



(12) 发明专利

(10) 授权公告号 CN 101847407 B

(45) 授权公告日 2013.01.02

(21) 申请号 201010126315.2

审查员 贾杨

(22) 申请日 2010.03.12

(73) 专利权人 中山大学

地址 510006 广东省广州市番禺区大学城中山大学东校区教学实验中心C401

(72) 发明人 赵仲明 罗笑南 杨彪

(51) Int. Cl.

G10L 15/22(2006.01)

G10L 15/18(2013.01)

G10L 15/26(2006.01)

(56) 对比文件

CN 1392473 A, 2003.01.22,

CN 101589427 A, 2009.11.25,

CN 101271689 A, 2008.09.24,

CN 101669116 A, 2010.03.10,

权利要求书 1 页 说明书 6 页 附图 5 页

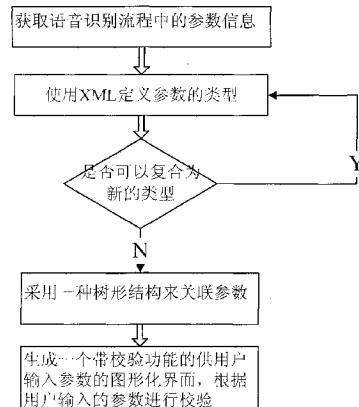
(54) 发明名称

一种基于 XML 的语音识别参数处理方法

(57) 摘要

本发明公开了一种基于 XML 的语音识别参数处理方法。该方法包括：1) 获取语音识别流程中的语音识别参数的信息；2) 通过 XML 定义不同语音识别参数的类型，并通过语音识别参数的逻辑关系复合出一个名称为 DeelType 的数据类型结构，其中包含了段号、语法文件号、反馈音文件、返回结果的参数信息；3) 采用树形结构表示各语音识别参数相互之间的联系；4) 生成一个带校验功能的供用户输入语音识别参数的图形化界面，根据用户输入的语音识别参数进行校验。本发明的技术方案可以使得语音识别参数的输入和传递更为方便与简洁，并且通过生成新的参数类型来更好的管理语音识别参数。

B



1. 一种基于 XML 的语音识别参数处理方法,其特征在于,包括以下步骤:

1) 获取语音识别流程中的语音识别参数的信息,所述语音识别参数的信息包括:在语音识别过程中所用到的文件信息、数据结构、数据之间的关系以及数据本身的约束信息;

2) 通过 XML 定义不同语音识别参数的类型,并通过语音识别参数的逻辑关系复合出一个名称为 DeelType 的数据类型结构,其中包含了段号、语法文件号、反馈音文件、返回结果的参数信息;

3) 采用树形结构表示各语音识别参数相互之间的联系;

4) 生成一个带校验功能的供用户输入语音识别参数的图形化界面,根据用户输入的语音识别参数进行校验。

2. 根据权利要求 1 所述的基于 XML 的语音识别参数处理方法,其特征在于:

所述采用树形结构表示各语音识别参数相互之间的联系包括:

首先,设置一个根节点 Root;

然后,根据语音识别参数读取的顺序,以及语音识别参数在传递过程中出现的次数设置根节点的子节点或者是叶子节点。

3. 根据权利要求 1 所述的基于 XML 的语音识别参数处理方法,其特征在于:

所述生成一个带校验功能的供用户输入语音识别参数的图形化界面,根据用户输入的语音识别参数进行校验,依照如下步骤执行:

步骤 1:使用软件生成一个 .sps 文件,并将该文件与已经生成的 .xml 文件和 .xsd 文件关联;

步骤 2:导入 .xsd 文件中的树形结构和参数类型,并将其以表格的形式显示,并调整整个界面;

步骤 3:判断是否存在需要在输入端检验的语音识别参数,如果有,则执行步骤 4,否则跳转至步骤 6;

步骤 4:在 .xsd 文件中嵌入正则表达式代码,实现对于输入语音识别参数的检验;

步骤 5:更新 .sps 文件,并跳转至步骤 3;

步骤 6:结束。

4. 根据权利要求 1 所述的基于 XML 的语音识别参数处理方法,其特征在于:

所述图形化界面列出所有需要用户输入的参数类型,并给出用户输入的提示信息以提示用户正确的输入语音识别参数;

