

2.0--目前的信息抽取词库表说明

进行大规模的知识图谱的信息抽取测试

对需要抽取的信息进行列表化，方便后续进行处理

- 古诗词相关
 - 唐朝常见诗人
 - 宋朝常见词人
- 地理主题相关
 - 著名历史山岳
 - 湖泊
 - 古代景点、世界遗产
 - 中国的非物质文化遗产
 - 二十四节气
- 物理学相关
 - 专业名词
- 生物学相关
 - 专业名词
- 传统文化主题
 - 传统节日
 - 古书名
 - 历史人物--按朝代分类
 - 中国传统戏曲

目前依赖的词库

- 后续需要持续对词库进行扩充

```
# 古诗词相关
# 1、唐朝诗人
TangPoet_list = ["虞世南", "王绩", "上官仪", "卢照邻", "骆宾王", "李峤", "杜审言", "张若虚", "苏味道", "刘希夷", "杨炯", "崔融",
                 "宋之问", "沈全期", "贺知章", "陈子昂", "王涯", "王之涣", "张九龄", "张旭", "王翰", "孟浩然", "李颀",
                 "寒山", "綦毋潜", "常建", "刘长卿", "祖咏", "王湾", "李颀", "崔颢", "祖咏", "高适", "王维", "裴迪", "李白", "高适",
                 "崔曙", "刘繇", "储光羲", "常建", "刘长卿", "杜甫", "岑参", "皇甫冉", "元结", "司空曙", "钱起", "顾况", "张继",
```

"僧皎然", "戴叔伦", "韦应物", "卢纶", "李益", "李端", "孟郊", "权德輿", "王建",
"韩愈", "张籍", "薛涛", "李绅", "白居易",
"刘禹锡", "柳宗元", "马戴", "元稹", "贾岛", "姚合", "李贺", "杜秋娘", "马戴",
"杜牧", "薛逢", "李群玉", "陈陶", "李商隐",
"温庭筠", "李频", "郑畋", "韦庄", "司空图", "聂夷中", "皮日休", "韩偓", "杜荀鹤",
"崔涂", "韩雄", "许浑", "张乔", "张祜",
"秦韬玉", "刘方平", "夏侯审"]

2、宋朝词人

SongPoet_list = ["张孝祥", "刘克庄", "石孝友", "史达祖", "欧阳修", "卢祖皋", "毛滂", "宋自逊",
"李煜", "韩元吉", "杨万里", "方岳", "王安石",
"柳永", "张先", "周邦彦", "王安国", "袁去华", "赵佶", "黄机", "叶梦得", "吴潜",
"王建", "晁补之", "薛昭蕴", "卢炳", "李珣",
"朱淑真", "戴复古", "宋祁", "蔡伸", "贺铸", "文及翁", "林逋", "张抡", "李清照",
"孙光宪", "俞国宝", "程垓", "牛希济", "冯延巳",
"徐昌图", "王清惠", "王观", "朱服", "时彦", "王禹偁", "严蕊", "张志和", "秦观",
"晏几道", "蒋捷", "周芝", "万俟咏", "曹勋",
"黄庭坚", "严仁", "陈亮", "姜夔", "朱敦儒", "吴文英", "陆游", "张炎", "向子諲",
"李好古", "刘禹锡", "李白", "汪藻", "戴叔伦",
"李之仪", "顾夔", "侯蒙", "皇甫松", "李璟", "张昇", "黄升", "杨无咎", "黄裳",
"温庭筠", "房舜卿", "寇准", "潘阆", "岳飞", "叶清臣",
"范成大", "陈东甫", "王雱", "仲殊", "黄公绍", "白居易", "韩缜", "李元膺", "牛峤",
"刘辰翁", "刘过", "张元干", "韦庄", "张耒",
"黄孝迈", "李重元", "谢逸", "曹组", "唐婉", "辛弃疾", "赵长卿", "吕本中", "张舜民",
"晏殊", "苏轼", "王炎", "韩偓", "陈与义", "范仲淹"]

地理主题相关

3、著名山岳

FamousMountains_list = ["黄山", "庐山", "雁荡山", "泰山", "华山", "衡山", "恒山", "嵩山",
"峨眉山", "五台山", "普陀山",
"九华山", "龙虎山", "武当山", "青城山", "齐云山", "武夷山", "神农架"]

4、湖泊

Lakes_list = ["洞庭湖", "西湖", "太湖", "鄱阳湖", "洪泽湖"]

5、世界遗产

WorldHeritage_list = ["嘉峪关", "玉门关", "八达岭长城", "承德避暑山庄", "桂林山水", "草堂",
"武侯祠", "故宫", "莫高窟", "秦始皇陵", "孔林",
"布达拉宫", "丽江古城", "苏州园林", "颐和园", "大足石刻", "龙门石窟", "都江堰",
"云冈石窟", "安阳殷墟", "开平碉楼",
"乐山大佛", "良渚古城", "九寨沟", "武当山古建筑群", "乐山大佛", "天坛", "丽江古城",
"平遥古城", "开平碉楼", "福建土楼"]

