



清华大学知识计算开放平台 (THUKC) 简介

2019年1月21日



清华大学人工智能研究院
知识智能研究中心
Knowledge Intelligence Research Center

开展理论研究。研究支持鲁棒可解释人工智能的大规模知识的表示、获取、推理与计算的基础理论和有效方法。

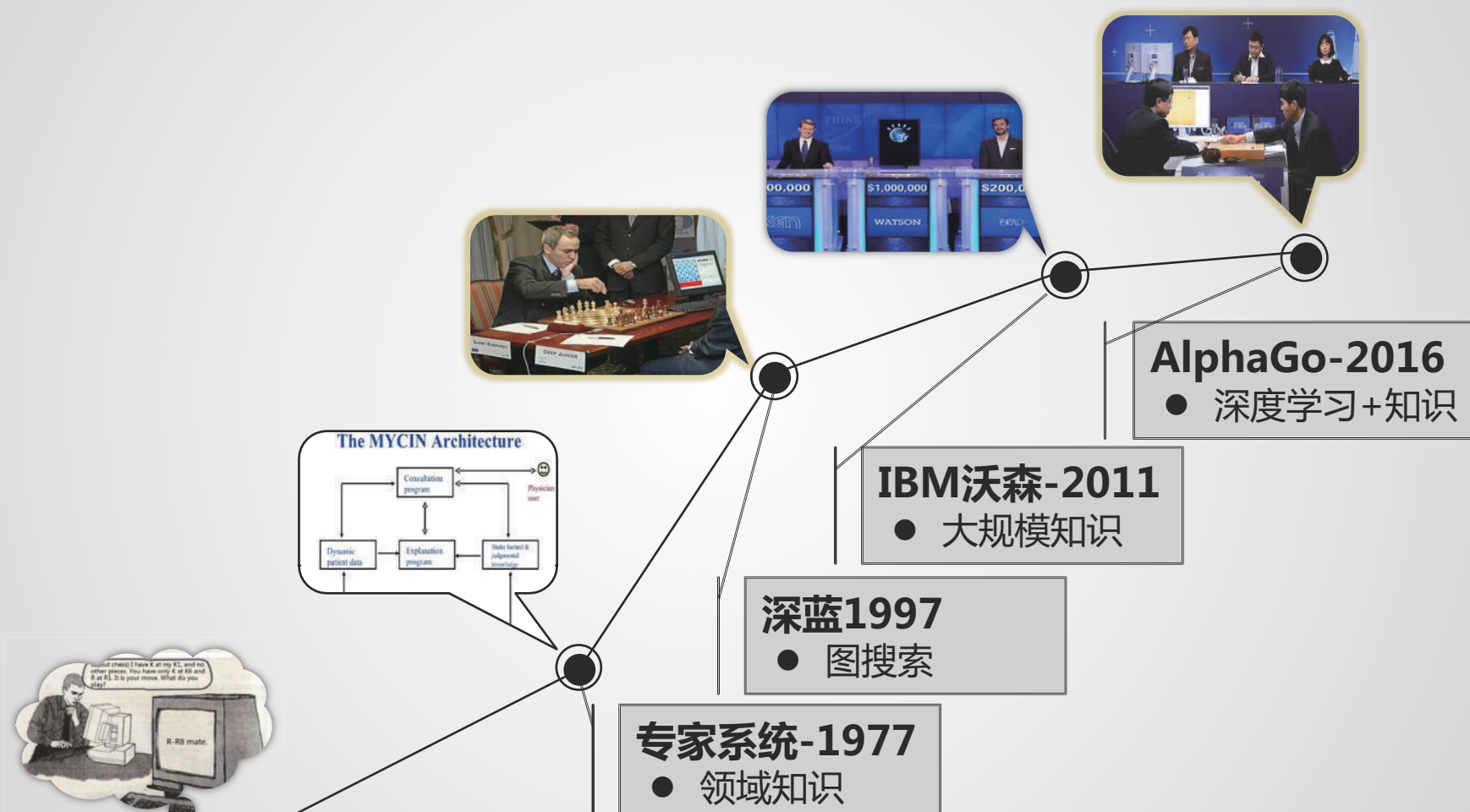
构建知识平台。建设中英平衡的、涵盖语言知识、常识知识、世界知识、认知知识与典型行业知识的大规模知识图谱，建成知识计算服务平台。