所述图形化界面提供对于语音识别参数的管理,包括进行添加、修改、删除操作。

一种基于 XML 的语音识别参数处理方法

技术领域

[0001] 本发明涉及数字家庭技术领域,具体涉及一种基于 XML 的语音识别参数处理方法。

背景技术

[0002] 随着科学技术的不断进步,经济的不断发展,家用电器的智能化程度越来越高,也有越来越多的人致力于数字化家庭的研究和探索。而在数字化家庭技术中的一个重要环节就是家用设备对于人的语音的识别。语音识别,简单而言就是让机器通过识别和理解过程把语音信号转变为相应的文本或命令。根据针对的发音人,可以把语音识别技术分为特定人语音识别和非特定人语音识别,前者只能识别一个或几个人的语音,而后者则可以被任何人使用。显然,非特定人语音识别系统更符合实际需要,但它要比针对特定人的识别困难得多。

[0003] 另外,根据语音设备和通道,可以分为桌面(PC)语音识别、电话语音识别和嵌入式设备(手机、PDA等)语音识别。不同的采集通道会使人的发音的声学特性发生变形,因此需要构造各自的识别系统。

[0004] 语音识别的应用领域非常广泛,常见的应用系统有:语音输入系统,相对于键盘输入方法,它更符合人的日常习惯,也更自然、更高效;语音控制系统,即用语音来控制设备的运行,相对于手动控制来说更加快捷、方便,可以用在诸如工业控制、语音拨号系统、智能家电、声控智能玩具等许多领域;智能对话查询系统,根据客户的语音进行操作,为用户提供自然、友好的数据库检索服务,例如家庭服务、宾馆服务、旅行社服务系统、订票系统、医疗服务、银行服务、股票查询服务等等。

[0005] 现有技术之中,在传统的语音识别过程中,语音识别参数是通过文本文件或是程序来定义,由于其本身的局限性或是为了方便运输,导致了某些参数被设定为不符合实际情况的类型;另外,在传统的语音识别参数传递过程中,参数之间是一种平行结构。也就是说,各个参数都是独立的被存储和传输的。这样,由于各种参数类型的不同和结构的变化使得在读取这些参数时十分的繁琐,并且容易出现遗漏。因此,现有技术的语言识别参数处理方法存在缺陷。

发明内容

[0006] 本发明的目的在于提供一种基于 XML 的语音识别参数处理方法,可以使得语言识别参数的输入和传递更为方便与简洁,并且通过生成新的参数类型来更好的管理语音识别参数。

[0007] 为实现本发明目的,本发明提供的技术方案如下:

[0008] 本发明提供一种基于 XML 的语音识别参数处理方法,包括以下步骤:

[0009] 1) 获取语音识别流程中的语音识别参数的信息,所述语音识别参数的信息包括:在语音识别过程中所用到的文件信息、数据结构、数据之间的关系以及数据本身約束信

息；

[0010] 2) 通过 XML 定义不同语音识别参数的类型，并通过语音识别参数的逻辑关系复合出一个名称为 Dee1Type 的数据类型结构，其中包含了段号、语法文件号、反馈音文件、返回结果的参数信息；

[0011] 3) 采用树形结构表示各语音识别参数相互之间的联系；

[0012] 4) 生成一个带校验功能的供用户输入语音识别参数的图形化界面，根据用户输入的语音识别参数进行校验。

[0013] 优选的，所述采用树形结构表示各语音识别参数相互之间的联系包括：

[0014] 首先，设置一个根节点 Root；

[0015] 然后，根据语音识别参数读取的顺序，以及语音识别参数在传递过程中出现的次数设置根节点的子节点或者是叶子节点。

[0016] 优选的，所述生成一个带校验功能的供用户输入语音识别参数的图形化界面，根据用户输入的语音识别参数进行校验，依照如下步骤执行：

[0017] 步骤 1：使用软件生成一个 .sps 文件，并将该文件与已经生成的 .xml 文件和 .xsd 文件关联；

[0018] 步骤 2：导入 .xsd 文件中的树形结构和参数类型，并将其以表格的形式显示，并调整整个界面；

[0019] 步骤 3：判断是否存在需要在输入端检验的语音识别参数，如果有，则执行步骤 4，否则跳转至步骤 6；

[0020] 步骤 4：在 .xsd 文件中嵌入正则表达式代码，实现对于输入语音识别参数的检验；

[0021] 步骤 5：更新 .sps 文件，并跳转至步骤 3；

[0022] 步骤 6：结束。

[0023] 优选的，所述图形化界面列出所有需要用户输入的参数类型，并给出用户输入的提示信息以提示用户正确的输入语音识别参数；所述图形化界面提供对于语音识别参数的管理，包括进行添加、修改、删除操作。

