

# 叙词表融合方法研究\*

吴雯娜 王 星

**摘 要** 多表融合过程中可能产生概念界定不一致,语义关系矛盾冲突、相互缠结的问题,加上多人协同工作模式的复杂性,致使知识体系梳理极为困难。从逻辑角度对“概念”、“词汇”及概念间“关系”进行界定,提出多表融合的技术路线。多表融合以多表概念映射为前提。由于词群构成的复杂性,需要对 closeMatch 映射进行语义分析,以区分出 exactMatch 映射。概念合并和新关系生成可能引发多种逻辑问题;关系冲突、等级关系循环、等级关系冗余、兄弟概念互参、上下位概念共参、交叉互参、等级关系多路径。这些逻辑问题可被区分为两大类:逻辑错误、逻辑优化,并进一步影响逻辑问题的解决步骤。讨论了多人协同工作模式下叙词表的数据锁定机制。图 9。表 1。参考文献 11。

**关键词** 叙词表融合 语义关系 概念映射

**分类号** G254

## A Method for Thesauri Merging

Wu Wenna & Wang Xing

**ABSTRACT** In thesauri merging, the inconsistency of definitions of concepts and the conflicts or entanglement of semantic relations would arise. It is difficult to reorganize the concepts and relations especially under the collaborative working mode. So, it is very important to develop a technology platform to support the process of thesauri merging. This paper has defined “concept”, “term”, and “relationship” of a thesaurus from a logical point of view. Concepts mapping between different thesauri is pre-stage of multi-thesauri merging. Because of the complex of “Terms-Group” members, it is necessary to analysis the semantics of “closeMatch” concepts, in order to distinguish “exactMatch” concepts. Concepts merging and new relationships establishment may lead to occurrence of logical problems: relationship conflict, hierarchical relations cycle, hierarchical relationships redundant, sibling concept associated mutually, broader concepts and narrower concepts associated mutually, cross-associated mutually, multi-path hierarchical relationships. The logical problems are divided into two categories: logic errors, logic optimization. It would affect the choice of solving steps of the logic problems. Locking mechanism of thesaurus data under multiplayer mode also be discussed. 9 figs. 1 tab. 11 refs.

**KEY WORDS** Thesauri merging. Semantic relationship. Concept mapping.

叙词表主要用于信息的标引与检索。叙词表的语义关系比较简单,这使其技术推广和大规模编制成为可能,全世界编制的叙词表达上千部。随着叙词表编制与应用技术日臻成熟,产生了一系列国际标准和国家标准<sup>[1]</sup>。

《汉语主题词表》<sup>[2]</sup>(以下简称《汉表》)是一部大型综合叙词表。继《汉表》之后,各专业领域采用与《汉表》类似的体制编制专业叙词表,我国叙词表数量已逾百部<sup>[3]</sup>。国内研制的专业表分属不同机构,体制上虽然和《汉表》基

\* 本文系国家社会科学基金“网络环境下叙词表的编制模式与应用方式研究”(批准号:10BTQ048)的研究成果之一。

通讯作者:吴雯娜,Email:wuwenna@istic.ac.cn

本一致,但知识结构的一致性未能得到控制。各专业表和《汉表》之间没有形成母表和子表的集成结构和应用模式,限制了《汉表》在专业领域的应用。此外,传统条件下编制的《汉表》在词汇遴选上主要依赖专业人员个人经验,在反映文献用词和用户用词方面欠缺客观性,入口词比例偏低<sup>[4]</sup>。这些缺陷会影响用户体验,也会影响计算机自动应用的效果。

2010年,中国科学技术信息研究所启动了《汉表》修订工作。新《汉表》一方面要吸纳各专业表中的专业概念,借鉴已有的参照关系;一方面要利用文献数据库中的关键词以及检索系统中的用户检索词增补大量词汇。

大规模的叙词表协同加工即便在当前技术环境下依然是一个很大的挑战。《汉表》编修采用了先分散后集中的方法:先分领域编制专业表,然后再逐步集成、融合为综合表。虽然叙词表编制技术相对成熟,有可参照的技术标准,但多表融合仍是一个有待研究和实践的领域。各专业表覆盖的学科领域既有不同又有交叉,概念体系有详有略,揭示概念关联的角度、表达概念的用词等方面各有不同,当这些概念体系集成到一起后,可能产生概念界定不一致,概念间关系矛盾冲突、相互纠结的问题。各种逻辑问题互相牵扯,互为因果,加上多人协同工作模式的复杂性,致使知识体系的梳理极为困难。因此,技术平台对多表融合流程的支持和控制显得尤为重要。