促进交流合作。经常性地组织知识智能有关学术活动，增进学术交流；普及知识智能技术，促进产学合作。

人工智能七十年

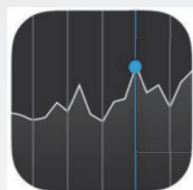


清华大学人工智能研究院
知识智能研究中心
Knowledge Intelligence Research Center



数据与知识融合是人工智能学科发展必然趋势

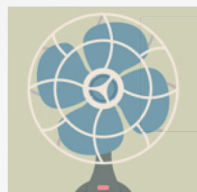
人工智能任务渴求知识



空调厂商股票可能大涨



行业知识



我们需要想办法降温



世界知识



这个夏天温度很高很热



常识知识

这夏天就像烤箱

语言知识



The summer is like an oven

对语言图像的深层理解需要复杂知识的支持

国际著名知识图谱



清华大学人工智能研究院
知识智能研究中心
Knowledge Intelligence Research Center

图谱名称	启动时间	语言	类型	规模
CYC	1984	英语	手工, 常识知识	百万级
WordNet	1985	英语	手工, 语言知识	十万级
ConceptNet	1999	英语	手工, 常识知识	千万级
Freebase	2005	英语	自动, 世界知识	十亿级
DBPedia	2007	英语	自动, 世界知识	十亿级
WikiData	2012	英语为核心	自动, 世界知识	千万级
BabelNet	2012	英语为核心	自动, 语言+世界知识	十亿级

我国人工智能发展需要中文核心知识的支持

知识智能研究在清华



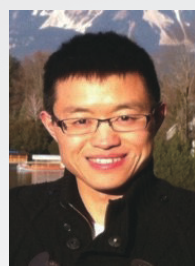
孙茂松



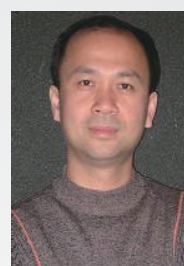
朱小燕



李涓子



唐杰



许斌



刘洋



黄民烈



刘知远

自然语言处理实验室 (THUNLP)：由孙茂松教授、刘洋副教授、刘知远副教授带领，主要从事中文信息处理、社会计算和知识图谱研究。最近在孙茂松教授牵头的973项目支持下，提出融合知识的统一语义表示框架和知识指导的自然语言处理的学术思想，取得一系列原创成果。

知识工程实验室 (THUKEG)：由李涓子教授、唐杰教授、许斌老师带领，主要从事知识工程、社会网络分析研究。最近研制的科技知识挖掘系统AMiner和跨语言知识图谱XLORE，获得国内外广泛关注与使用，曾获北京市科技进步一等奖，中国人工智能学会科技进步一等奖。

交互式人工智能实验室 (THUCoAI)：由朱小燕教授、黄民烈副教授带领。主要从事机器学习与自然语言处理应用基础研究，主要研究方向包括深度学习、强化学习、问答系统、对话系统、情感理解、逻辑推理、语言生成等。最近探索将常识知识融入对话系统与文本生成，取得一系列原创成果。



清华大学知识计算开放平台



「XLORE」是融合中英文维基、法语维基和百度百科，对百科知识进行结构化和跨语言链接构建的多语言知识图谱，是中英文知识规模较平衡的大规模多语言通用知识图谱。



「OpenHowNet」是一个以词语的概念为描述对象，以“义原”为基本语义单位，以揭示概念与概念之间以及概念所具有的属性之间的关系为基本内容的常识语言知识图谱。



「AMiner」是一个研究者学术搜索类网站，能够为计算机科学相关领域的研究者提供更全面的领域知识和更具针对性的研究话题和合作者信息，为科研人员提供信息获取和帮助。



