

# 基于有限自动机的航班计划句法检查技术研究

阚海建 夏洪山

(南京航空航天大学 南京 210016)

**摘 要** 为了提高航班计划正确性和完整性的自动检查和分析能力,在分析航班计划编排特点的基础上,给出了基于有限自动机的航班计划句法检查技术,进行航班计划句法检查中的词法分析、语法分析和规则检查,并进行了例证分析,验证了运用不确定有限自动机进行航班计划句法检查的可行性。

**关键词** 有限自动机;词法分析;语法分析;航班计划;规则检查

**中图分类号**:V8 **文献标志码**:A

航班计划编排是航空公司生产计划中的一项控制性工作,其质量不仅决定着运输生产能否安全、正点的运行,而且还关系到航空公司的运作成本。航班的编排是一系统工程,必须考虑各方面的问题和因素。在实际工作中由于航班计划编排错误造成的延误情况时有发生,同时因航班编排的不合理而使得一线生产部门无法提供及时有效保障的现象同样存在。针对这种情况,笔者提出了基于有限自动机的航班计划句法检查技术,这一技术的应用有助于提高航班计划编排过程的系统性、全面性和高效性。

## 1 航班计划编排过程

航空公司在航班生产的计划阶段,无论是旅客航班运输还是货邮航班运输,从航班计划编排到具体的飞机及机组人员安排,都需要将飞行人员和飞机落实到具体的航班飞行计划中。这就是航班计划编排的过程,包括航班计划编排、勤务编排、飞机排班、机组配对、机组排班等流程。每一过程的编排结果都必须满足航班编排规则要求。

## 2 航班计划句法检查

在民航生产经营的过程中,首先要保证航班计划表示形式的正确性,必须时刻对航班计划进行句法正确性检查,其过程与编译过程中对源程序的词法、语法规则检查过程极其相似。因此可以把编译技术中的有限自动机技术运用到航班计划的句法检查中。将航班计划看作是按民航特定规则编写出来的“语言”,其中航班号、起止时

间、起止地点、机型、机组等基本信息看作是“语言”中的关键字,通过构造相应的不确定有限自动机检查航班号、机场三字码等关键字是否符合其组成规则;采用语法分析对编排的计划进行语义上的分析,检查其是否按照规定的词序及格式编排,是否表达了正确的涵义。从而实现航班计划编排规则的句法检查。

下面以具体的航班计划为例,论证编译技术应用于航班计划句法检查中的可行性。

航班计划的基本内容包含航线、班期和班次、机型、机号等,各个航空公司的航班计划格式有所不同,但基本内容一致,为了便于讨论,本文假设航班计划格式统一为表1所列。一条航班计划就是规则检查的一个基本单元。

表1 航班计划格式

机号	航班号	起始站	到达站	时间		机型	舱位等级	班期
				离站	到站			
B1001	CA3501	PEK	SHA	08:30	10:20	320	FJY	1234567

### 2.1 字符串预处理

在计算机实现的过程中,将表1中所列的一条航班计划定义为字符串:“B1001 CA3501 PEK SHA 0830 1020 320 FJY 1234567”,以空格符为界,换行符\n结束,视为包含9个单词的语句。班期中不应该包含非数字型字符,数字不超过8。无航班的日用点表示,如“1.3..67”表示的班期为星期一、三、六、日。

在进行规则检查之前,程序以上述字符串语句为一个基本单位逐条分离获取出来,作为下一步的处理对象。

### 2.2 航班号的正规表达式与有限自动机

通过对字符串预处理,程序将航班计划逐条

收稿日期:2008-09-09



分离出来,下一步对每条语句中的关键词做规则检查,判断是否符合规则。首先介绍一下正规表达式。每个正规表达式  $r$  表示一个语言  $L(r)$ 。使用这种表示法,程序语言的标识符可以定义为:  $\text{letter}(\text{letter}|\text{digit})^*[1]$ 。

以航班号为例,首先需要规定航班号的组成规则。航班号的编排是由航空公司的2位代码加3位或4位数字组成,航空公司2位代码是报请国际航联确认分配的全世界范围内不得重复的惟一代码,由2位英文字母或1位英文字母加1位数字组成。对于国内航班的航班号,后面数字的第1位代表航空公司的基地所在地区,其中数字1代表华北、2为西北、3为华南、4为西南、5为华东、6为东北、8为厦门、9为新疆,第3、第4位表示航班的序号,单数表示由基地出发向外飞的航班,双数表示飞回基地的回程航班。本文只针对国内的航班进行分析,选用目前常用的国内航空公司2字码为例:国际航空CA、南方航空CZ、东方航空MU、上海航空FM、厦门航空MF、海南航空HU、深圳航空ZH、四川航空3U、山东航空SC。