本文将从逻辑角度对叙词表知识体系的构成和特征进行分析,提出多表融合的技术路线。对多表融合中可能产生的逻辑问题进行分析,讨论多人协同工作模式下逻辑问题控制的步骤和策略。这些讨论有助于在系统开发中认识和把握问题的复杂性,以获得更好的系统性能,改进系统对多表融合过程的支持力度。

## 1 叙词表基本构成要素及其特征

从微观结构上看,叙词表知识体系的构成要素包括“概念”和“关系”。“概念”用自然语言词汇来表示,概念间的关联被归纳为三种类

型,即等同关系、等级关系和相关关系。

### 1.1 词与概念

与分类法不同,叙词表用自然语言词汇来表示概念。自然语言词汇可能存在词义模糊甚至歧义的现象。如果不对词汇歧义问题进行控制,在检索时可能会造成“误检”。另外,自然语言词汇普遍存在同义词现象,一个概念有若干种表达方式。检索时如果没有对同义词进行控制,可能会造成“漏检”。对词汇进行控制是叙词表的主要功能之一。

在叙词表中一条词汇只表达一个概念,即“一词一义”。如果必须使用词形相同的词汇,则人为添加限定成分对词汇进行区分,以表示不同的概念<sup>[5]</sup>。这一点体现出了叙词表作为一种“人工语言”的特点。

对于同义词现象,即一个概念有多种表达形式时,叙词表从中选择一个词作为概念优选词(也称为叙词或正式主题词),其他词汇作为概念非优选词(也称为非叙词、非正式主题词、入口词),在概念优选词和非优选词间建立等同关系(也称用代关系)。叙词表等同关系是词汇与词汇间的关系,不是概念与概念间的关系。叙词表中也常将反义词、某些过于专指的下位词或某些相关词作为概念非优选词。因此,叙词表的“等同关系”不一定就是同义关系,“等同”词汇不一定就是同义词。

### 1.2 概念间的语义关系

叙词表概念之间的语义关系包括等级关系和相关关系。

#### 1.2.1 等级关系

等级关系也称属分关系。叙词表编制标准列举了三种类型的等级关系:属种关系、整体—部分关系、类—实例关系。

从逻辑上看,等级关系有两个特征:①等级关系是一种不对称关系,有方向性。②等级关系是一种可传递的关系,即C1、C2、C3为叙词表概念,如果C1是C2的上位概念,C2是C3的上位概念,则C1是C3的上位概念。

如果把概念看作节点,等级关系看作有向

边,叙词表等级结构就是由等级关系串接多个上下位概念形成的非循环有向图。从顶层概念到叶节点概念所形成的完整的等级结构称为词族。本研究中规定有向边由上位概念指向下位概念,即箭头指向下位概念。

等级关系在语义上须满足以下限定条件:

① 等级关系连接的上下位概念具有相同的语义类型,即上下位概念为同一类事物,或都为某种物(thing),或都为某种行为(action),或都为某种属性(property);② 满足“全部是一部分是”(All-and-Some)限定,即下位概念必须“全部”属于上位概念,上位概念只有一“部分”是下位概念。

叙词表等级结构可起到类聚概念,引导用户通过等级关系发现未知概念的作用。但用于引导用户的等级结构和叙词表标准规定的等级结构并不总是一致的。在叙词表中事实上存在着一些不符合标准规范的等级结构,比如在“事物”与其“方面”间建立等级关系。另外,由于事物具有多重属性,比如“人”有性别、年龄、职业等属性,用于表现事物的“概念”就具有了从多个角度进行细分的可能性,即概念具有多向成族的特征。专业不同,应用目标不同,都会影响概念细分的标准。概念多向成族的特性会在很大程度上增加概念体系融合的复杂度。

### 1.2.2 相关关系

相关关系在叙词表中没有明确界定。如果两个概念之间有必要建立关联,而这种关联关系又不属于等级关系,则用相关关系来表现<sup>[6]</sup>。由于未对概念相关的原因和类型进行明确形式化的说明,相关关系在逻辑上具有如下特征:① 相关关系是一种对称关系,无方向性;② 相关关系不宜作为一种可传递的关系类型。

