

自动提取信息模型的关键内容¹

张伟良 冯重熙

(清华大学电子工程系 北京 100084)

摘 要 信息模型是 TMN 的关键技术之一, 本文讨论了自动实现信息模型从 GDMO 模板描述到编程语言描述的重要性, 详细分析了 GDMO 模板的结构, 并叙述了自动提取信息模型关键内容的算法, 最后提出了对 GDMO 模板改进的 3 点建议。

关键词 电信管理网, 信息模型, 管理目标定义指南, 抽象语法定义 1

中图分类号 TN919.1

1 引 言

信息模型是电信管理网 (TMN) 中的基本概念之一, 它是电信网中物理或逻辑资源的抽象, 使得网络管理系统对网络资源有一个共同的认识, 基于这种公共的网络管理知识, 不同的网络管理系统可以很容易地实现互连和互操作。ITU 用“管理目标定义指南 (GDMO)”^[1] 描述信息模型。GDMO 采用一系列模板, 运用面向对象的方法来描述管理目标。

信息模型的 GDMO 模板描述并不能直接应用于实际的网络管理开发, 还需要通过一定的手段转换成编程语言描述。GDMO 模板的面向对象特性支持信息模型可以用面向对象的编程语言来描述。从现有 ITU 的标准和建议来看, 信息模型规定得很细, 种类繁多, 如果采用手工转换的方法显然不能胜任网络管理的开发工作, 可以考虑采用自动方式来实现信息模型的 GDMO 模板描述到编程语言描述的转换。自动转换的工作原理是: 以 GDMO 模板为参考基础提取关于信息模型的各种关键内容, 按照一定的转换机制, 把提取的信息模型关键内容用面向对象的编程语言来描述。因此, 自动转换的两个要素是信息模型的提取机制和转换机制。相对于手工转换, 自动转换具有以下优点: (1) 大大缩短转换时间, 提高了转换效率; (2) 可以避免人为错误, 提高了转换准确率; (3) 可以自动识别维护内容, 提高了维护效率; (4) 遵循 ITU 标准, 具有比较强的通用性。

下面将首先分析 GDMO 模板的结构, 然后叙述提取信息模型关键内容的算法, 最后从自动提取信息模型关键内容的算法出发, 对 GDMO 模板提出了 3 点改进建议。

2 GDMO 模板

ITU 为信息模型定义了 9 个模板, 包括管理目标类、特性组、属性组、属性、通知、动作、参数、行为和命名约束等。每个模板都各自具备一些关键内容, 而这些关键内容是用关键词提示的, 关键词及其所提示的关键内容共同组成一个构件 (Construct)。这里所谓关键词是指 GDMO 模板中大写的单词或词组, 如管理目标类中的 MANAGED OBJECT CLASS, DERIVED FROM, REGISTERED AS 等等, 与关键词相关的内容就是关键词提示的关键内容。下面具体分析 GDMO 模板中关键词及其提示的关键内容之间的关系分类 (由于 GDMO 模板已经在文献 [1] 中详细叙述, 在关系分类中就不具体举例):

- (1) 只有关键词;
- (2) 关键内容只有一项, 后面紧跟关键词。这种情况一般代表一个模板的开始;

¹ 1999-07-20 收到, 1999-11-21 定稿
国家自然科学基金资助项目 (69896242 号)

(3) 关键词后面紧跟关键内容, 关键内容只有一项;

(4) 关键词后面紧跟关键内容, 关键内容有若干项, 属于同一类型, 它们之间以逗号隔离, 以分号结束;

(5) 关键词后面紧跟关键内容, 关键内容有若干项, 但不属于同一类型, 主要包含以下 $a-j$ 10 种情况: (a) 特性组的 ATTRIBUTES 构件; (b) 特性组的 ATTRIBUTE GROUPS 构件; (c) 特性组的 ACTIONS 构件; (d) 特性组的 NOTIFICATIONS 构件; (e) 通知的 WITH INFORMATION SYNTAX 构件; (f) 命名约束的 SUBORDINATE OBJECT CLASS 构件; (g) 命名约束的 NAMED BY SUPERIOR OBJECT CLASS 构件; (h) 命名约束的 CREATE 构件; (i) 命名约束的 DELETE 构件; (j) 管理目标类的 CONDITIONAL PACKAGES 构件。