因此航班号的正规表达式可以表示为  $(C|M|F|H|S|Z)3(A|Z|F|U|M|C|H)(0|1|2|3|4|5|6|7|8|9)^*$ ,所有航班号的集合是其中的一个子集。

通过构造有限自动机可以把正规表达式编译成识别器。有限自动机分为确定和不确定2种,

都能识别正规表达式所表示的语言,一般来说确定有限自动机导出的识别器速度上快很多,但比等价的不确定有限自动机要大很多。由于转换为不确定有限自动机更直接,因此首先构造相应的不确定有限自动机。

由航班号的正规表达式可以得出如图1所示NFA状态转换图,之所以在这里引入具有动作的NFA,主要目的是为了把识别各类单词的有限自动机用矢线连接起来,组成一个单一的NFA,然后把所得的NFA确定,最后再根据此设计编译程序的词法分析器<sup>[2]</sup>。

定义在输入字符表  $\{A, C, M, F, H, S, Z, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$  上。对于  $*$  (由字符表中字符组合成的所有字符串) 中的任何字符串,若存在一条从初态结点到终态结点的通路,且这条通路上所有弧的标记字符依序连接成的字符串(忽略那些标记为  $\epsilon$  的弧)等于,则称字符串可为NFA所识别。

如航班号 CZ3501,在图1中的转化顺序为: X - 0 - 1 - 8 - 10 - 20 - 32 - 33 - 39 - 44 - 45 - 46 - 56 - 57 - 59 - 68 - Y,因此该航班号能够被识别;而航班号 AB1001 不能够被识别,即不符合航班号组成规则。这样通过构造不确定有限自动机实现了对航班号的规则检查,同理可以建立对机号、起始到达站、时间等其他8个关键词的不确定有限自动机、分析器。

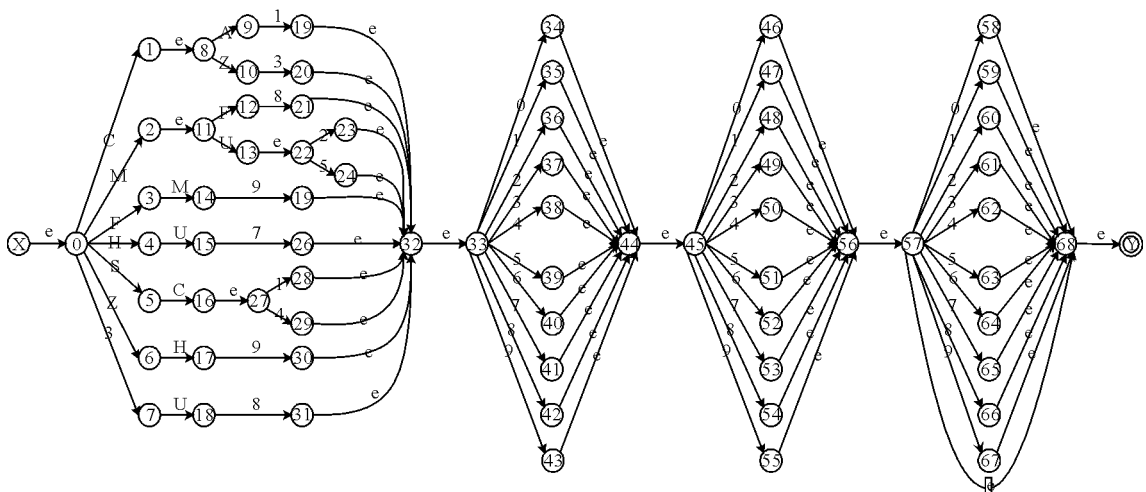


图1 航班号 NFA 状态转换图

### 2.3 不确定有限自动机 NFA 的确定化

一个确定的有限自动机(DFA)  $M$  是一个5元式,  $M = (S, \Sigma, f, S_0, Z)$ 。式中  $S$  为一个有

限集,它的每个元素称为一个状态; $\Sigma$  为一个有穷字母表,它的每个元素称为一个输入字符; $f$  为一个从  $S \times \Sigma$  至  $S$  的(单值)部分映射,  $f(s, a)$

$= s$  意味着:当现行状态为  $s$ , 输入字符  $a$  时, 将转换到下一个状态  $s$ 。我们把  $s$  称为  $s$  的一个后继状态;  $s_0 = s$ , 是惟一的一个初态;  $Z \in S$ , 是一个终态(可空)。

