

汉语开放依存句法树库(Chinese Open Dependency Treebank, CODT)

共享协议 (2021 年 3 月版本)

甲方：苏州大学人类语言技术研究所 (<http://hlt.suda.edu.cn>)

乙方：[研究机构或公司的名字]

CODT 项目（读音：[kəʊ'di:ti:]）旨在持续标注高质量、覆盖多领域多来源文本的大规模依存句法树库，以促进句法分析技术的广泛应用。为了促进句法分析和领域移植技术的进展，甲方将 CODT 免费共享给乙方，用于科学研究。

CODT 主页：<http://hlt.suda.edu.cn/index.php/CODT>

特别说明

- 1) 请您所在课题组、实验室、科研团队负责人打印本协议，填写相关信息并签字后，**使用官方邮箱**将扫描版发送给李正华(zhli13@suda.edu.cn)。
- 2) CODT 只免费提供给“高校、科研院所”用于科学研究使用。如果用于商业产品、或者以公司名义使用，请联系李正华购买。
- 3) 如果公司希望共享数据，请先发邮件联系李正华说明情况，需酌情考虑。
- 4) 未来我们还会继续共享更多数据，旧的数据也可能发生更新（修改错误等）。如果需要新的共享数据，通常不需要重新签署协议（除非协议内容有所更新），只需使用**官方邮箱再次**发邮件索要新的数据。对于新的共享数据，即使无需重新签署协议，乙方仍需遵守以下协议条款。
- 5) 如果将来发生一些变化，如 CODT 改名、主页网址变化、甲方名称改变等，不影响树库共享的持续性，乙方仍需遵守以下协议条款。

CODT2.0 数据的详情信息如下表：

	BC	PB	PC	FIN	LEG	ZX
Train	26.1K	5K	6.8K	5K	5K	1.5K
Dev	1K	1.3K	1.3K	1K	1K	0.5K
Test	2K	2.5K	2.6K	1.8K	1.9K	1K

本着互相促进及尊重知识产权的原则，乙方必须遵守下述条款：

- (1) 免费共享 CODT 后，仅用于学术研究。如果用于商业产品、或者以公司名义使用，请联系李正华购买。
- (2) 该资源仅在签署协议的组织内部使用，不以任何方式将资源扩散给第三方。例如申请共享数据的组织名为苏州大学人类语言技术研究所，并且所长张民老师签字，那么该数据仅限于人类语言技术研究所内部使用，不可以转给苏州大学计算机学院的其他课题组。如果申请共享数据的老师没有课题组或其他组织，也可以以个人名义申请共享

数据，那么数据仅限于该老师个人团队内部使用。通常签署协议的组织应该为从事相关研究的课题组，而不应该是学院或学校级别。

- (3) 高校或研究所获得 CODT 后，在和公司进行非商业性质的、以发表文章为目的的科研合作中，允许使用我们的数据。如果科研合作终止，除非取得我们的许可，并重新签署协议，否则公司不可以单独继续使用我们的数据。
- (4) **代码公开和实验结果可重现条款：**我们希望通过共享数据，推动汉语句法分析、领域移植方面的研究。代码公开和实验结果可重现非常重要，可以保证研究的良性发展。如果公开发表**句法分析或领域移植**相关论文（包括 Arxiv 文章），且实验中使用了我们的数据，则需要：
 - a) 需发布完整的实验代码，包括实验配置、实验日志等所有信息，保证可以重现出论文中的结果。如果由于特殊原因确实无法发布代码或无法发布完整代码，也请积极帮助后续研究者尽快重现方法和结果。
 - b) 如果在实验中使用其他人无法获得的额外资源，研究人员应同时报告未使用额外资源情况下的详细结果，使后续研究人员能够清楚地了解额外资源的影响。
 - c) 不能随意改变数据的划分（train/dev/test）和评价指标（UAS/LAS，对于局部标注的句子，分母只计算标注了核心词的词），以便不同工作之间的实验比较。如果由于特殊原因需要改变，请在论文中详细说明原因，并尽可能汇报正常设置的结果。
 - d) 需尽可能与前人的结果进行适当、公平比较，促进研究圈子良性发展。我们计划在 CODT 主页上及时整理和更新在 CODT 上的结果。
- (5) 如果文章或其他成果中使用到 CODT，请引用至少一篇 CODT 主页上指定的论文。

乙方负责人需完整、清晰填写如下信息。

姓名：

职务或职称：

电话：

电子邮件：

通讯地址（邮编）：

乙方负责人签名：

年 月 日