



古籍文献智能化及其API调用

——以sou-yun.com、cnkgraph.com为例

陈逸云

2024年7月

成果简介与演示

搜韵网 (<https://sou-yun.cn>)

- 2009年上线至今 (年500万终端5亿浏览量)
- 赋予计算机韵书、对仗、诗词格律之类的知识，建立专业诗词平台：
 - 从字、词、篇三个角度，提供辅助创作服务：用韵，用典，遣词，对仗，格律校验
 - 针对人文或诗词学者，提供专业资料库及在线工具：诗词库 (115万首)、唐宋文学编年地图、出处及全诗自动笺注功能

知识图谱网 (<https://cnkgraph.com>)

- 2019年上线至今
- 赋予计算机文史类知识，建立古典文献知识体系，建立平台协助学者从事研究：
 - 古籍库 (7千多部古籍)
 - 古今纪时转换、文献按时间索引
 - 人物交往、影响力研究
 - 古今地名查询
 - 文献自动笺注

应用示例：影响力研究

作品

- 出处、引用、诗评
- 句子字面相似性
- 步韵

作家

- 作品中常提及的前人
- 被同代或后世作品提及
- 在其它文献中被提及

城市

- 人才引进、人才输出
- 景观历代题咏

应用示例：笺校

- [诗集校雠](#)
 - 例：kanripo.org 石珙诗集
- 文影对照
- 自动笺注
 - 诗作
 - 文献
- 出处与化用分析

何谓文献智能化?



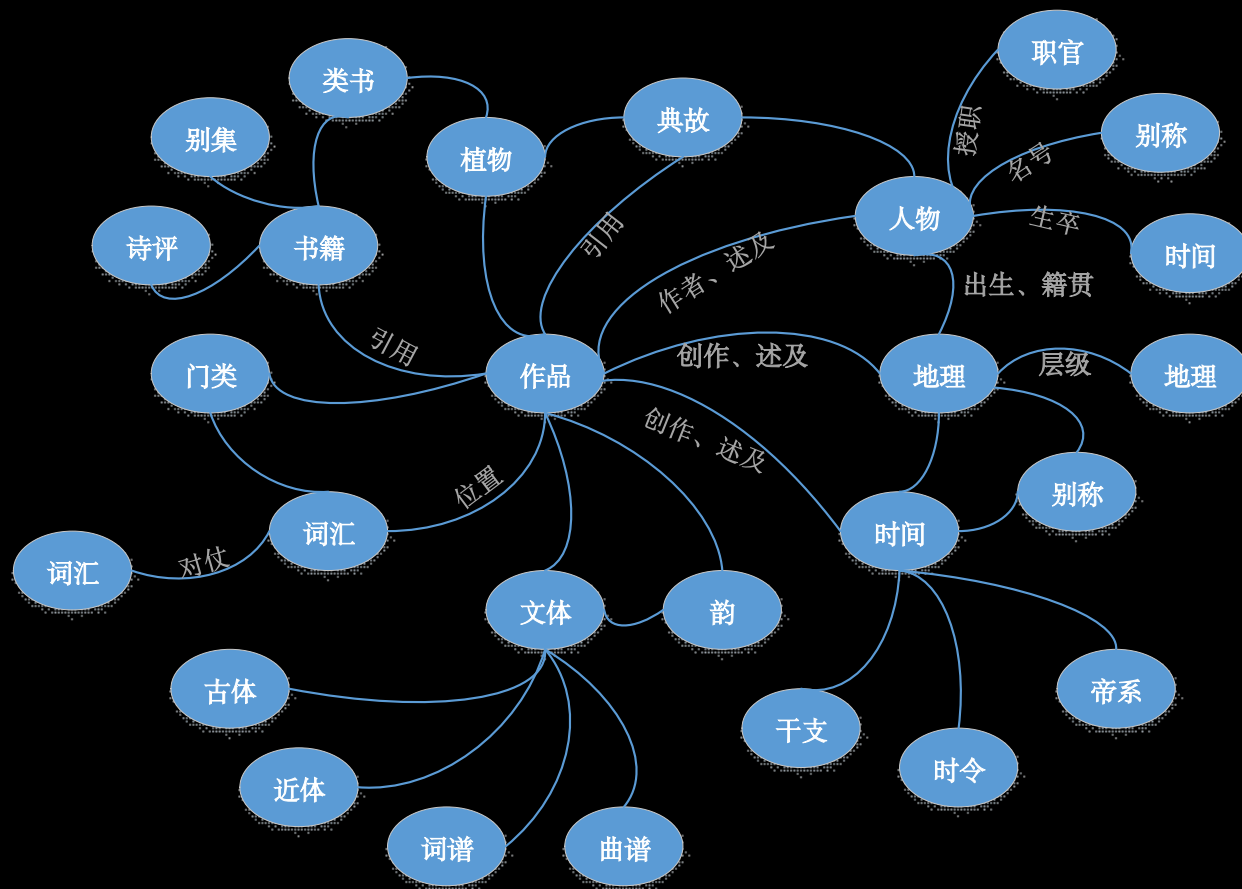
从文献中提取知识点



将知识点有机地关联在一起，形成知识网络



基于知识网络，从一或多个维度，开发各式应用，以辅助学习或研究工作



从关键词检索到语义检索

传统关键词检索（基于汉字编码匹配）

- 不知道“陶潜”、“陶渊明”、“五柳先生”、“陶靖节”等等是指同一人
- 不知道“元祐八年”与“大安九年”是同一年
- 没法根据上下文推断纪时信息，例如：
 - 宣和元年春正月戊申朔，日下有五色云。壬子，进建安郡王枢为肃王
- 北宋 · 欧阳修《去思堂会饮得春字（甲午四月颍州张唐公座上）》

