

该项目使用基于 bert 的 OpenKE 工具从浙大校史中提取到了大量人物三元组信息。文本信息来源于《名流浙大(百年求是)》、《图说浙大》、《浙江大学图史》。具体流程如下：

首先使用了 OpenKE 的 NER 功能，即命名实体识别功能，从非结构化文本中提取得到大量实体信息。例如，对于文本

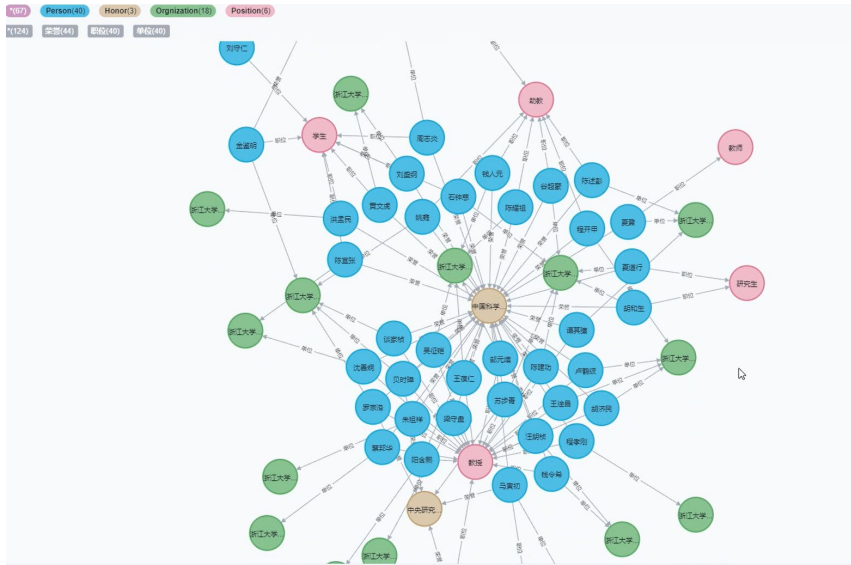
“作为乡人和被器重的留洋科学家，竺可桢很快被任命为浙江大学的校长，因为他来出任浙大校长，地缘、人缘等都一应俱全。”

我们可以在这句话中抽取到人名“竺可桢”、机构名“浙江大学”、职位“校长”“科学家”等。接着对所有文本中的识别到的实体进行了记录。

然后使用了 OpenKE 的 RE 功能，即关系抽取功能，从非结构化文本中抽取实体间的关系。我们已经知道有这些实体：人名“竺可桢”、机构名“浙江大学”。那么它们之间的关系是怎样的呢？

我们将每条句子中出现的实体两两配对，进行关系预测，如模型认为“竺可桢”和“浙江大学”是雇员和单位的关系，“竺可桢”和“校长”是职位的关系，同时模型还给出了该预测的可信度，我们将抽取到的关系对按可信度排序，取高于 0.9 的部分，对于结果进行了部分人工辨别，得到了基于浙大校史人物的关系三元组。

可视化结果



查询所有获得中国科学院院士荣誉的校友，可以输入查询语句：

```
match p=(n:Honor{hon:"中国科学院院士"}-[r:荣誉]-()) return p;
```

简单 QA

