# 协陪义种子词扩展算法

## 一、协陪义种子词

带领 带队 带 领导 领 率 率领 引 引导 拉 牵 伴同 伴 陪同 随同 陪伴 陪 **携** 搀 搀扶 扶 护送 送 帮 帮助 臂助 辅助 襄助 协同 协助

## 二、《知网》介绍

《知网》是一部比较详尽的语义知识词典，受到了人们普遍的重视。它有如下特点：

1) 每一个词可能含有多个概念，每一个概念的义原组成不同。

例如：词“携”包含两个概念A和B，概念A由一个基本义原“携带”构成，概念B由一个基本义原“引导”构成。

A：携 V bring|携带

B：携 V guide|引导

2) 词语的语义描述中各个义原并不是平等的，它们之间有着复杂的关系，通过一种专门的知识描述语言来表示。

在《知网》中，对一个概念的描述可以表示为四个特征：第一个基本义原、其他基本义原、关系义原、关系符号义原。

《知网》一共采用了1618个义原，这些义原既表达了词语的语义信息，也表达了概念之间的联系。

## 三、词汇语义相似度计算规则

规则1：协陪义种子词扩展实验中，只考虑动词的相似度计算；

规则2：对于一个汉词词语*W1*和*W2*，如果*W1*有*n*个概念*{w11, w12, w13, …, w1n}*，*W2*有*m*个概念{*w21, w22, w23,…, w2m*}，则*W1*和*W2*的相似度是它们之间概念相似度的最大值，计算如下：

规则3：概念相似度计算

其中，是可调节参数，且，；

规则4：每一部分相似度计算：

：第一基本义原相似度计算，即义原相似度计算；

：其他基本义原相似度计算，转换为两个基本义原集合的相似度计算；

：关系义原相似度计算

：关系符号相似度计算

规则3：义原相似度计算：由于所有的义原根据上下位关系构成了一个树状的义原层次体系，故通过语义距离计算相似度。假设两个义原在这个体系中的路径距离为，可以得到两个义原之间的语义距离：

其中表示两个义原，是在义原层次体系中的路径长度，是一个正整数，是一个可调节的参数。

注：词汇语义相似度计算规则引用自文献《基于《知网》的词汇语义相似度计算》。

四、协陪义种子词扩展规则

列举协陪义种子词在《知网》中的概念描述（见附件“协陪义种子词-<知网>.xlxs”），提取协陪义动词的共现义原，如表1所示：

**表1 协陪义动词共现义原表**

|  |  |
| --- | --- |
| **共现义原** | **协陪义动词** |
| guide|引导 | 带领、带队、带、领导、领、率、率领、引、引导、携、牵 |
| follow|跟随 | 伴、陪同、随同、陪伴、陪、送 |
| help|帮助 | 帮、帮助、臂助、辅助、襄助、协助、拉 |
| HoldWithHand|搀扶 | 搀、搀扶 |
| HoldWithHand|搀扶 help|帮助 | 扶 |
| cooperate|合作 | 协同 |
| protect|保护 | 护送 |

以共现义原“guide|引导”、“follow|跟随”、“help|帮助”、“HoldWithHand|搀扶”、“cooperate|合作”、“protect|保护”为基础，根据“词汇语义相似度计算规则”查找相似动词。

## 参考文献

[1] 刘群, 李素建. 基于《 知网》 的词汇语义相似度计算[J]. 中文计算语言学, 2002.