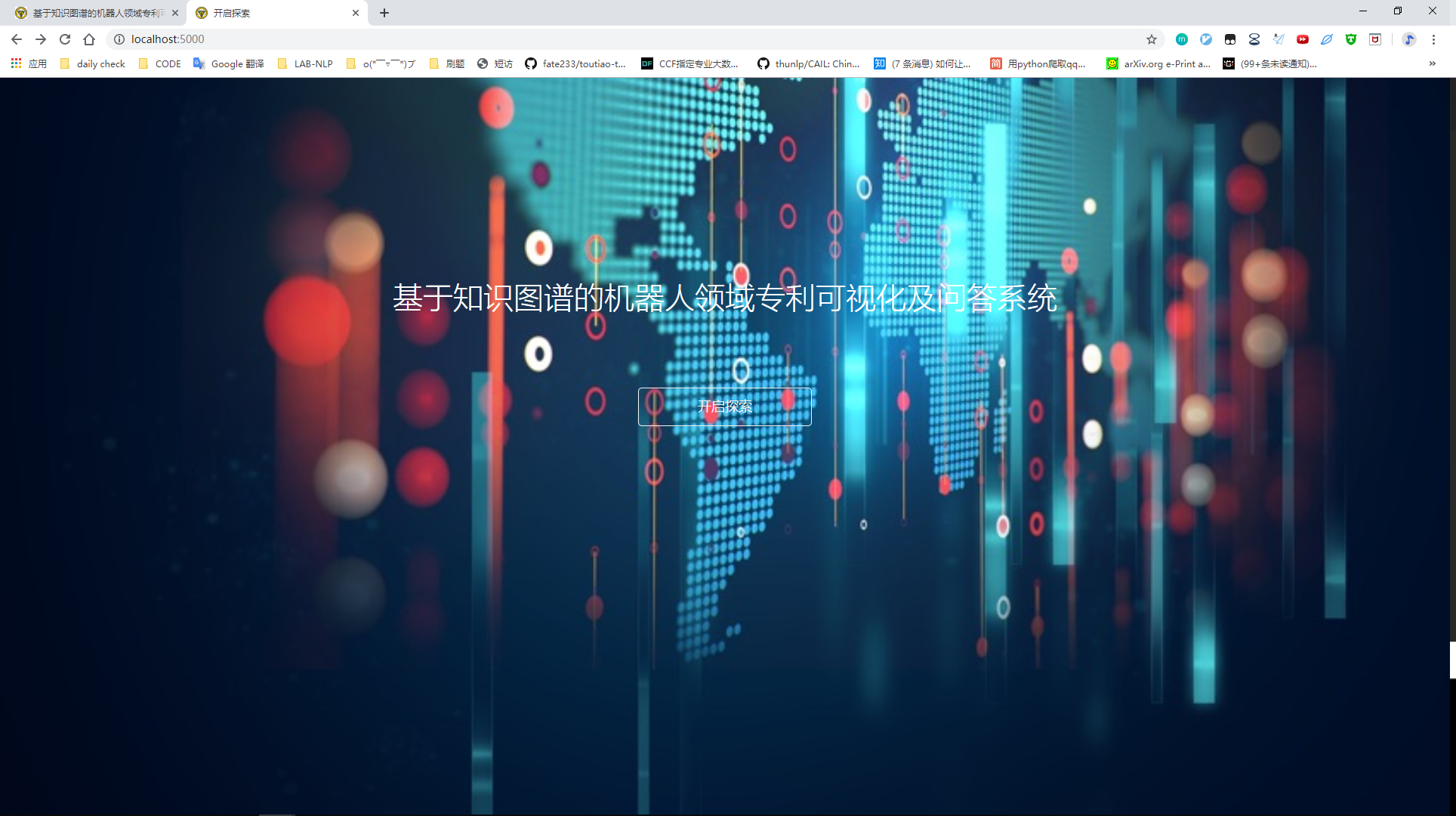
# 机器人专利知识图谱



## 机器人专利知识图谱功能模块

对机器人领域专利进行类别划分。根据类别和专利之间的相似度和CPC分类号分别赋予一定权值，根据权值大小对每个专利文本进行分类。

对机器人手册中介绍的机器人设计构造流程进行结构化提取。

关键技术查询。结合网页结构化语句和分类文本和引用关系等进行关键技术信息返回。

专利文本关键技术提取。从每个专利文本中提取得到关键技术和包含关系等信息。