设已给具有 动作的非确定有限自动机

$$M = (K, \quad, f, s_0, Z)$$

则利用子集法构造相应的 DFA

$$M = (K, \quad, f, q_0, Z)$$

基本思路是:首先把从  $s_0$  出发, 仅经过任意条 矢线所能到达的状态所组成的集合作为  $M'$  的初态  $q_0$ , 然后分别把从  $q_0$  出发, 经过对输入符号  $a$  的状态转移所能到达的状态(包括转移后再经 矢线所能到达的状态)组成的集合作为  $M$  的状态, 如此等等, 直到不再有新的状态出现为止。具体地说, 构造  $K$  及  $f$  的步骤可递归地描述如下:

1) 令  $K = \text{CLOSURE}(s_0)$ ; (给出  $M$  的初态  $q_0$ )。把从某一给定的状态  $q$  出发, 仅经过所有标记为 的矢线所能到达的状态所组成的集合, 包括  $q$  本身, 记做  $\text{CLOSURE}(q)$ 。

2) 对于  $K$  中任一尚未标记的状态  $q_i = \{s_{i1}, s_{i2}, \dots, s_{im}\}, s_{ik} \in K$  做:

(1) 标记  $q_i$ 。

(2) 对于每个  $a$ , 置

$$T = f(\{s_{i1}, s_{i2}, \dots, s_{im}\}, a)$$

$$q_j = \text{CLOSURE}(T)。$$

(3) 若  $q_j$  不在  $K$  中, 则将  $q_j$  作为一个未加标记的状态添加到  $K$  中, 且把状态转移

$$f(q_i, a) = q_j \text{ 添加到 } M。$$

3) 重复进行步骤 2, 直到  $K$  中不再含有未标记的状态为止。对于由此构造的  $K$ , 把那些至少含有一个  $Z$  中的元素的  $q_i$  作为  $M$  的终态<sup>[3]</sup>。

## 2.4 编写航班计划文法

语法分析器接收词法分析器提供的记号串, 检查它们是否能由源程序语言的文法产生。编译器常用的是自顶向下和自底向上的方法。自顶向下语法分析器沿着从顶向底的方向建立分析树, 而自底向上语法分析器则沿着从底向顶的方向建立分析树。

首先对航班计划编写上下文无关文法。上下文无关文法由终结符、非终结符、开始符号和产生式组成。针对最基本的语法规则, 即检查航班计划中的关键词是否按照格式规定的词序构成来编

写一个简单的文法。根据航班计划约定:

下列符号是终结符: (1) 字母表中的大写字母, 如 A、B、C 等。(2) 数字 0、1、2、...、9。

小写字母是非终结符,  $s$  为开始符号。

按照上述约定, 编写航班计划的上下文无关文法: (1)  $s \rightarrow abccdddce$ ; (2)  $a \rightarrow gd$ ; (3)  $d \rightarrow fd|f$ ; (4)  $f \rightarrow 0|1|2|3|4|5|6|7|8|9$ ;

(5)  $b \rightarrow hgd$ ; (6)  $h \rightarrow gl|3$ ; (7)  $c \rightarrow gc|g$ ; (8)  $g \rightarrow A|B|C|D|E|F|G|H|I|J$

$|K|L|M|N|O|P|Q|R|S|T|U|V|W|X|Y|Z$ ; (9)  $e \rightarrow je|j$ ; (10)  $j \rightarrow .|f$

## 2.5 构造语法分析树

语法分析树可以看成是文法推导的图形表示。语法分析树是一棵有序有向树, 每个节点都有标记。根结点代表文法的识别符号; 每个内部结点(非叶结点)表示一个非终结符号, 其子结点由这个非终结符号在这次推导中, 所用产生式的右部各个符号代表的结点组成; 每个末端结点(叶结点)代表终结符号或非终结符号。

语法树的根结由开始符号所标记。随着推导的展开, 当某个非终结符被它的某个候选式所替换时, 这个非终结符的相应节就产生出下一代新节, 候选式中自左至右的每个符号对应一个新结, 并用这些符号标记其相应的新结。每个新结和其父结间都有一条连线。在一棵语法树生长过程中的任何时刻, 所有那些没有后代的段末结自左至右排列起来就是一个句型。

以表 1 所列航班计划为例:

"B1001 CA3501 PEK SHA 0830 1020 320 FJ Y 1234567",

经过词法分析后, 得到处理后的字符串:

"B1001CA3501PEKSHA08301020320 FJ Y 1234567",

此时其中的词法已经被检查过, 提交给语法分析器的字符串在词法上符合相应的规则。按照以上文法可以得到图 2 所示的语法分析树。

## 2.6 采用 LR 语法分析器归约

LR 分析法是一种有效的自底向上语法分析技术,  $L$  指的是从左向右扫描输入字符串,  $R$  指的是构造最右推导的逆过程。LR 语法分析器可以识别几乎所有能用上下文无关文法描述的语言结构, 是已知的最一般的无回溯移动归约分析法, 能及时发现语法错误<sup>[4]</sup>。

自底向上的分析使用先进后出的分析栈存放

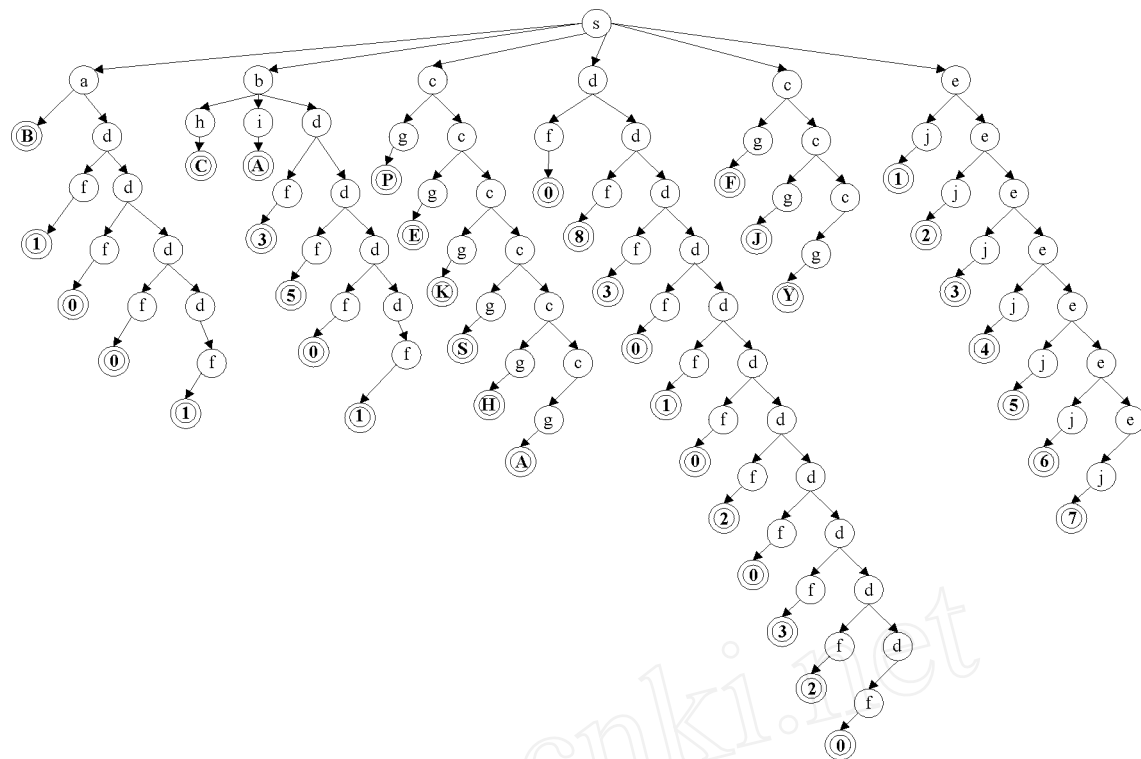


图 2 航班计划语法分析树

文法符号。分析开始时,在栈底放置一个届符 # ,然后将输入符号逐个移入栈内,一旦在分析栈的栈顶出现句柄,即进行一次归约;反之则继续将后续的输入符号移入栈内,并重复上述过程;若最终能将全部输入符号(不包括右界符 # )移掉,且分析栈中只留下栈底符号 # 及最后一步归约所得到的文法开始符号,则表明对输入串的分析已成功,反之则表明输入符号串不是文法的一个句子,存在语法错误。

以表 1 所列航班计划为例:

“B1001 CA3501 PEK SHA 0830 1020 320 FJ Y 1234567 ”,

经过词法分析后,得到处理后的字符串:

“B1001CA3501PEKSHA08301020320FJ Y 1234567 ”,

表 2 表明了对通过词法分析后得到字符串用 LR 分析进行归约的过程。S 表示“移进”动作,R<sub>j</sub>表示采用文法的第 j 条产生式进行归约,ACC 表示分析成功。