「THUATIR」以AMiner全球科技情报大数据挖掘服务平台为基础，聘请领域顶级专家作为顾问，采用人工智能自动生成技术，汇集而成的科技领域洞察报告。

世界知识：XLORE



清华大学人工智能研究院
知识智能研究中心
Knowledge Intelligence Research Center



聚力两大中文百科
中英文平衡的图谱



更丰富的语义关系
基于isA关系验证



多种查询接口
助力第三方使用



16,284,901 实体



2,466,956 概念



446,236 关系



1亿+ 访问



160万+ 年API调
用



53 国家,
113 国内城市

语言、常识知识：OpenHowNet



清华大学人工智能研究院
知识智能研究中心
Knowledge Intelligence Research Center



首次将知网（HowNet）核
心数据开源



世界独创的基于义原的语
义描述体系



提供丰富的调用接口，方
便查询使用



229,767个中英文词条



35,202个概念

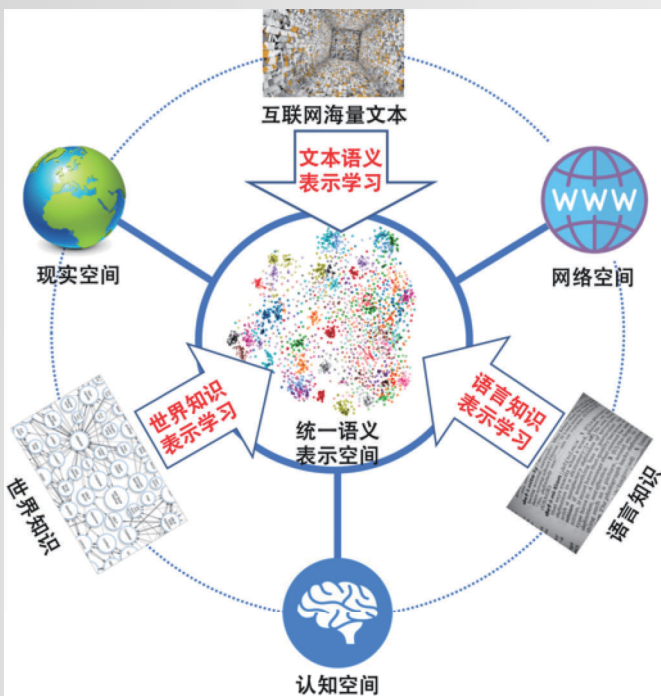


2,196个义原

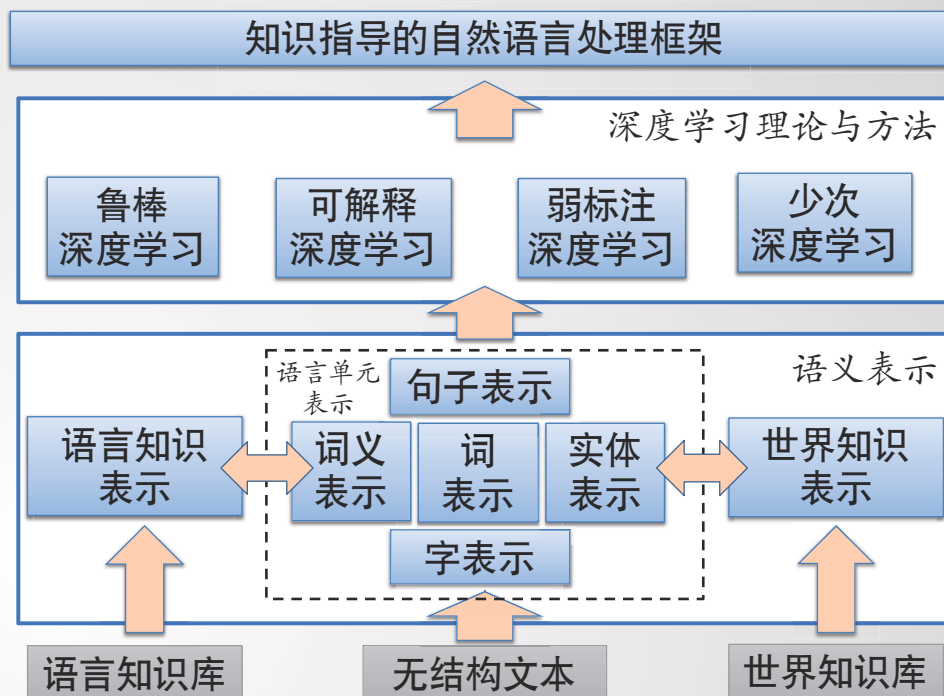
30

构建时间近30年

知识计算框架



融合知识的统一语义表示框架

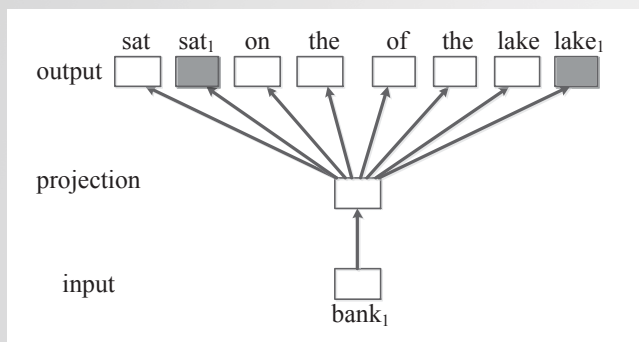


知识指导的自然语言处理框架

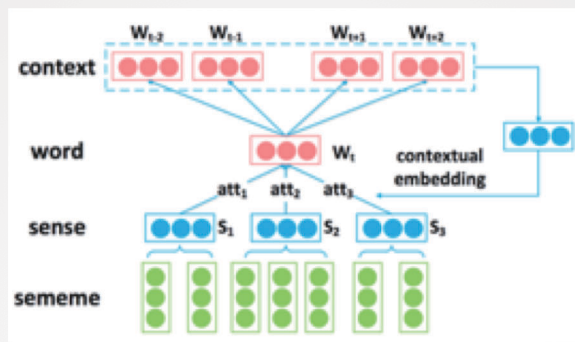