[0024] 上述的技术方案可以看出，本发明的有益效果在于：

[0025] 通过将 XML 引入到语音识别的参数处理过程中，实现了对于语音识别参数的高质量管理。并且，通过 XML 文件、SPS 文件和 XSD 文件之间的相互关联，为语音识别参数的传递和扩展提供了保障。而由于 XML 自身的高度可扩展性，使得这种语音识别参数处理方法也具有很强的可扩展性，为日后语音识别参数的扩展打下了良好的基础。同时，还实现了语音识别参数的在输入端的校验工作。

附图说明

[0026] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其它的附图。

[0027] 图 1 是本发明基于 XML 的语音识别参数管理方法的流程图；

- [0028] 图 2 是本发明定义的 DeelType 数据类型示意图；
- [0029] 图 3 是本发明语音识别参数的树形结构示意图；
- [0030] 图 4 是本发明 xml 文件, .xsd 文件和 .sps 文件三者之间的关系示意图。
- [0031] 图 5 是本发明最终完成的图形化语音识别参数输入界面的示意图。

具体实施方式

[0032] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0033] 本发明提供一种基于 XML 的语音识别参数处理方法,可以使得语音识别参数的输入和传递更为方便与简洁,并且通过生成新的参数类型来更好的管理语音识别参数。而由于 XML 自身的高度可扩展性,使得这种语音识别参数处理方法也具有很强的可扩展性,为日后语音识别参数的扩展打下了良好的基础。同时,还实现了语音识别参数的在输入端的校验工作。

[0034] 本发明方法,利用了现今国际标准流行的描述语言 -XML(Extensible Markup Language, 可扩展标记语言),来描述语音互动过程中需要的语音识别参数文件及比较复杂的流程结构,并且能够在用户层提供对于语音识别参数的检验功能,使得语音识别参数的传递具有清晰、可扩展等优良特性。

[0035] XML 与 HTML 一样,都是 SGML(Standard Generalized Markup Language, 标准通用标记语言)。XML 是 Internet 环境中跨平台的,依赖于内容的技术,是当前处理结构化文档信息的有力工具。XML 也是一种简单数据存储语言,使用一系列简单的标记描述数据,而这些标记可以用方便的方式建立,并且易于掌握和使用。XML 的简单使其易于在任何应用程序中读写数据,这使 XML 很快成为数据交换的唯一公共语言。XML 保留了 SGML 的结构化功能,这样也使得网站设计者可以定义自己的文档类型,XML 同时也推出一种新型文档类型,使得开发者也可以不必定义文档类型。

- [0036] 本发明的方法主要采用了以下处理方式：

- [0037] 1) 使用 XML 定义各种语音识别参数的类型；

[0038] 在传统的语音识别过程中,语音识别参数是通过文本文件或是程序来定义,由于其本身的局限性或是为了方便运输,导致了某些语音识别参数不得不被设定为不符合实际情况的类型。在本发明中,使用 XML 来定义各种参数类型(通过生成一个与 XML 相关联的 XSD 文件, XSD 是指 XML 结构定义 (XML Schemas Definition)),不仅为每个语音识别参数选取了恰当的类型,还将几个语音识别参数根据其相互之间的逻辑关系复合起来生成一个新的数据结构。

[0039] 在本发明中,将各种语音识别参数定义为一些基本数据类型,并且根据语法文件、反馈语音文件、识别返回结果数等参数信息,生成一个新的名为 DeelType 的数据类型。使用这个新的数据类型,在 XML 中定义类型名为 DeelType 的数据,极大方便了语音数据在 XML 中的存储和传输。

