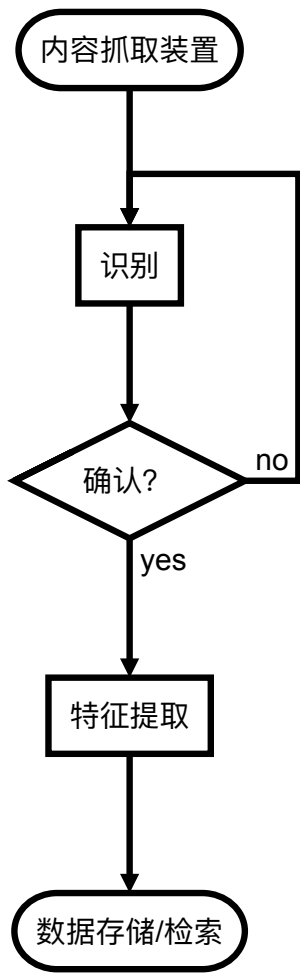


X实验 结构图



遇到问题

- ☐ 内容获取器
 - ☐ 爬虫?
- ☐ 词向量过拟合
 - ☐ 如何拆解内容（识别）

下载器/爬虫/内容获取

内容下载器

目标选择-> 时间， 地点， 事件， 人[?] 新闻

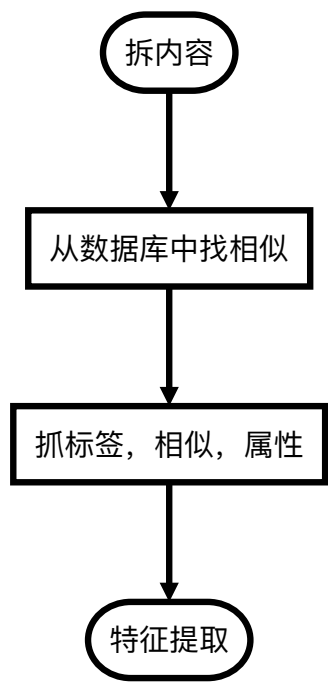
内容提取

内容清洗

1. 全部新闻

2. 压缩图片

识别器



神经网络训练内容的分类（无监督）

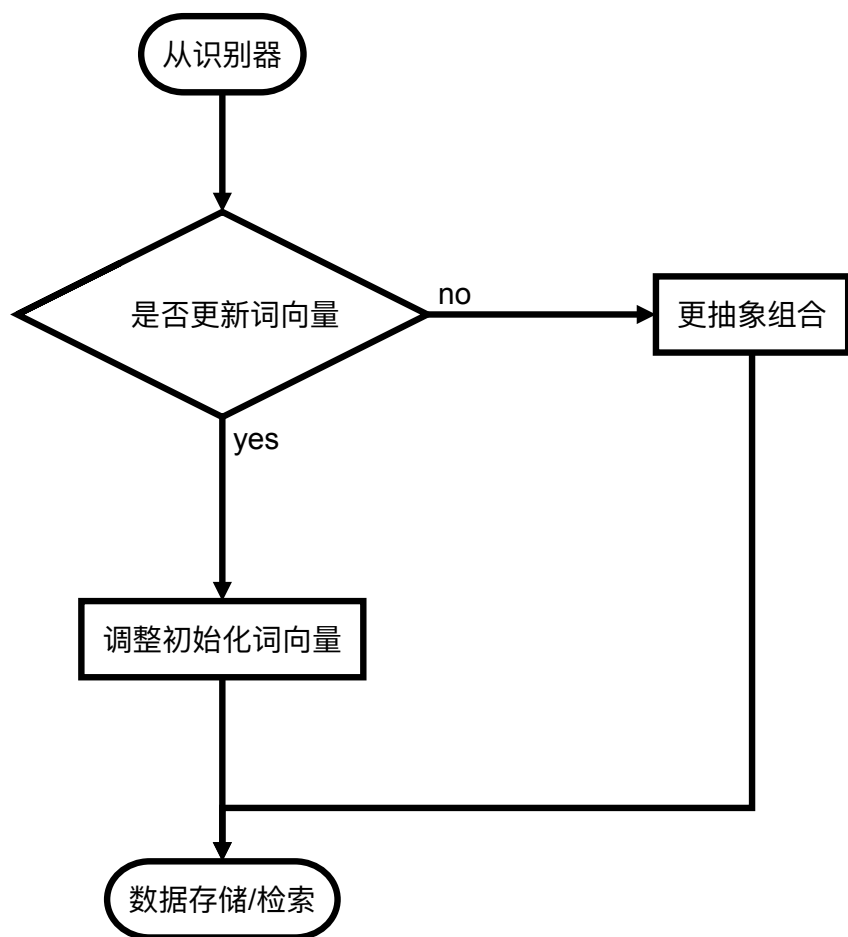
词向量的条件下

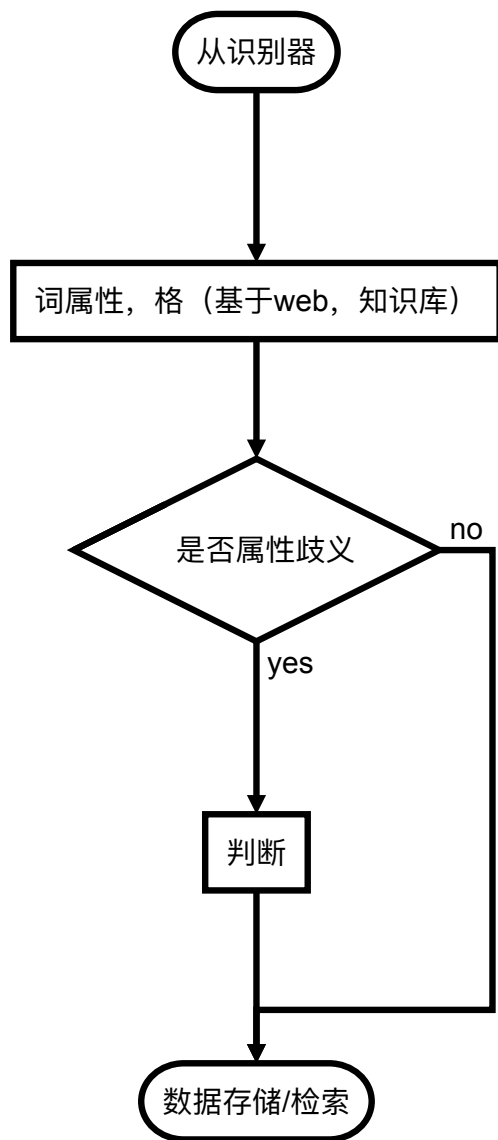
step1: 提取内容 | 提取描述（来源描述，html，docx）

step2: 拆分内容

特征提取器

- 词向量
- 词属性





数据存储

wikibase 知识库

结果的可视化

Wiki search :

文章

WDQS: wiki data query service

地图, 表, 等的可视化

支持**API**

- python
- bash

- rest