<p>≡ OpenKE</p> <p>An Open-Source Package for Knowledge Embedding (KE)</p> <p>● Python ★ 715 🍴 253</p>	<p>≡ OpenNE</p> <p>An Open-Source Package for Network Embedding (NE)</p> <p>● Python ★ 697 🍴 246</p>	<p>≡ OpenNRE</p> <p>An Open-Source Package for Neural Relation Extraction (NRE) implemented in TensorFlow</p> <p>● Python ★ 1.1k 🍴 416</p>
<p>≡ KRLLPapers</p> <p>Must-read papers on knowledge representation learning (KRL) / knowledge embedding (KE)</p> <p>● TeX ★ 455 🍴 101</p>	<p>≡ NRLPapers</p> <p>Must-read papers on network representation learning (NRL) / network embedding (NE)</p> <p>● TeX ★ 1.5k 🍴 468</p>	<p>≡ GNNPapers</p> <p>Must-read papers on graph neural networks (GNN)</p> <p>★ 1.4k 🍴 202</p>

融合语言知识的语言表示学习

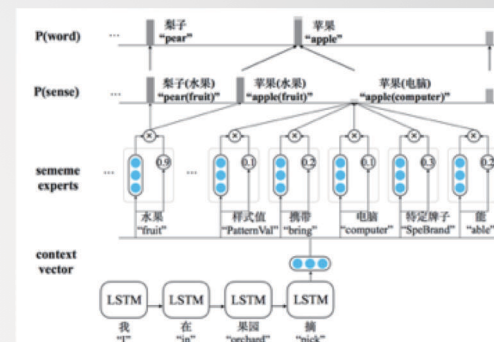
融合WordNet词义知识、HowNet义原知识等，实现更加鲁棒有效的语言单元表示学习，并可以反过来进行义原知识预测等