- [0040] 2) 采用一种树形结构来表示语音识别参数相互之间的联系；

[0041] 在传统的语音识别参数传递过程中,语音识别参数之间是一种平行结构。也就是说,各个语音识别参数都是独立的被存储和传输的。这样,由于各种语音识别参数类型的不同和结构的变化使得在读取这些语音识别参数时十分繁琐,并且容易出现遗漏。

[0042] 而在本发明中,结合 XML 的特性,采用一种树形结构来保存各种语音识别参数及相互之间的关系,不仅方便了数据的保存,而且使得整个语音识别参数的便利过程变得十分简单。

[0043] 3) 设计一个带校验功能的供用户输入语音识别参数的图形化界面;

[0044] 传统的语音识别参数传递大多依靠单一的文本文件,在语音识别参数输入方面十分不便,并且在输入语音识别参数的同时也无法对于输入语音识别参数的正确性进行验证。

[0045] 在本发明中,通过使用 XML 相关软件设计出一个类似于 HTML 的可供用户输入语音识别参数的图形化界面(即生成一个与 XML 文件相关联的 .SPS 文件)。在该图形化界面(即 .SPS 文件)上可以清楚的列出所有需要用户输入的参数类型,并给出用户输入的提示信息以提示用户能正确的输入语音识别参数;由于该界面是直接与 XML 参数文件向关联,所以在该界面上还可以直接实现对于语音识别参数的管理,如添加,修改,删除。

[0046] 另外,通过在 xsd 文件嵌入正则表达式代码,可以在图形化界面上即时的对于用户所输入的语音识别参数进行检验,极大的减轻了后台语音识别参数处理程序的负担,同时也高质量地保证了语音识别参数的输入。

[0047] 以下结合附图详细介绍本发明的方案。

[0048] 如图 1 所示是本发明处理方法的流程图,主要包括以下步骤:

[0049] (1) 获取语音识别流程中的语音识别参数信息;

[0050] 在本发明中所指的语音识别参数信息,主要是指在语音识别过程中所用到的文件信息、数据结构、数据之间的关系以及数据本身的一些约束等。例如:语法文件、语音文件、语法文件个数等等。基于对语音识别流程的充分理解,将在语音识别过程中所要用到的所有语音识别参数都提取出来,为后面的操作做好准备。

[0051] (2) 使用 XML 定义各种语音识别参数的类型;

[0052] 在本发明中,用 XML 来定义各种语音识别参数类型(通过生成一个与 XML 相关联的 .xsd 文件)。

[0053] 首先,为已知的各种语音识别参数设定一个数据类型。

[0054] 然后,判断出一些语音识别参数之间的逻辑关系,如:语法文件、提示音、动作编号等语音识别参数总是会以一个整体的形式出现。

[0055] (3) 判断是否可以复合为新的类型,若是,返回步骤(2),若否,进入(4);

[0056] 若可以复合为新的类型,则将一些具有逻辑相关性的语音识别参数复合为一个整体,然后继续判断是否可以在这个整体中添加新的语音识别参数,直到将所有的语音识别参数都按其逻辑关系分类,并生成新的参数类型。

[0057] 在本发明中,新生成一个名为 DeelType 的数据类型,其中包含了段号、语法文件号、反馈音文件、返回结果等多个参数信息。使得 DeelType 成为语音识别中基本的数据单元被添加,删除或修改,而且也作为基本的数据单元被传输。

[0058] 在附图中 2 中可以看到新定义的 DeelType 数据类型的格式。

[0059] (4) 采用一种树形结构来关联语音识别参数；

[0060] 在本发明中,结合 XML 的树形结构特性,本发明使用了一种树形结构来保存各种语音识别参数及相互之间的关系。这样不仅方便了数据的保存,而且使得整个语音识别参数的便利过程变得十分简单。

[0061] 如图 3 所示,是本发明语音识别参数的树形结构示意图。

[0062] 首先,为整个 XML 设置一个根节点 Root ;

[0063] 然后,根据语音识别参数读取的先后关系和顺序,以及语音识别参数在传递过程中出现的次数分别设置为根节点的子节点,或者是叶子节点,例如图中根节点 Root 下面分为不同子节点 :Deel, Number_of_deel, audi_files。节点 Deel 和 audi_files。下面再接着设置子节点。整个语音识别参数的树形结构如图 3 所示。