技术术语推荐。根据键入短语的相似度在知识图谱中检索最相关的单词和短语。

问题解决回答。根据问题提取实体，结合返回设计流程。

### 1.1 关键技术查询

输入关键技术术语，返回得到相对丰富的查询结果。

结果内容包括：

* 关键技术本身介绍

从网页如维基百科摘取半结构化文档信息进行实体对齐等方式作为介绍。

* 关键技术主题分类

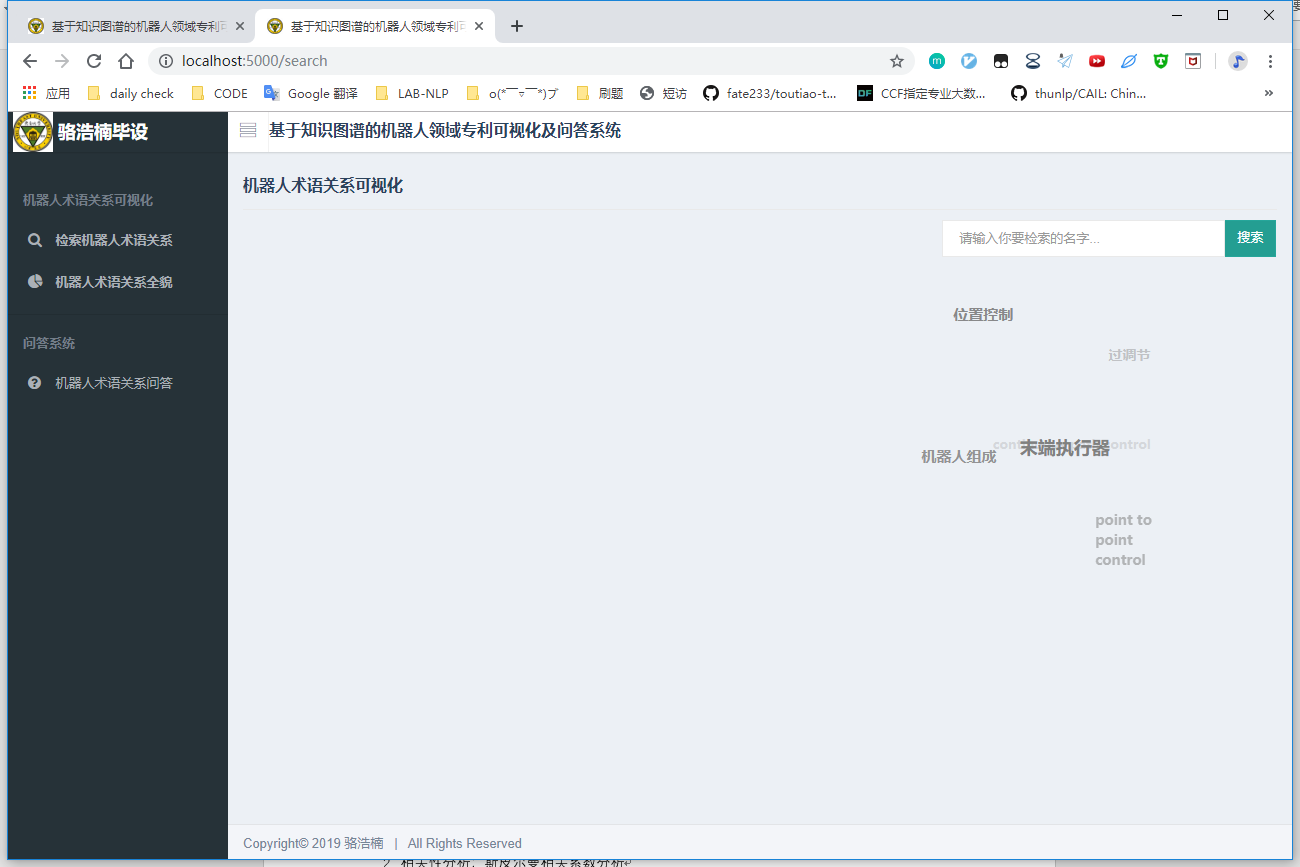
对于语料中目前得到的技术信息进行主题提取和分类，分类算法可采用深度学习方式不断进行优化。