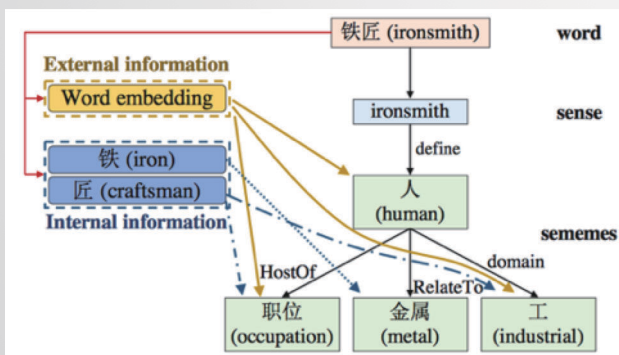
融合WordNet词义知识的
词义表示学习(EMNLP 2014)



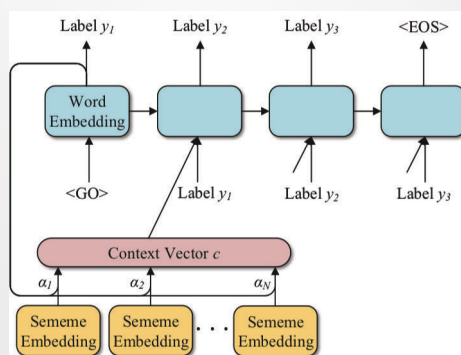
融合HowNet义原知识的
词义表示学习(ACL 2017)



融合HowNet义原知识的
句子表示学习 (EMNLP 2018)



基于词表示的义原知识预测
(IJCAI 2017, ACL 2018)



融合HowNet义原知识的
词典扩展 (AAAI 2018)

INTERFACE 机器之心
智能系统系列讲

主题：
在深度学习时代
用HowNet搞事情

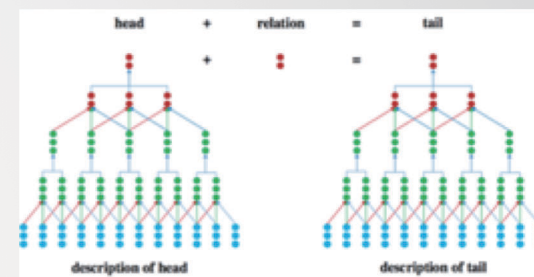
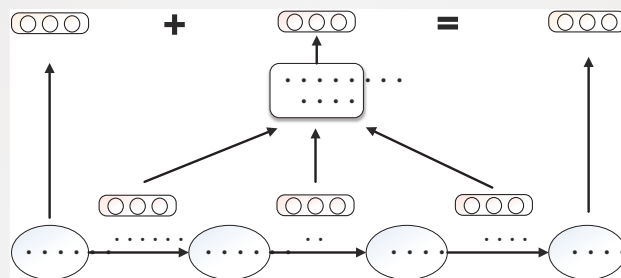
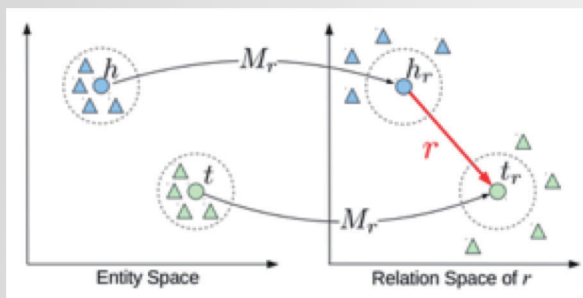
分享嘉宾：刘知远
清华大学计算机系副教授、博士生导师