6、二十四节气

SolarSeason_list = ["立春", "雨水", "惊蛰", "春分", "清明", "谷雨", "立夏", "小满", "芒种",
"夏至", "小暑", "大暑", "立秋", "处暑",
"白露", "秋分", "寒露", "霜降", "立冬", "小雪", "大雪", "冬至", "小寒", "大寒"]

物理学

7、物理专业名词

PhysicalTerms_list = ["扩散", "密度", "摩擦力", "凝固", "凝固点", "凝华", "能量", "内能",
"频率", "气化", "熔点", "热运动", "声波", "升华",
"万有引力", "液化", "压强", "原子", "紫外线", "质量", "重力", "重心", "正电荷",
"沸点", "晶体", "焦点", "电子",
"电荷", "电磁波", "电磁感应", "绝缘体", "磁场", "比热容", "热运动", "电压",
"波长", "超声波", "次声波", "杠杆",
"浮力", "惯性", "介质"]

```
# 生物学相关
# 8、生物名词
BiologicalTerms_list = ["吸收光谱", "乙酰辅酶A", "体腔动物", "动作电位", "活化能", "活动站
点", "主动运输", "适应", "适应性辐射", "腺嘌呤",
    "三磷酸腺苷", "藻类", "异源物种形成", "氨基酸", "羊膜", "厌氧", "类似结
构", "解剖学", "动物", "抗生素", "细胞凋亡",
    "蜘蛛学", "古菌", "人工选择", "无性繁殖", "天体生物学", "自身免疫", "自
养", "B细胞", "细菌", "噬菌体",
    "孢子", "干细胞", "类固醇", "叶黄素", "木质部", "免疫球蛋白", "感染",
    "胰岛素"]

# 传统文化相关
# 9、传统节日
TraditionalFestivals_list = ['小年', '盂兰节', '立春', '惊蛰', '立夏', '火把节', '腊八节',
    '大雪', '七夕', '霜降', '大暑',
    '小暑', '中元节', '春节', '地藏节', '立秋', '小雪', '鬼节', '立
    冬', '白露', '秋分', '芒种',
    '龙抬头', '大寒', '小寒', '处暑', '夏至', '祭祖节', '小满', '雨
    水', '春分', '除夕', '寒露',
    '中秋节', '谷雨', '元宵节', '阔时节', '冬至', '清明节', '寒食节',
    '重阳节', '翻经节', '清明', '姑姑节']

# 10、古书名
AncientBooks_list = ["史记", "汉书", "后汉书", "三国志", "晋书", "宋书", "南齐书", "梁书",
    "陈书", "魏书", "北齐书", "周书", "南史", "北史",
    "隋书", "旧唐书", "新唐书", "旧五代史", "新五代史", "宋史", "辽史", "金史",
    "元史", "明史", "诗经", "尚书", "礼记", "周易",
    "周礼", "仪礼", "论语", "孝经", "尔雅", "孟子", "公羊传", "谷梁传", "春秋左
    氏传", "山海经", "水经注", "资治通鉴", "梦溪笔谈",
    "本草纲目", "文心雕龙", "三国演义", "水浒传", "红楼梦", "西游记", "国语",
    "战国策", "黄帝内经", "难经", "伤寒杂病论",
    "神农本草经", "搜神记", "淮南子", "论衡", "皇庭经", "文史通义", "聊斋志异",
    "心经", "金刚经", "道德经", "茶经", "世说新语",
    "孙子兵法", "增广贤文", "百家姓", "老子", "庄子", "弟子规"]

# 11、历史人物
HistoricalFigures_list = ["关羽", "张飞", "张苞", "关兴", "关平", "卢植", "公孙瓒", "甘
    氏", "刘禅", "诸葛瞻", '诸葛亮', "刘备", "黄月英", "黄承彦",
    "诸葛瑾", "公孙越", "马腾", "韩遂", "曹操", "刘胜", "刘辩", "姜维",
    "孙权", "孙尚香", "糜氏", "糜芳", "糜竺", "庞德",
    "马谡", "祝融", "孙韶", "刘协", "董卓", "王允", "貂蝉", "丁原", "吕
    布", "刘表", "蔡氏", "蔡瑁", "蒯越", "张宝", "张梁",
    "袁绍", "袁术", "袁谭", "袁熙", "袁尚", "吴国太", "孙策", "大乔", "小
    乔", "周瑜", "曹嵩", "吕伯奢", "邹氏", "清河公主",
    "夏侯楙", "夏侯渊", "曹爽", "荀彧", "司马师", "司马昭", "司马炎", "曹
    仁", "曹纯", "曹昂", "超昂", "卞氏", "曹丕", "曹植", "孙坚"]

# 12、传统戏曲
TraditionalOpera_list = ["京剧", "越剧", "黄梅戏", "评剧", "豫剧", "昆曲"]
```