**清华大学自然语言处理与社会人文计算实验室**

**共享资源协议**

**甲方：清华大学自然语言处理与社会人文计算实验室**

**乙方：（此处填写申请人所在单位）**

**协议内容：**

1. THULAC面向国内外大学、研究所、企业以及个人研究者免费开放源。乙方提交该协议表示同意：（1）该资源或程序库不用于商业目的；（2）不将资源或程序库扩散给第三方。
2. 为了增进同行交流，促进自然语言处理技术的发展，甲方将相关自然语言处理算法代码和模型资源免费共享给乙方，列表如下：

|  |  |
| --- | --- |
| **共享内容** | **详细内容** |
| 算法源码 | THULAC中文分词和词性标注训练算法、测试算法 |
| 模型资源 | 1-中文分词模型，由北京大学人民日报分词语料库训练产生。 |
| 2-中文分词与词性标注联合模型。由多个分词语料库训练数据联合学习产生。 |
| 3-中文分词与词性标注联合+词表。更加复杂精确的模型，并含有多个大规模词表。 |

1. 乙方同意在发表论文或申报成果时声明“使用了清华大学THULAC工具”，并按如下格式引用：

中文： 孙茂松, 陈新雄, 郭志芃, 张开旭, 刘知远. THULAC：一个高效的中文词法分析工具包. 2016.

英文： Maosong Sun, Xinxiong Chen, Zhipeng Guo, Kaixu Zhang, Zhiyuan Liu. THULAC: An Efficient Lexical Analyzer for Chinese. 2016.

1. 甲方将继续丰富和完善相关功能模块，为了能够给乙方及时提供最新资源和程序库，乙方需要完整提供如下信息：

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **姓名** | **机构名称** | **职务** | **电话** | **电子邮件** | **机构所在国别地区** | **网站** | **通讯地址** | **邮编** |
|
| **乙方负责人** |  |  |  |  |  |  |  |  |  |
| **乙方联系人** |  |  |  |  |  |  |  |  |  |

乙方负责人签名：

年 月 日