### 1.3 数据模型

传统叙词表以词汇为基本描述对象,通过描述词汇属性来表现叙词表知识体系。词汇属性包括范畴、用、代、属、分、参关系,英文译称,定义等。非优选词与优选词之间为用代关系,在优选词间构建属、分、参关系。ISO 25964 和 SKOS(Simple Knowledge Organization System)采

用以概念为核心的描述机制,即把词汇作为概念的“标签”,一个概念可以有多个“标签”。在概念之间构建等级关系或相关关系。基于概念的描述机制在逻辑上更加清晰,能更好地支持对叙词表知识体系的复杂操作。

本研究采用了基于概念的描述方式,对词汇和概念分层进行描述——在词汇层面描述词汇译称、定义等信息,在概念层面描述范畴、等级关系和相关关系。另外,本研究提出了“词群”概念,“词群”即一个概念的多个表达词汇所形成的词汇集。概念与词群一一对应,一个词群包含一条或多条词汇,每条词汇唯一属于一个词群。词群中有且只有一条词汇是优选词,其余为非优选词。

为控制多表融合进程,避免逻辑问题因堆积而难于梳理,本研究默认融合之初各来源表中的每条词、每个概念都是唯一的,不管它们在今后的融合中是否被确定为等同。

## 2 叙词表融合的技术路线

### 2.1 映射与融合

语义层面的多表集成模式包括映射与融合。这两种模式既有技术上的共同之处,又有目标和过程上的不同。从目标上看,“映射”是要揭示各表概念的对应关系,不生成新表;“融合”是将多表合并,最终生成一部新表。从过程上看,“映射”要尽量在目标表中找到与来源表概念对应的概念,对应关系可能是等同、等级、相关或其他;“融合”不但要找到概念之间的映射关系,还要进一步合并等同概念,或在非等同映射的概念间建立相应的关系。

虽然目标不一致,但两种集成模式在技术上有共通之处:发现概念之间的映射关系。本文将此过程称为“关联点识别”。“映射”可视为“融合”的一个技术环节。

映射阶段要解决的问题是在不同概念体系中找到最相关的概念,即找到不同概念体系的“关联点”。SKOS 提出了概念映射的五种类型:closeMatch,exactMatch,broadMatch,narrowMatch,relatedMatch<sup>[7]</sup>。broadMatch、narrowMatch 表示相

互映射的概念为等级关系;relatedMatch 表示相互映射的概念为相关关系;exactMatch 表示相互映射的概念为等同关系;closeMatch 表示相互映射的概念非常接近。可把 exactMatch 视为 closeMatch 的一种特殊情况。

在融合阶段,需要把等同映射(exactMatch)的概念合并,对于其他映射类型的概念则建立相应的关联关系。融合阶段概念的合并和关联会改变概念体系的原有结构,可能导致语义关系出现逻辑问题,因此需要对语义关系的逻辑问题进行控制。

朱礼军提出了叙词表语义集成包括“同义词和多义词分析”、“概念映射建立”、“概念合并”、“相关关系发现”、“融合后词汇表的一致性处理”五个步骤<sup>[8]</sup>,薛春香对以上步骤所应包含的具体项目进行了补充<sup>[9]</sup>。本研究将叙词表融合的步骤重新进行梳理,把“同义词和多义词分析”、“概念映射建立”合并为“关联点识别”(包含“相关关系发现”)。多表融合的步骤为:关联点识别——概念合并或概念关联——语义关系逻辑问题控制——新的关联点识别。

## 2.2 关联点识别与处理

关联点识别就是找到各来源表中互为等同、等级或相关关系的概念。但这三类映射的优先级是不一样的,等同映射优先级最高,因为等同概念识别是概念语义和词群边界的界定问题,是识别与处理其他映射关系的基础。融合完成后新的概念体系中不允许等同的概念分列多处。

各来源表由于存在学科交叉,因此存在共有概念,这些概念在不同词表中可能表现为相同或不同的词汇,这是进行词表融合的基础。等同概念识别有多种方法,可通过各来源表中的交叉词群(即有着相同词汇的词群)来找到可能的等同概念,也可借助同义词识别工具,如同义词词典或同义词识别算法。关联点识别过程中,由于词汇歧义,以及词群中非同义词的影响,等同映射获得的映射关系只能作为候选等同关系,可视为 closeMatch 映射来处理。由于词群构成的复杂性,两个概念体系间甚至可能形