3 结束语

本文根据航班计划编排的实际需要,提出应用有限自动机解决航班计划编排中的航班句法检查问题,给出了航班计划句法检查不确定有限自

动机的设计思路,验证了使用有限自动机进行航班计划句法检查的可行性。

表 2 航班计划 LR 语法分析器归约过程

步骤	分析栈的内容	余留输入符号串	下步动作
0	#	B1001CA3501PEKSHA 08301020320FJ Y1234567 #	S
1	# B	1001CA3501PEKSHA 08301020320FJ Y1234567 #	S
2	# B1	001CA3501PEKSHA 08301020320FJ Y1234567 #	R4
3	# Bf	001CA3501PEKSHA 08301020320FJ Y1234567 #	S
4	# Bf0	01CA3501PEKSHA 08301020320FJ Y1234567 #	R4
5	# Bff	01CA3501PEKSHA 08301020320FJ Y1234567 #	S
6	# Bff0	1CA3501PEKSHA 08301020320FJ Y1234567 #	R4
...	...	...	...
105	# abcdcjijje		# R10
106	# abcdcjije		# R10
107	# abcdcje		# R10
108	# abcdce		# R10
109	# abcdce		# R1
110	# S		# ACC

参考文献

[1] 李劲华,丁洁玉.编译原理与技术[M].北京:北京邮电大学出版社出版,2006

[2] 陈火旺,刘春林,谭庆平,等.程序设计语言——编译原理[M].北京:国防工业出版社,2003

[3] 徐 红. 对确定有限自动机最小化算法的改进[J]. 桂林航天工业学校学报, 2005, 10(4): 14-16

[4] 蒋立源, 康慕宁. 编译原理[M]. 西安: 西北工业大学出版社, 2002

## Finite Automata-based Technology in Syntax Checking Methods of the Flight Scheduling

KAN Haijian XIA Hongshan

(Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

**Abstract :** In order to improve the accuracy and integrity of automatic inspection and analysis capabilities of the flight scheduling, this paper analyzed the characteristics of the flight scheduling. Then, it presented the nondeterministic finite automata-based technology in syntax checking methods of the flight scheduling, and carried out the lexical analysis, grammar analysis and inspection rules for Syntax check of flight scheduling with the analysis of examples. The feasibility of the method to use nondeterministic finite automata in the flight scheduling syntax check was proved.

**Key words :** finite automata; lexical analysis; syntax analysis; flight schedule; rules inspection

(上接第 34 页)

[8] 潘东来. 城市轨道交通枢纽交通衔接研究[D]. 武汉: 华中科技大学, 2005

社, 1994

[9] 何新贵. 模糊数据库系统[M]. 北京: 清华大学出版

[10] 魏 延. 主动模糊数据库中的事件与规则[J]. 重庆师范学院学报, 2002(12): 41-45

## Automatic Estimation of Railway Passenger Station's Traffic Connection

YANG Jiaqi ZHOU Ying

(Wuhan University of Technology, Wuhan 430063, CHINA)

**Abstract :** First, the current situation and existing problems about the traffic connection of railway passenger station were analyzed. Then, the four specific indexes including modes of transfer matching, transfer time, comfort of transfer, and transfer economy were put forward. A fuzzy relation data model was built. The active fuzzy database technology was applied to fuzzily implementing the automatic estimation of traffic connection.

**Key words :** railway passenger station; traffic connection; fuzzy database; estimation

## 振兴物流业确定九大重点工程

据新华社消息, 国务院总理温家宝 2 月 25 日主持召开国务院常务会议, 审议并原则通过物流业调整振兴规划。

会议指出, 物流业是融合运输、仓储、货运代理和信息等行业的复合型服务产业, 涉及领域广, 吸纳就业人数多, 促进生产、拉动消费作用大。但是我国物流业总体水平落后, 严重制约国民经济效益的提高。必须加快发展现代物流, 建立现代物流服务体系, 以物流服务促进其他产业发展。一要积极扩大物流市场需求, 促进物流企业与生产、商贸企业互动发展, 推进物流服务社会化和专业化。二要加快企业兼并重组, 培育一批服务水平高、国际竞争力强的大型现代物流企业。三要推动能源、矿产、汽车、农产品、医药等重点领域物流发展, 加快发展国际物流和保税物流。四要加强物流基础设施建设, 提高物流标准化程度和信息化水平。

会议确定了振兴物流业的九大重点工程, 包括多式联运和转运设施、物流园区、城市配送、大宗商品和农村物流、制造业和物流业联动发展、物流标准和技术推广、物流公共信息平台、物流科技攻关及应急物流等。会议要求, 各地区、各有关部门要加强组织协调, 深化物流管理体制改革, 完善政策法规体系, 多渠道增加投入, 加快物流人才培养, 促进我国物流业平稳较快发展。

摘自《中国交通报》2009 年 2 月 27 日