下面以管理目标类模板为例, 说明 GDMO 模板是如何用关键词来提示关键内容的 (如表 1 所示)。

表 1 管理目标类模板的关键词及其提示的关键内容

关键词	关键内容	备注
MANAGED OBJECT CLASS	class-label	管理目标类, 有且仅有一个
DERIVED FROM	class-label	管理目标类, 没有或有一个或几个
CHARACTERIZED BY	package-label	必要特性组, 没有或有一个或几个
CONDITIONAL PACKAGES	package-label	条件特性组, 没有或有一个或几个
PRESENT IF	condition-definition	条件特性组的条件, 与条件特性组一一对应
REGISTER AS	object-identifier	注册名, 有且仅有一个

管理目标类模板如下:

```
<class-label> MANAGED OBJECT CLASS
  [DERIVED FROM <class-label>[, <class-label>]*;]
  [CHARACTERIZED BY <package-label>[, <package-label>]*;]
  [CONDITIONAL PACKAGES <package-label> PRESENT IF condition-definition
    [, <package-label> PRESENT IF condition-definition]*;]
REGISTERED AS object-identifier;
```

在 GDMO 模板中, [] 表示的内容可以存在, 也可以不存在; * 表示的内容可以重复, 逗号用于间隔同一关键词提示的关键内容, 分号用于结束一个关键词提示的关键内容。

通过以上对 GDMO 模板的详细分析, 可以看出 GDMO 模板的描述具有很强的规律性和条理性; 因此通过一定的算法来实现信息模型关键内容的自动提取是可行的。

另外, 在读取 GDMO 模板时, 还应该注意几个问题:

(1) 模板互相嵌套的情况。主要有两种, 特性组模板出现在管理目标类模板中和行为模板出现在特性组模板中。

管理目标类和特性组互相嵌套的情况是, 特性组出现在管理目标类的 CHARACTERIZED BY 和 CONDITIONAL PACKAGES 构件中, 这样使得 (4) 和 (5)(j) 两种情况变得复杂。这时需要考虑: (a) 特性组是独立的, 还是出现在管理目标类的 CHARACTERIZED BY 构件或 CONDITIONAL PACKAGES 构件中; (b) 特性组模板的开始和结束标志, 特别是当特性组没有 REGISTERED AS 构件时。

行为模板嵌套在特性组模板中的情况稍微简单一些, 行为模板只有两个构件 BEHAVIOUR 和 DEFINED AS, 在行为模板中, 它们是连续的, 可以把它们看作一个关键词, 该关键词的前面为行为名称, 后面为行为描述。

(2) 关键内容是字符串的情况。包括 BEHAVIOUR 的 DEFINED AS 构件、属性组的 DESCRIPTION 构件和管理目标类中条件特性组的 PRESENT IF 构件。

前两种构件中的字符串内容一般是以分号结束,第3种构件中的字符串内容以分号或逗号结束。但是在字符串自身的描述过程中也可能会出现分号和逗号,应该从文法上来区分这两种情况,即:如果字符串中的逗号或分号是真正的标点符号,它们与句子之间一般没有间隔,而作为字符串结束标志的逗号或分号,它们的前面一般会出现其他的标点符号,如句号或引号等。

(3) 标点符号的作用。在模板的描述中,逗号一般用来隔离同一关键词提示的若干个等同的关键内容,而分号表示关键词提示的关键内容的结束;

(4) 一个模板开始和结束的标志。一个模板的开始比较容易判断,因为各个模板都有相应的关键词标志模板的开始,如管理目标类模板的 MANAGED OBJECT CLASS 和特性组模板的 PACKAGE 等等。一个模板的结束一般都以构件 REGISTERED AS 结束,但是在有的模板没有 REGISTERED AS 构件时,情况稍微复杂一些,有两种判断条件,一是当一个模板还没有结束而另一个模板开始时,则结束前一个模板;二是当一个模板还没有结束但出现了别的模板关键词时,结束前一个模板。