成“一对多”closeMatch 映射。

为明确映射类型,需要对 closeMatch 映射涉及的概念及其词群进行语义分析。对于词群有交叉、但词群对应概念不等同的,需要对词群进行拆分重组,包括对歧义词词形进行区分,保证概念和词群间一一对应。需要注意的是,词群“清理”过程不能改变概念的语义(本研究中规定概念的语义主要由优选词确定)。词群确定后可对映射类型进行判断,确定是否为 exactMatch 映射。如果不是 exactMatch 映射,可进一步判断是否为 broadMatch 或 narrowMatch 映射。如果以上映射均不成立,可确立为 relatedMatch 映射或解除映射关系。

exactMatch 映射的判断依据为:将 closeMatch 映射的两个概念互换后,原有概念体系的语义关系仍然成立,则该映射为 exactMatch 映射。

两个概念体系的 exactMatch 映射为一对一映射。和 exactMatch 映射不同,broadMatch、narrowMatch、relatedMatch 映射都有可能出现一对多的情况,这在逻辑上是允许的。Martin Doerr 提出了 broadMatch、narrowMatch“最优映射”的约束条件<sup>[10]</sup>:①叙词表 A 中的概念 C 在叙词表 B 中没有对应的等同概念,但概念 C 在叙词表 B 中有至少一个对应的上位概念和至少一个对应的下位概念;②概念 C 对应的上位概念是最小的,即在叙词表 B 中没有一个概念比该上位概念小,同时又比概念 C 大;③概念 C 对应的下位概念是最大的,即在叙词表 B 中没有一个概念比该下位概念大,同时又比概念 C 小。

relatedMatch 也存在“最优映射”的判断问题。由于相关关系无方向性,“最优”relatedMatch 映射很难从逻辑角度进行判断,需要对概念进行语义分析。

exactMatch 映射的等同概念可合并。等同概念合并包括词群合并和属性合并。词群合并就是将等同概念对应词群的词汇查重合并,并选择新词群的优选词。优选词选择的原则之一是不能改变概念语义。属性合并又分为词汇层面的属性合并和概念层面的属性合并。词汇层面的属性合并是指将等同概念对应词群中存在

的同形同义词的译称、注释等属性信息进行查重合并。概念层面的属性合并是指将等同概念的语义关系和范畴分别进行查重合并。范畴合并后需要进行逻辑检查,如有多个范畴,要确保范畴间不构成上下级关系。

broadMatch、narrowMatch 映射的概念间可建立等级关系;relatedMatch 映射的概念间可建立相关关系。

### 3 叙词表语义关系的逻辑问题

概念合并和关联都会加大概念体系的复杂度。由于叙词表融合有着多来源表、多专业、分期分批、多人协同工作的特征,随着概念体系复杂度增加,由于不恰当的概念合并或关联,或者由于来源表本身存在语义模糊甚至错误,在融合过程中产生各种逻辑问题在所难免。

Marios Sintichakis 提出了叙词表融合中必须避免的两种逻辑错误:等级关系循环和等级关系冗余<sup>[11]</sup>。本研究对叙词表融合中的逻辑问题进行了完善,将叙词表语义关系的逻辑问题分为错误和优化两个级别。逻辑错误包括关系冲突、等级关系循环、等级关系冗余三种情况。逻辑优化包括兄弟概念互参、上下位概念共参、交叉互参、等级关系多路径四种情况。

#### 3.1 逻辑错误

##### 3.1.1 关系冲突

如果两个概念间既有等级关系又有相关关系,即上下位概念之间存在相关关系,则为关系冲突,如图 1 所示。图 1 及以下各图中的圆点表示概念节点,有向边表示等级关系,无向边表示相关关系。图 1 中(a)图的 A 与 B 之间既有等级关系,又有相关关系;(b)图中 A 是 B 的上位概念,B 是 C 的上位概念,则 A 也是 C 的上位概念,在 A 和 C 之间同时存在等级关系和相关关系。这两种情况都属于关系冲突错误。

##### 3.1.2 等级关系循环

互为等级关系的概念相互串接后形成了首尾相接的环则为等级关系循环错误,如图 2 所示。最小的循环只涉及两个概念,如图 2 中的

(a)图。

##### 3.1.3 等级关系冗余

当两个概念间的等级关系可通过其他等级关系推导得出时,这两个概念间的等级关系即为冗余,如图 3 所示。图 3 中,C 是 B 的下位概念,B 是 A 的下位概念,可推出 C 是 A 的下位概念,因此,从 A 直接指向 C 的等级关系是冗余的。