[0064] (5) 生成一个带校验功能的供用户输入语音识别参数的图形化界面,根据用户输入的语音识别参数进行校验。

[0065] 依照如下步骤执行 :

[0066] 步骤 1 :使用 XML 相关软件生成一个 .sps 文件,并将该文件与已经生成的 .xml 文件和 .xsd 文件关联。

[0067] 步骤 2 :导入 .xsd 文件中的树形结构和参数类型,并将其以表格的形式显示,并调整整个界面。

[0068] 步骤 3 :判断是否存在需要在输入端检验的语音识别参数,如果有,则执行步骤 4,否则跳转至步骤 6。

[0069] 步骤 4 :在 .xsd 文件中嵌入正则表达式代码,实现对于输入语音识别参数的检验。

[0070] 步骤 5 :更新 .sps 文件,并跳转至步骤 3。

[0071] 步骤 6 :结束。

[0072] 这样,通过 .sps 文件与 .xml 文件的关联,使得可以通过在图形化界面 (.sps 文件) 上完成对于语音识别参数的输入和检验,并最终将数据保存到 .xml 文件中。

[0073] 附图 4 显示了在整个方法过程中 .xml 文件,.xsd 文件和 .sps 文件三者之间的关系。如图 4 所示,.xml 文件和 .xsd 文件的关联通过控制参数类型和结构实现,.xsd 文件和 .sps 文件的关联通过控制参数的检验实现,通过在图形化界面 (.sps 文件) 上完成对于语音识别参数的输入和检验,并最终将数据保存到 .xml 文件中。

[0074] 附图 5 是一个界面的示意图。附图 5 显示了最终完成的图形化语音识别参数输入界面。如图中显示的界面字段包括:“段号”、“语法文件名”、“提示语音文件”、“该段输出结果数”、“识别结果”、“反馈音文件”、“动作编号”等。

[0075] 那么,通过该可供用户输入语音识别参数的图形化参数输入界面,可以清楚的列出所有需要用户输入的参数类型,并给出用户输入的提示信息以提示用户能正确的输入语音识别参数;由于该界面是直接与 XML 参数文件向关联,所以在该界面上还可以直接实现对于语音识别参数的管理,如添加,修改,删除。

[0076] 另外,通过在 xsd 文件嵌入正则表达式代码,可以在图形化界面上即时的对于用户所输入的语音识别参数进行检验,极大的减轻了后台语音识别参数处理程序的负担,同时也高质量地保证了语音识别参数的输入。

[0077] 上述的技术方案可以看出,本发明具有以下有益效果 :

[0078] 本发明通过将 XML 引入到语音识别的参数处理过程中,实现了对于语音识别参数的高质量管理。并且,通过 XML 文件、SPS 文件和 XSD 文件之间的相互关联,为语音识别参数的传递和扩展提供了保障。而由于 XML 自身的高度可扩展性,使得这种语音识别参数处理方法也具有很强的可扩展性,为日后语音识别参数的扩展打下了良好的基础。同时,还实现了语音识别参数的在输入端的校验工作。

[0079] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通过程序来指令相关的硬件来完成,该程序可以存储于一计算机可读存储介质中,存储介质可以包括:只读存储器 (ROM, Read Only Memory)、随机存取存储器 (RAM, Random Access Memory)、磁盘或光盘等。

[0080] 以上对本发明实施例所提供的一种基于 XML 的语音识别参数处理方法,进行了详细介绍,本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本发明的限制。

