不同领域的词库构建

目前存在的问题：

- 词袋的难点在于目前人工构建的词袋较小，扩展起来较为麻烦。

解决方法：

- 尝试通过提取目前常用的输入法的词库进行热词、不同领域的词的扩展，
    - 优势：快速、高效的进行词库扩充

领域词汇表研究

词汇表定义：

- 词汇表是专业或技术词汇及其含义的列表。

参考：

- 维基百科--目录

分类

1. 文化艺术
    - 文化
    - 艺术
    - 小说和文学
    - 运动
    - 收藏和爱好
2. 地理和地方
    - 地方
3. 健康和健身
    - 药物、临床研究、精神病学、解剖学
4. 历史和事件
    - 历史、考古
5. 数学和逻辑
    - 数学领域、几何、
6. 自然科学和物理科学
    - 生物科学
        - 生物学、植物学、生态学、病毒学
    - 物理科学
        - 天文(气象)、化学、气候变化、地理、地质学、物理学
7. 哲学与思考

    - 哲学
8. 社会和社会科学
    - 商业、政治、法律
    - 语言和俚语
9. 技术与应用技术
    - 技术与应用科学
    - 计算和信息技术
    - 运输

领域含义说明

数据源：

- wikipedia

维基百科：内容/文化与艺术

文化，意思是“居住、培养或荣誉”，一般来说文化指人类活动，文化的不同定义反映了不同的理解理论或评价人类活动的标准。现在的人类学家使用该术语来指代人类对经验进行分类并以符号方式对其进行编码和交流的普遍能力。

艺术是文化的一个巨大分支，由许多创造性的努力和学科组成。从更广泛的意义上讲，艺术是创造力或想象力的表达，艺术也可以被理解为与创造力、美学和情感的产生有关。

文化

- 艺术
    - 文化
    - 视觉艺术
    - 表演艺术
- 美食
    - 食物
    - 饮食
- 娱乐
    - 节假日
    - 游戏
    - 体育
        - 球类
        - 格斗运动
        - 赛车运动
- 人文学科
    - 地区研究
    - 古典研究

维基百科：内容/地理和地点

地理学，是对地球及其特征、居民和现象的研究。地理研究的四大历史传统是空间分析、自然和人类现象的研究(地理作为分布研究)，区域研究(地方和地区)，人地关系研究和地球科学研究。尽管如此，现代地理学是一门包罗万象的学科，它首先寻求了解世界及其所有人类和自然的复杂性--不仅是物体在哪里，而且需要解释它们是如何变化和形成的。作为“人文科学与自然科学之间的桥梁”，地理学分为两个主要分支--人文地理学和自然地理学。

地理

- 地理学分支
  - 自然地理学
  - 人文地理学
  - 区域地理学
- 地理和地方

  地球、世界、水域、城市、社区、大陆、国家、沙漠、湖泊、地貌、山脉、海洋、人口稠密的地方、保护区、地下、城市、村庄
- 自然地理特征
  - 地貌
  - 水体
- 人造地理特征
- 世界
  - 按半球
  - 按文化区域
  - 按经济发展
  - 按大陆

维基百科：目录/社会和社会科学

社会是一群人组成的一个半封闭的系统，最简单的社会一词指一大群分享自己的文化和制度的人，改词起源于拉丁语，意指“与他人的友好关系”。社会的意义与被认为是社会的东西密切相关。隐含在社会的含义是：它的成员可能分享一些共同的关注和利益，共同的目标或共同的特征。社会科学是一组研究世界人类方面的学科。它们与艺术和人文科学的不同之处在于，社会科学倾向于强调在人文研究中使用科学方法，包括定量和定性方法。

社会科学

- 社会科学
  - 人类学、考古学、认知科学、传播研究、批判理论、文化研究、经济学
  - 教育、地理、历史、语言学、法律、政治学
- 社会
  - 社区
  - 社会发展
  - 社会机构

        - 家庭
        - 基础设施
        - 经济与商业
        - 教育
        - 文明社会
        - 政府与政治
- 社会网络

维基百科：目录/自然与物理科学

在科学中，自然科学一词是指研究宇宙的理性的方法，被理解为遵守自然起源的规律或规则。“自然科学”一词也区别于使用科学方法研究自然的领域与使用科学方法研究人类行为和社会的社会科学领域，并且来自使用不同方法论的形式科学，例如数学和逻辑。与生物科学相比，物理科学是自然科学和研究非生命系统的一个涵盖性术语。

生物学

- 解剖学、生物化学、植物学、细胞生物学、计算生物学、生态学、进化生物学、
- 基因学、组织学、免疫学、微生物学、生命起源、生理学、动物学

物理科学

- 天文学
- 化学
- 物理学
    - 原子、基本粒子、力、引力、质量、物质、光学、时间、声音、能量、经典力学

维基百科：内容/历史和事件

历史是对过去事件、社会和文明的解释。历史一词来自希腊语，“一个人的调查记录”。1911年的《大英百科全书》指出：“广义上的历史就是发生的一切，不仅是人类生活的所有现象，还有自然界的现象，它是发生变化的一切。”

历史

- 按时期划分
    - 史前
    - 古代历史
        - 古代东方
        - 古代西方
    - 中世纪
    - 文艺复兴
    - 早期现代史

    - 现代历史
    - 全球化
- 按地区划分
    - 古埃及
    - 古罗马
    - 拜占庭帝国
    - 奥斯曼帝国
- 按主题划分
    - 按领域划分
        - 艺术史
        - 商业历史
        - 地理学史
        - 科学史
        - 技术史
    - 历史科学
        - 考古学
        - 地质学
        - 生物学
        - 天文学

参考资料：

- THUOCL：清华大学开放中文词库
    - IT、财经、成语、地名 历史名人、诗词、医学、饮食、法律、汽车、动物
- 搜狗收入法-词库
    - 城市信息、自然科学、人文科学、社会科学、工程与应用科学、农林渔畜、医学、艺术、生活.....

- 百度输入法
  - 人文社科、生活百科、文化艺术、人名专区.....

- QQ输入法
  - 自然科学、社会科学、医学医药、生活百科.....

搜索算法的词库解码方案：

- [x] 搜狗输入法
  - scelParser
- [x] 百度输入法
  - bdictParser
- [ ] QQ输入法