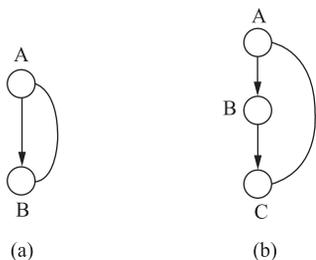


图 1 关系冲突

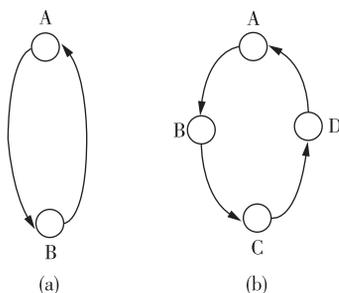


图 2 等级关系循环错误

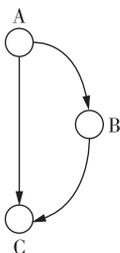


图 3 等级关系冗余错误

#### 3.2 逻辑优化

有一些语义关系虽然在逻辑上是合理的,但也可能存在优化的需要。

##### 3.2.1 兄弟概念互参

同一个概念的多个直接下位概念互称兄弟概念,如图 4 所示。图 4 中 B、C、D 同为 A 的直

接下位概念, B、C、D 互称“兄弟概念”。如果在 B、C 之间又有相关关系, 则为“兄弟概念互参”。

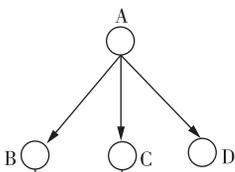


图 4 兄弟概念互参

### 3.2.2 上下位概念共参

互为等级关系的多个概念如果具有共同的相关概念即为“上下位概念共参”, 如图 5 所示。图 5 中 B 和 D 为上下位概念, 同时又都与 A 有相关关系。

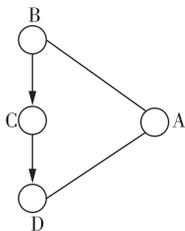


图 5 上下位概念共参

### 3.2.3 交叉互参

四个概念间形成如图 6 所示的关系: A 是 C 的上位概念, E 是 H 的上位概念, A 与 H 相关, 同时 C 与 E 相关, 相关关系出现了“交叉”, 这种情况即为“交叉互参”。

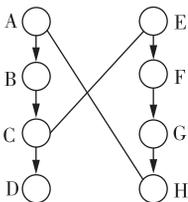


图 6 交叉互参

### 3.2.4 等级关系多路径

等级关系多路径即两个概念节点之间有多条等级关系链, 如图 7 所示。

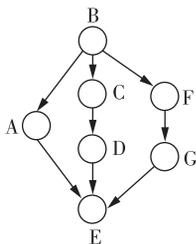
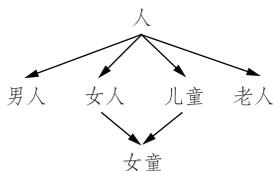


图 7 等级关系多路径

等级关系多路径有其合理性因素, 因为概念存在多个分面, 有多种属性, 存在多向成族的可能性。如下例:



“人”下分的依据包括两方面, 一个是性别, 将“人”分为“男人”、“女人”; 一个是年龄, 将“人”分为“儿童”、“老人”。“女童”这个概念中则同时反映了性别和年龄这两种属性, 因此出现了多属(多个直接上位词)。在“人”与“女童”两个概念间有两个等级链。

等级关系多路径也有可能是各来源表概念颗粒度不同, 概念体系融合不到位造成的, 如图 8 所示。图 8 中, 如果 A 是 C 的下位概念, 同时又是 D 的上位概念, 则等级关系可进一步优化, 从而解决多路径问题。