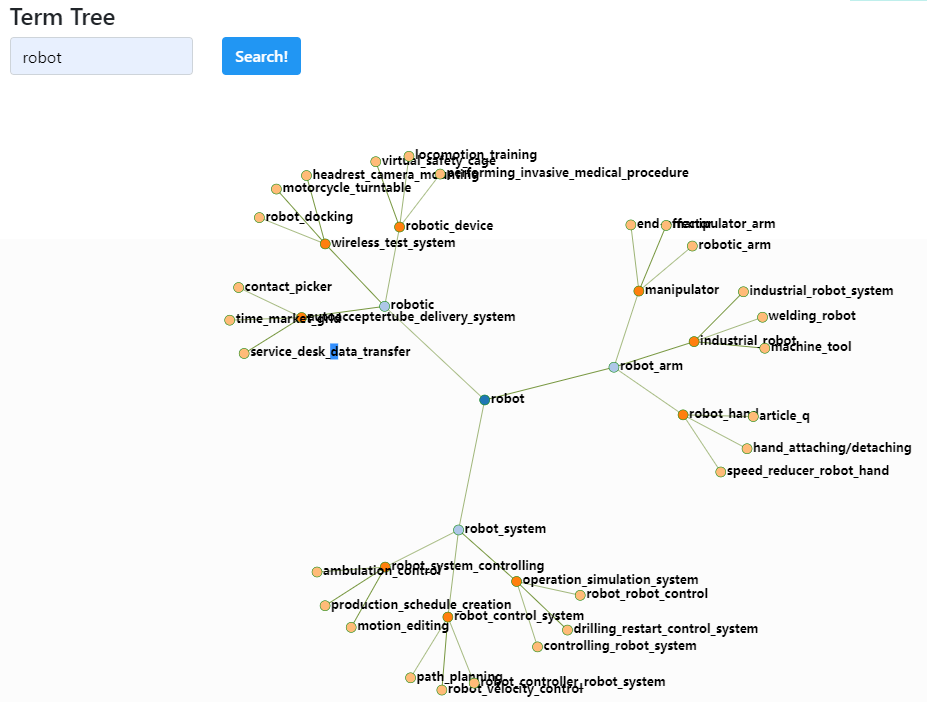
* 相关专利发展演变趋势

通过时间维度对相关专利进行梳理，形成按照时间顺序的相关专利发展脉络，揭示该技术到今天为止的发展变化趋势。

* 相关专利引用关系

通过引用维度对相关专利进行梳理，形成专利引用树，揭示相关技术专利的发展变化情况。





### 1.2 关键技术之间的相似度查询

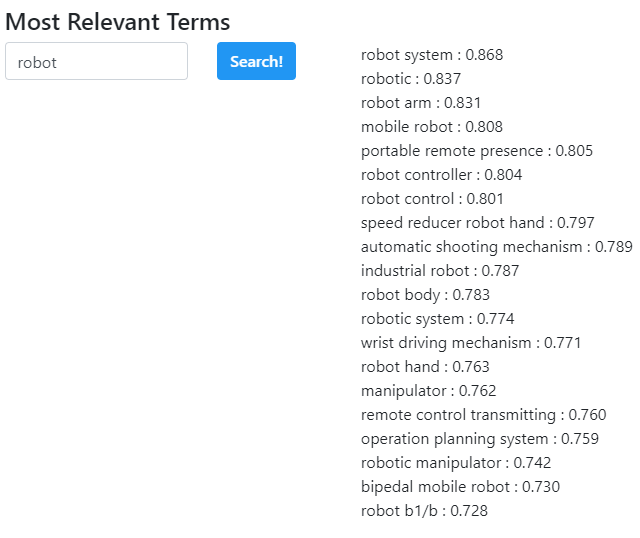
输入两个关键技术，返回得到相似度。



实现方式：基于在整个语料库中查询词频共现情况，在线计算，得到相似度结果输出。

* 单词的相似度，可直接通过余弦相似度计算得到；
* 短语的相似度计算，可通过识别得到的短语实体，通过返回文档重新生成词向量的方式，进行相似度计算。

### 1.3 技术术语推荐



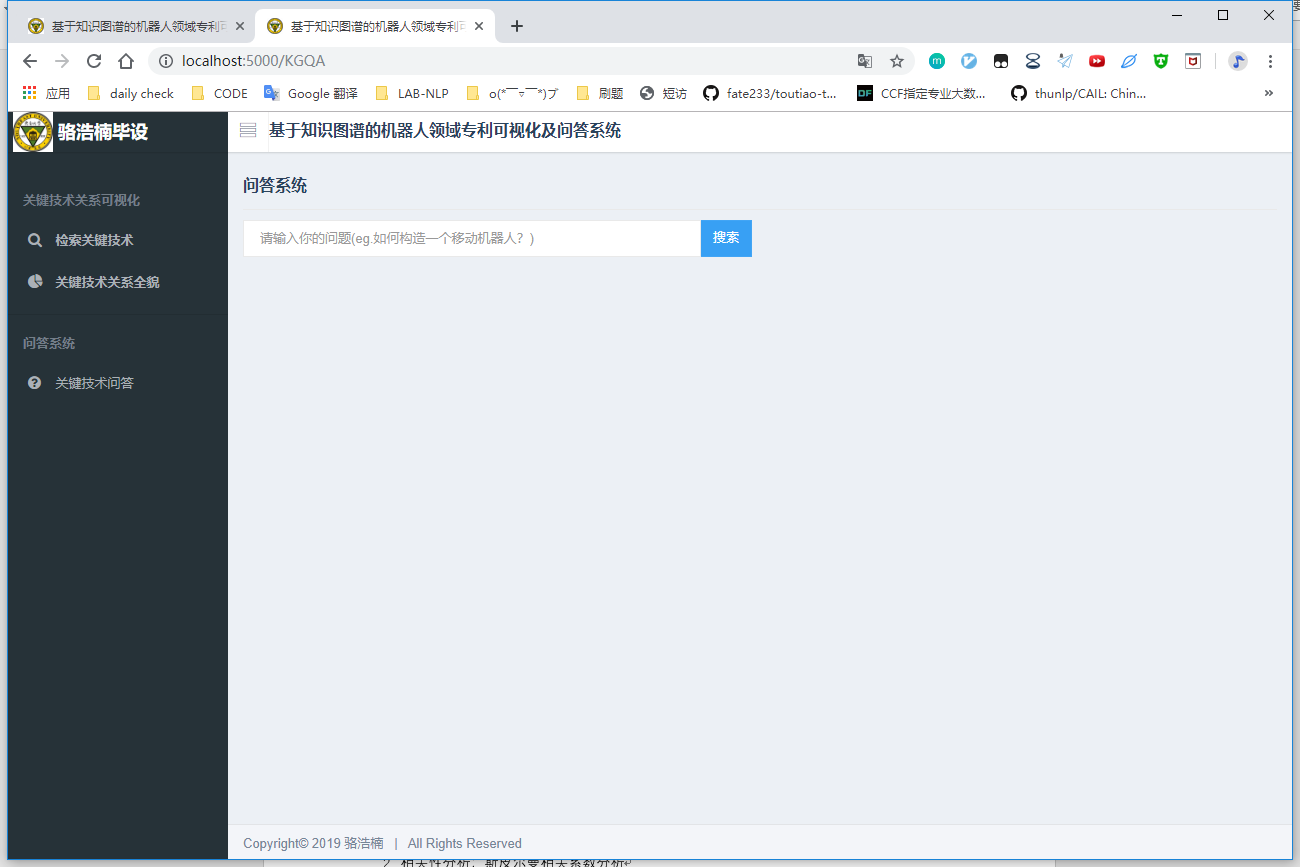
根据相关专利和技术相似度的计算结果，对相关技术进行推荐。