3 提取信息模型关键内容的算法

自动提取信息模型关键内容的算法思路是:在分析 GDMO 模板结构的基础上,从描述信息模型的文本文件出发,搜索 GDMO 模板的关键词,根据相应构件的结构提取关键词提示的关键内容。在信息模型的文本文件中搜索关键词之前,还需要完成识别字符和提取单词的工作。算法的流程如图 1 所示,主要包括四个步骤:读取并识别字符、读取并识别单词、定位关键词和提取关键内容。

首先讨论字符的读取和识别。在抽象语法标记 ASN.1^[2]中规定了字符集合,包括:

A to Z
a to z
0 to 9
:=, { } ;
() [] - ' "

在该字符集合中,可以出现在单词中的字符有 A to Z; a to z; 0 to 9 和“-”,把它们称为单词字符 (Word Character); 其他字符可以用来标志一个单词的开始或结束,把它们称为标志字符 (Flag Character)。但是这种情况并不是绝对的,虽然“-”可以出现在单词中,但在 ITU 文件中,“---”是用来表示注释的;还有,虽然句号一般是一个句子的结束标志,但在 ITU 文件中,句号可以用来连接两个单词代表一个整体,等等。这两种情况算法都已考虑。

然后讨论单词的读取,如图 2 所示。读取单词的过程为:先读取一个字符,如果它是单词字符,则写入缓存区;如果它是标志字符,取出缓存区中的内容作为一个单词并返回该单词的标志字符(如果缓存区为空,则认为没有读到单词),同时清除缓存区,准备读取另一个单词。一个单词的标志字符很有用,用它可以判断一个关键词所提示的关键内容是否结束。

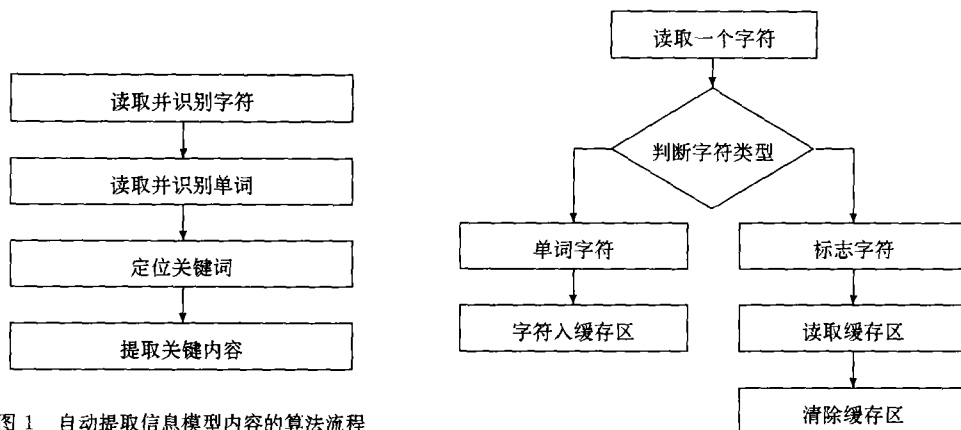


图 1 自动提取信息模型内容的算法流程

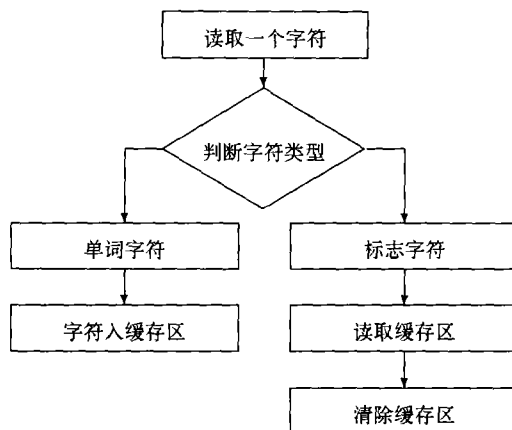


图 2 读取单词的算法

第三步讨论关键词的定位(单词匹配)。每次匹配后的结果称为单词匹配状态,可以分为3种情况,如表2所示。

匹配状态	对应的情况
NOTMATCH	单词不是关键词的子串
PARTMATCH	单词是关键词的真子串
COMPLETEMATCH	单词与关键词完全一样

关键词定位(单词匹配)的过程(如图3所示)为:

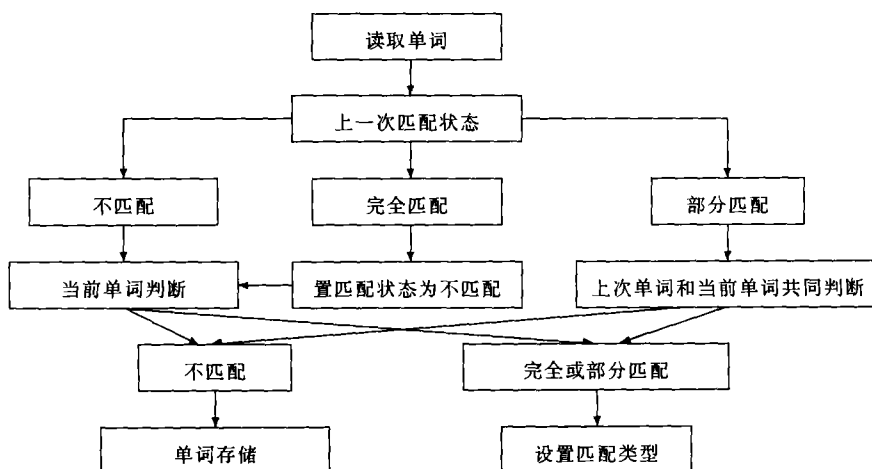


图 3 关键词的定位算法

(1) 读取到一个单词后,首先检测上一次的匹配状态,如果是不匹配,则直接进行当前单词的匹配;如果是完全匹配,则设置匹配状态为不匹配,然后进行当前单词的匹配;如果是部分匹配,情况稍微复杂一些,需要把上一次的单词和当前单词结合进行匹配,例如 MANAGED 是 MANAGED OBJECT CLASS 的一部分,匹配状态为部分匹配,当再读取到 OBJECT 时,

MANAGED OBJECT 仍然是 MANAGED OBJECT CLASS 的一部分, 直到读取到 CLASS 时, 才是完全匹配状态;

(2) 判断该次匹配状态, 如果是不匹配, 存储单词, 这样做是因为有时关键词提示的关键内容在关键词的前面, 例如当匹配到 MANAGED OBJECT CLASS 时, 需要提取它前面的 class-label 作为管理目标类的名称; 如果是部分匹配, 存储单词, 以备和下个单词结合进行匹配; 如果是完全匹配, 则表示已经匹配到一个关键词, 进入相应的关键内容提取;

(3) 重复步骤 (1) 和 (2)。

最后讨论关键内容的提取。由于各个关键词提示的关键内容结构并不完全一样, 具体的关键词需要分析具体的关键内容结构 (参考本文提供的 GDMO 模板结构分析), 因此没有一个通用的算法, 但有一点是共同的, 就是利用单词的标志字符来追踪关键内容的结构。下面还是以管理目标类模板为例解释提取关键内容的过程:

(1) 匹配到 MANAGED OBJECT CLASS 后, 直接提取它前面的单词作为管理目标类名称, 同时也表示一个管理目标类模板的开始;

(2) 匹配到 DERIVED FROM 后, 反复读取单词作为管理目标类名称, 直到单词的标志字符是分号为止;

(3) 匹配到 CHARACTERIZED BY 后, 反复读取单词作为必要特性组名称, 直到单词的标志字符为分号为止;

(4) 匹配到 CONDITIONAL PACKAGES 后, 简单地设置标志;

