

牧民 牧人 牧工

渔民 渔翁 渔家 渔夫 渔父

为了使用上的方便，对于第四级和第五级的分类也需要编码。新增的第四级和第五级的编码与原有的三级编码和并构成一个完整的编码，唯一的代表词典中的出现的词语。如：

Ba01A02= 物质 质 素

Cb02A01= 东南西北 四方

Ba01A03@ 万物

Cb06E09@ 民间

Ba01B08# 固体 液体 气体 流体 半流体

Ba01B10# 导体 半导体 超导体

编码的方法说明如下：

第四级用大写英文字母表示，第五级用二位十进制整数表示。由于第五级的分类结果需要特别说明，例如，有的行是同义词，有的行是相关词，有的行只有一个词，可以分出具体的三种情况。在使用上，有时需要对这三种情况进行区别对待，所以有必要再增加标记来分别代表着几种情形。具体的标记参见表 2。

表 2 词语编码表

编码位	1	2	3	4	5	6	7	8
符号举例	D	a	1	5	B	0	2	= \ # \ @
符号性质	大类	中类	小类		词群	原子词群		
级别	第 1 级	第 2 级	第 3 级		第 4 级	第 5 级		

表中的编码位是按照从左到右的顺序排列。第八位的标记有 3 种，分别是“=”、“#”、“@”，“=”代表“相等”、“同义”。末尾的“#”代表“不等”、“同类”，属于相关词语。末尾的“@”代表“自我封闭”、“独立”，它在词典中既没有同义词，也没有相关词。

四、词典的完善

目前推出了《哈工大信息检索研究室同义词词林扩展版》的 1.0 版本，已经可以满足很多研究领域的应用。