息框截取结果，维基百科推荐使用本平台预处理信息框截取结果。

3.2 实例分类

功能介绍：实例分类主要是从资源数据中得到实例和类别关系。本平台实例分类是对进行过预处理的百科原始数据进行处理得到实例类别信息，实例分类处理结果为三元组。

上传数据格式要求：百度百科、中文维基百科和英文维基百科上传文件为预处理后格式化的数据，其中百度百科推荐使用预处理概念截取、摘要截取、简介截取、歧义截取结果，维基百科推荐使用本平台预处理概念截取、信息框截取结果。

3.3 实例属性抽取

功能介绍：实例属性抽取主要是从资源数据中得到实例属性数据。本平台实例属性抽取是对进行过预处理的百科原始数据进行处理得到实例属性信息，实例属性抽取结果为三元组。

上传数据格式要求：百度百科、中文维基百科和英文维基百科上传文件为预处理后格式化的数据，其中百度百科推荐使用预处理概念截取、简介截取、信息框截取、歧义截取结果，维基百科推荐使用本平台预处理信息框截取结果。

3.4 实例摘要抽取

功能介绍：实例摘要抽取主要是从资源数据中得到实例摘要数据。本平台实例摘要抽取是对进行过预处理的百科原始数据进行处理得到实例属性信息，实例摘要抽取结果为三元组。

上传数据格式要求：百度百科、中文维基百科和英文维基百科上传文件为预处理后格式化的数据，其中百度百科推荐使用预处理摘要截取、简介截取、歧义截取结果，维基百科推荐使用本平台预处理概念截取、信息框截取结果。

3.5 实例近似实体获取

功能介绍：实例近似实体获取主要是从资源数据中得到实例近似实体数据。本平台实例近似实体获取是对进行过预处理的百科原始数据进行处理得到实例近似实体信息，实例近似实体获取结果为三元组。

上传数据格式要求：百度百科、中文维基百科和英文维基百科上传文件为预处理后格式化的数据，其中百度百科推荐使用预处理摘要截取、简介截取、概念截取、歧义截取结果，维基百科推荐使用本平台预处理摘要截取、简介截取、信息框截取结果。

4.1 概念跨语言对齐

功能介绍：概念跨语言对齐是一种以上资源中概念数据进行对齐的功能。本平台是对百度百科、中文维基百科和英文百科的概念数据进行对齐的功能，概念跨语言对齐结果为三元组。

上传数据格式要求：百度百科、中文维基百科和英文维基百科上传为预处理后的格式化数据。

4.2 实例跨语言对齐

功能介绍：实例跨语言对齐是一种以上资源中概念数据进行对齐的功能。本平台是对百度百科、中文维基百科和英文百科的实例数据进行对齐的功能，概念跨语言对齐结果为三元组。

上传数据格式要求：百度百科、中文维基百科和英文维基百科上传为预处理后的格式化数据。

5动态更新

动态更新，主要是对数据源中的数据随时进行更新得到最新的百科知识。

结果展示： 见git

6知识图谱自动生成

本功能是对平台所有的功能进行整合管理，供用户实现一键傻瓜式操作。

上传数据格式要求：百度百科上传文件为html， 中文维基百科和英文维基百科上传文件为dump 。

2.2 性能

2.3 安全性

系统具有较好的安全保密机制，每人在使用系统前必须先登录，依自己的权限使用系统。

三 运行环境

四 使用说明