语义检索（基于古籍文献知识图谱）

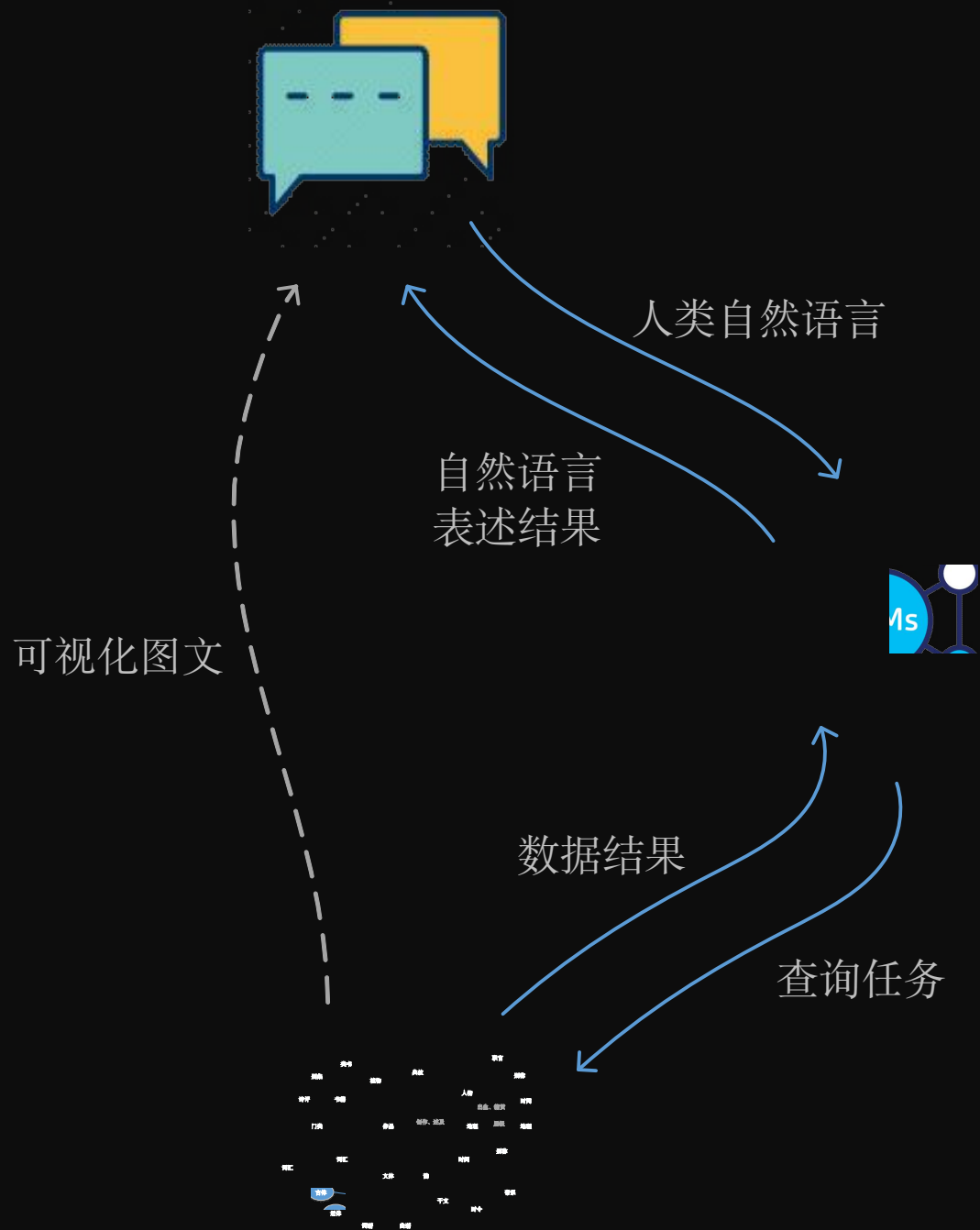
- 同一人物各种别称、省称，皆可一视同仁，例如，白居易有四十多种不同称呼，皆可在文献中标示出来
- 能根据上下文及作者信息，推断日期，并对日期先后，进行比较、排序
- 可按人物、日期、地点等信息对数据进行重组、编排

开放数据与接口

- 开放数据包
 - 诗文库: [下载 \(JSON格式\)](#)
- 接口网址
 - 开放接口网址:
<https://open.cnkgraph.com> (除古籍库功能外)
 - 古籍库接口网址:
<https://api.cnkgraph.com>
- 接口示例
 - PostMan格式示例: [下载](#)
- 更多内容详见: [“知识图谱”网开放资源](#)

古籍文献知识图谱与LLM

- LLM优势
 - 理解用户意图、流畅表达
- 古籍文献知识图谱优势
 - 知识深度与准确度
- 结合方法
 - 借助LLM理解用户意图，查询知识图谱后，或继续借助LLM用自然语言表述结果，或直接用可视化图表、文字的方式返回给用户



感谢！

诚挚欢迎各类交流合作，联系方式：cozychen@hotmail.com