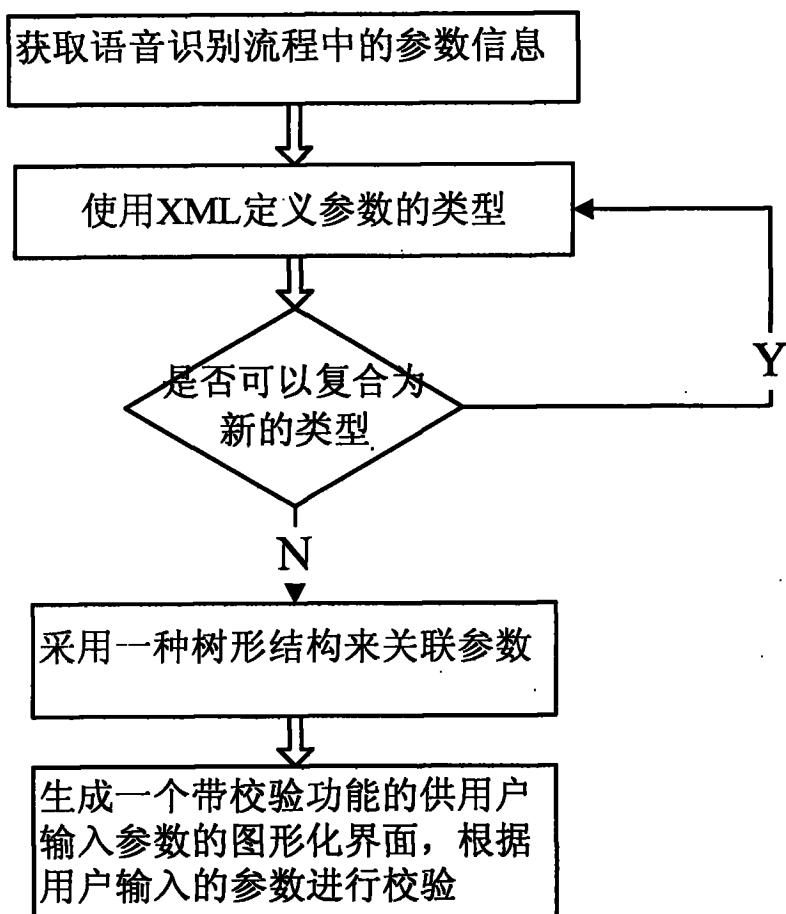


图 1

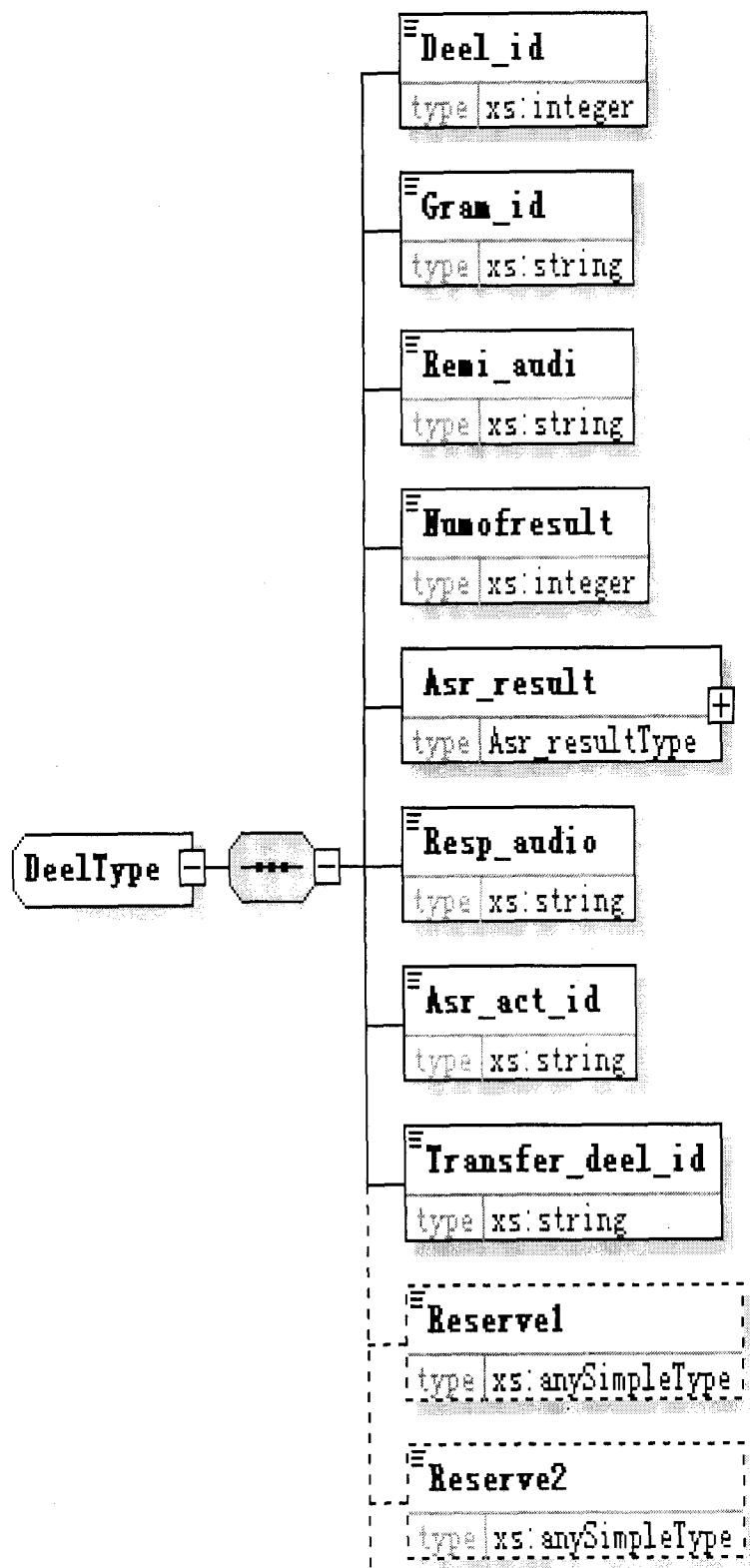


图 2

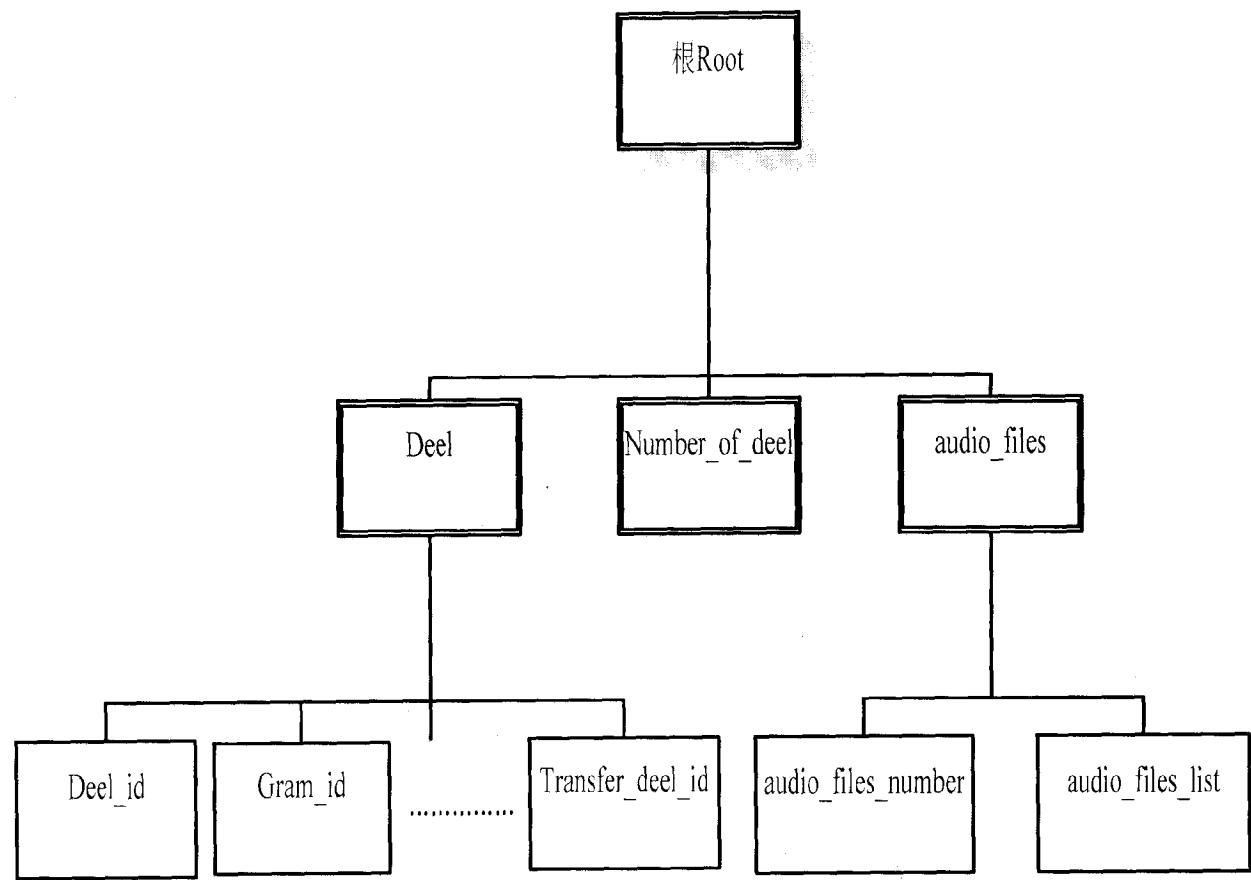


图 3