(5) 匹配到 PRESENT IF 后, 读取它前面的单词作为条件特性组名称, 读取后面的字符串作为条件特性组的条件, 直到读取到分号, 结束 CONDITIONAL PACKAGES 标志;

(6) 匹配到 REGISTERED AS 后, 以分号为结束标志读取一个字符串, 然后分析该字符串得到管理目标类的注册名, 同时也表示该管理目标类模板的结束。

当然, 在规律性和条理性背后, GDMO 模板仍然存在一些需要改进的地方, 如模板之间的嵌套、字符串的描述以及标点符号的使用等问题, 这些问题不会影响算法的实现 (见下一节), 但是会增加算法的复杂度。

4 对 GDMO 模板的要求

在 ITU 信息模型文件中出现一些特殊情况, 使得在自动提取关键内容的时候遇到一些困难, 主要有 3 种这样的情况:

(1) 标点符号和句子之间可能存在间隔, 这样的间隔可能是空格、跳格或回车。这种情况在该算法中可能导致无法正确判断一个关键词所提示关键内容的结束, 例如:

DERIVED FROM top;

DERIVED FROM top ;

前一种情况读取到 top 后它的标志字符是分号, 表示 DERIVED FROM 关键词提示的关键内容已经结束; 而后一种情况读取到 top 后它的标志字符为空格, 算法将继续读取后面的单词作为 DERIVED FROM 的关键内容, 而实际上该 DERIVED FROM 的关键内容只有 top, 从而导致错误。

解决该问题有两种方法, 一是直接以分号为结束标志读取一个字符串, 然后在该字符串中提取关键内容; 二是首先消除 ITU 信息模型文件中句子和标点符号之间的间隔, 然后再利用该算法进行关键内容的提取, 间隔的消除可以通过另外一个算法来实现。本文基本上采用后一种方法。

本文认为, 作为大量信息模型基础的 GDMO 模板, 应该具有严格性, 在句子及其结束标点符号之间不存在间隔是符合语法的。虽然在句子和标点符号之间插入间隔可以达到美观的效果, 但是这样做并不严格, 因为可以插入任何间隔字符, 也可以插入任意个间隔字符。

(2) 信息模型的描述中存在各种模板内容互相嵌套的情况, 使得自动提取算法不能简单地按照 GDMO 模板来设计。例如下面将要提到的管理目标类模板, 匹配到关键词 CHARACTERIZED BY 后, 正常的情况是反复读取其后的单词作为必要特性组名称, 直到单词的标志字符是分号为止, 而在该模板中, 读取到 attributeValueChangeRecordPackage 后, 将出现关键词 PACKAGE, 于是算法被迫转入读取 PACKAGE 模板的内容, 扰乱了正常的算法思路。

解决该问题有两种方法, 一是考虑已经出现的模板互相嵌套情况来实现算法, 二是修改 ITU 信息模型文件, 消除模板互相嵌套的情况, 然后按照 GDMO 模板设计算法思路。本文采取的是前一种方法, 但是这种方法很不安全, 因为该方法并没有考虑所有可能出现的模板嵌套情况, 而考虑各种模板嵌套情况将使得自动提取算法变得很复杂, 采用后一种方法既简单又安全。

本文认为, 信息模型的描述应该具有条理性, 应该严格按照 GDMO 模板来描述。在信息模型描述中采用模板嵌套的方式, 虽然显得紧凑, 但是有些混乱, 即使人工阅读也有些困难, 如果把互相嵌套的模板分离开来, 同时又把它们放在一起, 既不失紧凑性又具备条理性。

分离互相嵌套模板的例子如下所示:

```
attributeValueChangeRecord    MANAGED OBJECT CLASS
DERIVED FROM    eventLogRecord;
CHARACTERIZED BY
attributeValueChangeRecordPackage    PACKAGE
        BEHAVIOUR    attributeValueChangeRecordBehaviour
        BEHAVIOUR
```

DEFINED AS " This managed object is used to represent logged information that resulted from attribute value change notifications or event reports "; ;

ATTRIBUTES attributeValueChangeDefinition GET ; ;

