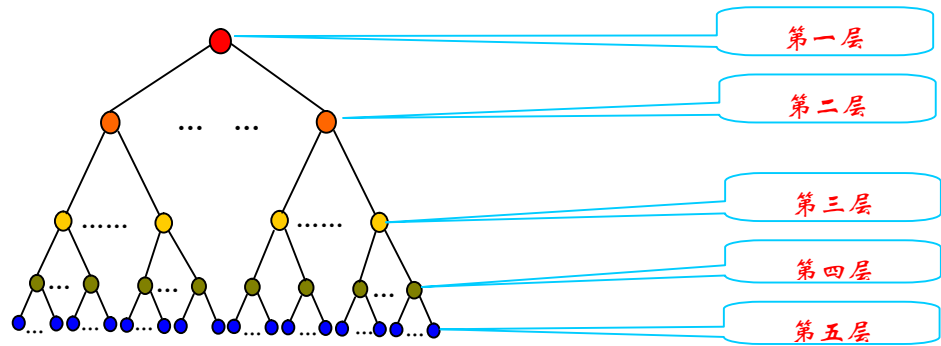


以为自然语言处理提供不同的服务，例如第四层的分类和第五层的分类在信息检索、文本分类、自动问答等研究领域得到应用。有研究证明，对词义进行有效扩展，或者对关键词做同义词替换可以明显改善信息检索、文本分类和自动问答系



统的性能。

词典《同义词词林》中保留下来的 39,099 条词语也保留了原有的分层结构，而新增的 36,267 条词语没有这样的结构。对于这些词，按照《同义词词林》的结构体系进行分类，工作量十分巨大。分类的某些环节可以使用机器自动完成，但是自动完成的结果不是很理想，各个环节主要还是依靠人工来完成。

三、编码

《同义词词林》只提供了三层编码，即大类用大写英文字母表示，中类用小写英文字母表示，小类用二位十进制整数表示。例如：“Ae 07 农民 牧民 渔民”，“Ae 07”是编码，“农民 牧民 渔民”是该类的标题。标题是由一个或者多个第四层的“段首（即每个段的第一个词）”组成。根据标题词可以知道小类有分成多少个第四级类，参见表 1。

表 1 词典结构示例

Ae07 农民 牧民 渔民	
农民 农夫 农人 农 庄稼人 庄稼汉 田父 泥腿子 农家 耕夫 老乡	<p>○第四级类 ○可分为若干个第五级类</p>
小农 个体农民	
佃农 佃户	
上中农 富裕中农	
* * 菜农 棉农 茶农 烟农 蔗农 花农 药农 林农	
雇农 贫农 下中农 中农 上中农 富农	
自耕农 半自耕农 集体农民 人民公社社员	