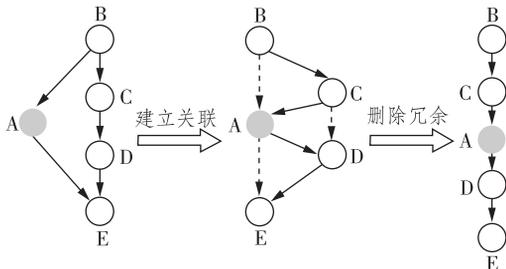


图 8 可优化的等级关系多路径问题

## 4 多人协同工作模式下数据锁定与逻辑问题控制

### 4.1 数据锁定机制

多表融合中,有的操作对象是词汇,有的是概念,有的是关系;有的操作在一个词族中进行,有的是跨词族进行。在多人协同工作模式下,多表融合系统要依据操作类型确定数据对象的最佳锁定范围,尽量避免对同一数据对象的并发操作,并保证有较高的工作效率。

本研究采用了基于概念的描述方式,因此

表1 叙词表融合中各类操作的锁定范围

操作对象	操作类型	锁定范围
词汇	词形更改 英文、注释的增、删、改	当前概念
概念	概念节点的增、删 概念范畴的增、删 概念词群元素的增、删 概念优选词确定与更换	当前概念
概念、关系	等级关系、相关关系的增、删、改	当前概念及与其直接关联的概念
概念、关系	关联概念的识别与处理	候选关联概念
概念、关系	逻辑问题的解决	逻辑问题涉及的概念

### 4.2 逻辑问题控制策略

叙词表语义关系必须符合特定的逻辑限定。逻辑问题成因比较复杂,这些问题可能环环相扣,互为因果,需要有一定的处理流程和优先级顺序,以保证融合过程有条不紊。另外,逻辑问题涉及范围各有不同,有的发生在一个词族中,有的则波及多个词族。如果涉及范围过大,系统计算耗时较长,为排除错误一次性锁定的数据较多,会导致工作效率降低。因此,对逻辑问题的控制要有一定的灵活性,需要根据情况采用不同的策略。

本研究中将逻辑问题的控制方式分为“先控”和“后控”。“先控”就是对欲建立的关系进行逻辑检查,如果该关系的建立不会引发逻辑问题,则允许构建;如引发了逻辑问题,则要求

将概念作为最小锁定单位。概念锁定意味着概念对应词群中所有词汇都被锁定。当操作对象不包括关系时,锁定相对简单,只需锁定当前需要加工的概念即可。概念增删、范畴增删、词群元素增删、概念优选词确定与更换都只需对当前概念进行锁定。如果操作对象包括关系,锁定相对复杂。编辑等级关系和相关关系时需要当前概念及其直接关联的概念进行锁定。识别与处理关联概念时需要锁定候选的多个关联概念;解决逻辑问题时需要锁定问题涉及的多个概念,具体见表1。

解决,否则不予构建关系。“后控”是指在构建关系时不检查某类逻辑问题,留待以后检查和解决。本研究列举的逻辑问题中,“逻辑错误”问题需要进行“先控”,“逻辑优化”问题可进行“后控”。“后控”还可根据逻辑问题涉及范围的不同采用不同的方式:在一个词族中发生的可在词族构建过程中进行检查和解决;涉及多个词族的则对全部数据阶段性地逻辑检查和解决。

对于同一类型的逻辑问题,可优先处理涉及概念节点最少的。比如在图9(a)中有两个等级关系冗余错误,一个包含A、B、C三个节点,另一个包含A、B、D、E四个节点。可优选解决包含节点少的。

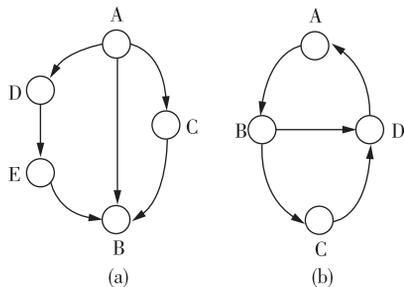


图9 多个逻辑问题并存

对于不同类型的逻辑错误,或不同类型的逻辑优化问题,也可以区分优先级顺序。可先解决相对比较简单,因为简单问题的解决可能化解复杂的问题。如图9(b)中有两种类型的错误,等级关系冗余和等级关系循环。等级关系冗余错误包含概念节点B、C、D。等级关系循环有两个,一个循环包含A、B、C、D四个节点,另一个包含A、B、D三个节点。可以先解决等级关系冗余错误,解决等级关系冗余错误的同时有一个循环错误也就化解了。

逻辑错误处理的优先级顺序建议为:关系冲突 > 等级关系冗余错误 > 等级关系循环错误。由于等级关系确定了概念体系的基本框架,对相关关系逻辑问题的判断需要以等级关系作为前提。因此,逻辑优化问题处理的优先级顺序建议为:等级关系多路径问题 > 兄弟概念互参 > 上下位概念共参 > 交叉互参。