实现方式：在后台预先计算当前提取得到的术语，每个术语最相关的其他技术术语词，并选择阈值进行存储。通过查询索引，返回每个词以及相似度。

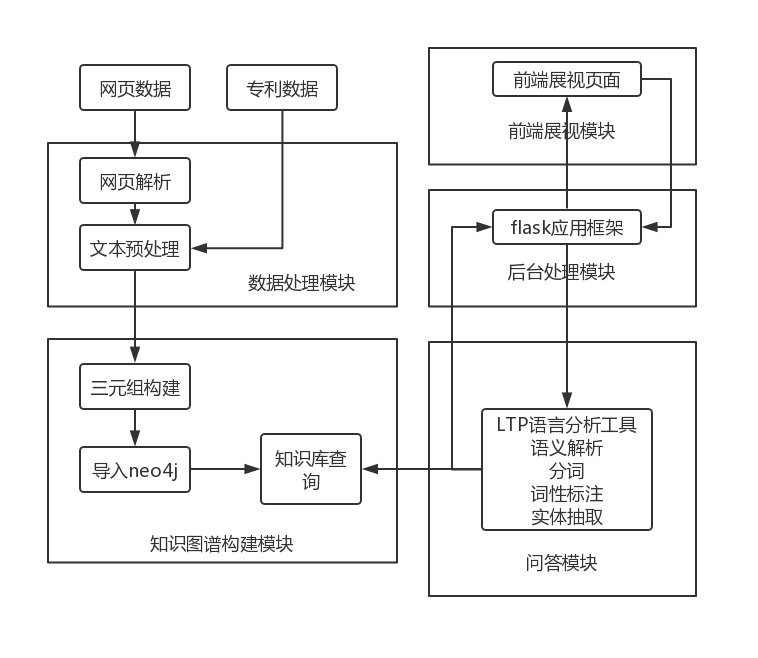
### 1.4 问题解决回答

输入问题文本，通过NER对其中的关键问题词进行识别，通过对识别得到的相关术语进行检索，返回得到关键技术查询结果。

通过文本生成返回得到该问题的参考解决方案。方案主要包括相关技术的专利展示。



## 2 机器人专利知识图谱构建过程



* 数据处理模块：

将网页数据和专利数据进行分词词性标注和去除停用词等预处理工作。

* 知识图谱构建模块：

将预处理得到的数据生成三元组，导入neo4j以供知识库查询。

* 后台处理模块：

使用python的flask应用框架，将知识图谱和前端连接结合。

* 前端展示模块：

展示机器人关系术语和知识图谱查询等页面。

* 问答模块：

将前端用户输入的问题，经过后台处理模块之后，进行分词和实体抽取等工作，进入知识库进行查询。