时间：2018.05.26
地点：北京市海淀区中关村鼎好a座三楼主科技路演厅

DEEP LEARNING
NEURAL NETWORKS
HOWNET

考虑复杂结构的世界知识表示学习

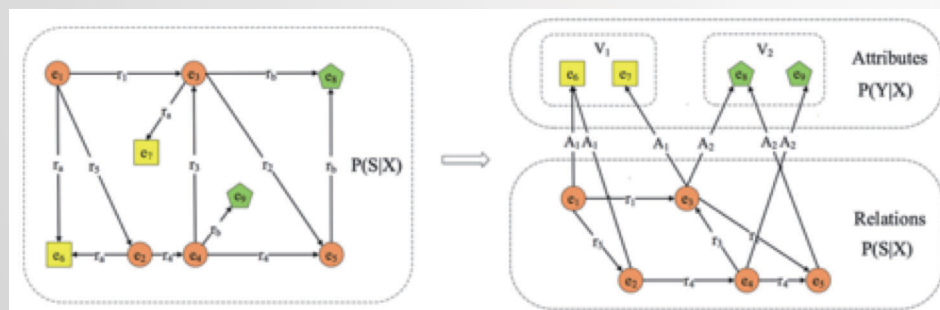
利用知识图谱和实体描述、类别和图像等外部信息，实现高效知识表示学习，建立低维语义表示空间与世界知识的关联



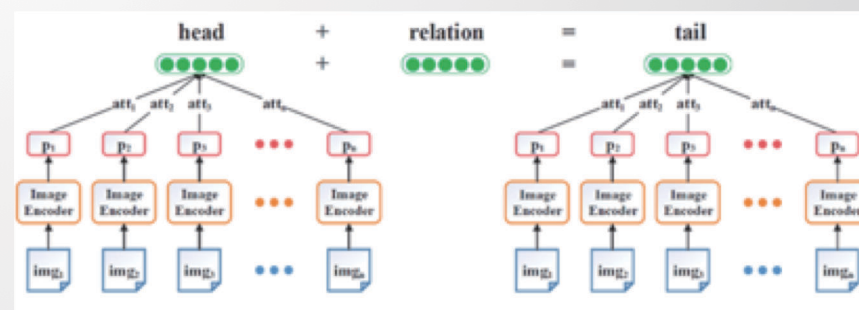
考虑复杂关系类型的知识表示
TransR (AAAI 2015)

考虑关系路径的知识表示
PTransE (EMNLP 2015)

考虑实体描述信息的知识表示
DKRL (AAAI 2016)



综合考虑实体、属性与关系的知识表示
KR-EAR (IJCAI 2016)



考虑实体图像信息的知识表示
IKRL (IJCAI 2017)

知识计算工具包



清华大学人工智能研究院
知识智能研究中心
Knowledge Intelligence Research Center



<https://github.com/thunlp/OpenKE>

- 工具包：统一接口，包括 TransE、TransH、TransR、TransD、RESCAL、DistMult、Hole、ComplEx等算法的高效实现
- 表示模型：面向WikiData和Freebase两大通用KG全量数据的预训练好的知识表示模型下载



<https://github.com/thunlp/OpenNRE>

- 工具包：统一接口，基于TensorFlow和PyTorch两个通用深度学习平台，实现不同的Embedding、Encoder、Selector、Classifier模块，方便用户使用和扩展
- 神经网络关系抽取算法：RNN、CNN、PCNN、+ATT等

科技知识：AMiner & THUAITR



清华大学人工智能研究院
知识智能研究中心
Knowledge Intelligence Research Center



科技知识挖掘平台



学术排行榜



开放学术图谱(OAG)



AI实验室



学者迁移图



会议汇总



人工智能研究系列报告

2019年公开报告



1、人工智能之数据挖掘



2、人工智能之知识图谱

2018年公开报告



1、自动驾驶与人工智能研究报告 (基础版)



2、机器人研究报告 (基础版)



3、区块链基础理论与研究概况



4、行为经济学与人工智能研究报告



5、机器翻译研究报告



6、通信与人工智能研究报告



7、自动驾驶与人工智能研究报告



8、自然语言处理研究报告



9、计算机图形学研究报告



10、超级计算机研究报告



11、3D打印研究报告



12、智能机器人 (前沿)



13、人脸识别研究报告



14、人工智能芯片研究报告



用知识为AI赋能



清华大学人工智能研究院

知识智能研究中心

Knowledge Intelligence Research Center

ai.tsinghua.edu.cn/kirc/