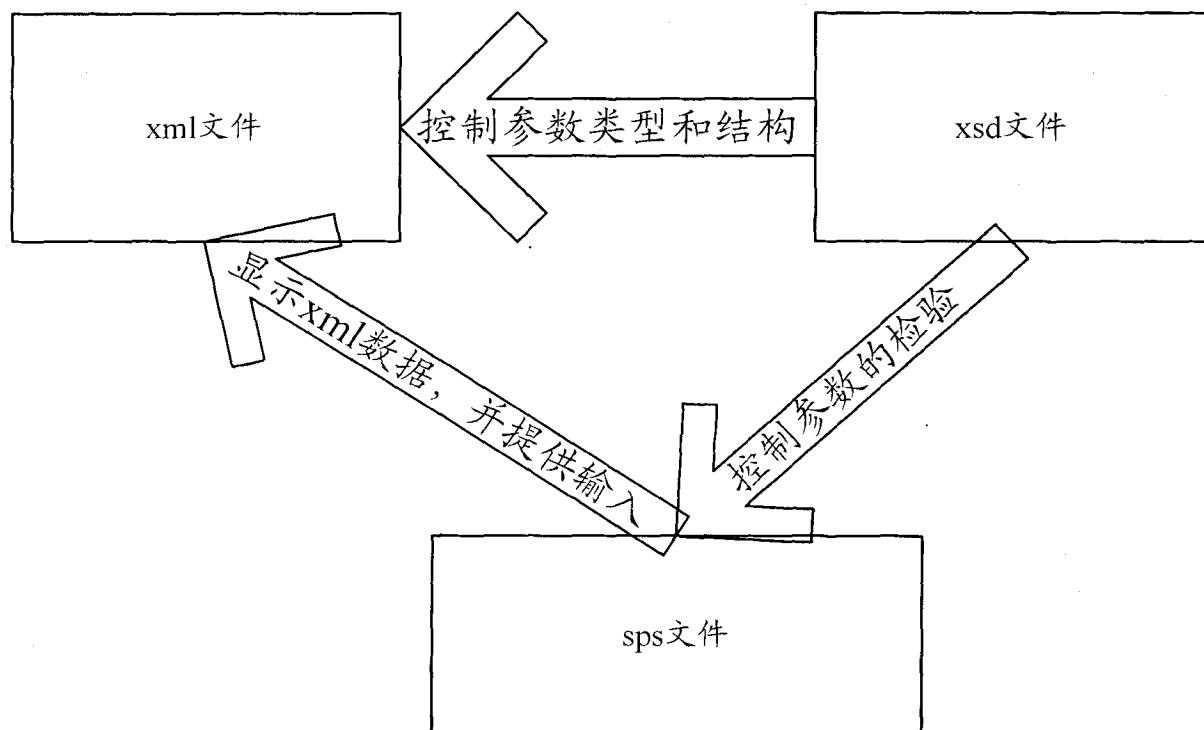


图 4

段号	语法文件名	该段输出结果 提示数	识别结果	反馈音文件	动作编号	跳转
			所有的输出情况	对应的编号		
0	s1 a1 3	完毕 操练完毕 操练完 毕		fsdffff1	a000 a211 ae31 e439	100 222 1
1	s2 a2 2	操练 军事操练 向左看 齐		-1	ferr	800 132
2	s3 a3 5	打 表演武打 滑冰 花 样滑冰 结束 娱乐结束 结束		-1	bh f	19 19 19 19
3	s6 a8 7	猜谜语 算算术题 讲故 事 侦探 扮大侦探 逻 辑		afad f1d1 f2d1	-1	21 24 27 28

图 5