## 5 问题与讨论

锁定方式与加工模式有关。等级关系加工一般有两种常用的工作模式,一种是表单方式,在表单中编辑关系,提交表单时进行逻辑检查。这种方式只需对当前概念及其直接关联的概念进行锁定,锁定范围小。但表单模式难以顾及整个词族语义结构的合理性,整体性较差,易于发生逻辑问题,编者负担较重。另一种方式是在多个概念间进行拖拽等操作,对每一步操作进行逻辑控制,在“可见”的状态下逐步形成树状等级结构。这种方式比较直观,整体性好,但需锁定词族中所有概念及其相关概念。对相关

概念的锁定又会导致包含相关概念的词族被锁定,这样会波及过多的数据,导致工作效率降低。如果不锁定相关概念又可能存在并发操作问题,具体的解决方案还需要进一步研究和实践。

## 参考文献:

- [1] 刘华,曾建勋,沈玉兰. 网络环境下叙词表编制标准的国际发展趋势[J]. 情报杂志,2009,28(11). (Liu Hua, Zeng Jianxun, Shen Yulan. International development trend of standard of thesauri establishment under the networked environment [J]. Journal of Information, 2009, 28(11).)
- [2] 中国科学技术情报研究所. 汉语主题词表[M]. 自然科学增订本. 北京:科学技术文献出版社,1991. (Institute of Scientific and Technical Information of China. Chinese thesaurus [M]. Updated version of the natural sciences. Beijing: Scientific and Technological Literature Press, 1991.)
- [3] 常春,卢文林. 叙词表编制历史、现状与发展[J]. 农业图书情报学刊,2002(5). (Chang Chun, Lu Wenlin. History, current situation and development of thesauri construction [J]. Journal of Library and Information Sciences in Agriculture, 2002(5).)
- [4] 吴雯娜,曾建勋. 叙词表微观结构的描述与评价:EI叙词表与中文叙词表的对比分析[J]. 图书情报工作,2009,53(8). (Wu Wenna, Zeng Jianxun. Description and evaluation of thesaurus microstructure: Contrastive analysis of EI thesaurus and Chinese thesauri [J]. Library and Information Service, 2009, 53(8).)
- [5] 全国文献工作标准化技术委员会. GB 13190-91汉语叙词表编制规则[M]. 北京:中国标准出版社,1992. (China National Technical Committee of Standardization for Documentation. GB 13190-91 Guidelines for establishment and development of Chinese thesauri [M]. Beijing: Standards Press of China, 1992)
- [6] ISO 25964-1:2011 Information and documentation-Thesauri and interoperability with other vocab-

ularies-Part 1: Thesauri for information retrieval [M]. [2011-08-20]. [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=53657](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53657).

- [7] SKOS Simple Knowledge Organization System Reference[M]. [2011-08-20]. <http://www.w3.org/TR/skos-reference/>.
- [8] 朱礼军,赵新力,乔晓东,等. 跨领域多来源主题词表集成与服务研究[J]. 现代图书情报技术, 2007(1). (Zhu Lijun, Zhao Xinli, Qiao Xiaodong, et al. Research of multidisciplinary and multi-sources thesauri integration and service architecture [J]. New Technology of Library and Information Service, 2007(1).)
- [9] 薛春香,乔晓东,朱礼军. 基于集成的领域知识组织系统构建初探[J]. 现代图书情报技术, 2009(11). (Xue Chunxiang, Qiao Xiaodong, Zhu Lijun. Preliminary study on the construction of domain knowledge organization system based on integration [J]. New Technology of Library and Infor-

mation Service, 2009(11).)

- [10] Martin Doerr. Semantic problems of thesaurus mapping[J/OL]. Journal of Digital Information, 2001, 1(8). [2011-08-20]. <http://journals.tdl.org/jodi/article/view/31/32>.
- [11] Marios Sintichakis, Panos Constantopoulos, Fazli Can, et al. A method for monolingual thesauri merging [C]. Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval, 1997: 129 - 138.

**吴雯娜** 中国科学技术信息研究所副研究馆员。通讯地址:北京市复兴路15号323室。邮编:100038。

**王星** 中国科学技术信息研究所工程师。通讯地址同上。

(收稿日期:2011-10-08;修回日期:2011-12-01)