CONDITIONAL PACKAGES

sourceIndicatorPackage PRESENT IF " the sourceIndicator parameter is present in the attributeValueChange notification or event report corresponding to the instance of attribute value change record ",

attributeIdentifierListPackage PRESENT IF " the attributeIdentifierList parameter is present in the attributeValueChange notification or event report corresponding to the instance of attribute value change record ";

REGISTERED AS {smi2MObjectClass 2};

经过处理得到以下互相分离的模板描述:

```
attributeValueChangeRecord    MANAGED OBJECT CLASS
DERIVED FROM    eventLogRecord;
CHARACTERIZED BY
        attributeValueChangeRecordPackage;
```

CONDITIONAL PACKAGES

sourceIndicatorPackage PRESENT IF " the sourceIndicator parameter is present in the attributeValueChange notification or event report corresponding to the instance of attribute value change record ",

attributeIdentifierListPackage PRESENT IF " the attributeIdentifierList parameter is present in the attributeValueChange notification or event report corresponding to the instance of attribute value change record ";

REGISTERED AS {smi2MObjectClass 2};

```
attributeValueChangeRecordPackage    PACKAGE
        BEHAVIOUR    attributeValueChangeRecordBehaviour;
        ATTRIBUTES attributeValueChangeDefinition GET;
attributeValueChangeRecordBehaviour    BEHAVIOUR
```

DEFINED AS “ This managed object is used to represent logged information that resulted from attribute value change notifications or event reports ”;

(3) 信息模型的 GDMO 模板中存在一些字符串的描述, 如行为的 DEFINED AS 构件、属性组的 DESCRIPTION 构件和管理目标类的 PRESENT IF 构件等等。字符串的开始很容易判断, 可以有上面几个关键词提示, 但是字符串的结束不是很容易判断, 在本文的 GDMO 模板分析中提出了相应的解决方法。但有一点应该注意, 即在读取一个字符串遇到一个逗号或分号时, 并不能断定一个字符串已经结束。

本文认为, 对于字符串的表达, 应该提供两个意义明确而又唯一的字符, 来提示一个字符串的开始和结束, 可以用特殊字符, 如 @, # 等等。

5 结 束 语

本文算法以 ITU 信息模型的文本文件作为输入, 输出存储信息模型关键内容的文件。该算法已经对 3 个具有代表性的信息模型文件 (X.721, M.3100 和 G.774) 很好地提取出信息模型的关键内容。由于试验结果篇幅过大, 这里无法给出。这三个信息模型文件中存在一些不规范的地方, 算法为了解决这些不规范, 做了一些额外的工作。如果在其他信息模型文件中出现算法未考虑的不规范时, 可能会出现不正常现象。本文将进一步对其他信息模型文件进行试验, 使得自动提取算法更加完善。

信息模型的 GDMO 模板描述到编程语言的自动转换提高了转换的效率、准确性以及可维护性, 这些优点可以从信息模型关键内容的自动提取中体现出来。本文的自动提取关键内容算法为进一步实现信息模型的编程语言描述提供了必要的前提, 本文将进一步研究并实现信息模型的自动转换机制。

参 考 文 献

- [1] X.722, Guidelines for definition of managed objects(GDMO), 1992.
- [2] X.208, Abstract syntax notation one(ASN.1), 1988.

AUTO-EXTRACTING KEY CONTENTS OF INFORMATION MODELS

Zhang Weiliang Feng Zhongxi

(Dept. of Electron. Eng., Tsinghua University, Beijing 100084)

Abstract Information model is one of the key technologies of TMN(Telecommunication management network). This paper discusses the advantages of auto-translating information model from GDMO description to programming language description, analyzes the construction of GDMO templates, brings out an algorithm for auto-extracting key contents of information models, and puts forward three proposals for improving GDMO templates.

Key words TMN, Information model, GDMO, ASN.1

张伟良: 男, 1973 年生, 博士生, 研究方向为电信管理网络的实现。

冯重熙: 男, 1930 年生, 教授, 博士生导师, 研究方向为传输技术、接入网及网络管理。