

软件开发文档的数据化浅析

张 检

西南交通大学 四川成都 610000

摘 要: 软件开发中伴随着大量的软件文档, 软件文档质量管理成为软件开发中重要目标。质量的软件文档管理有助于提高软件开发的效率; 有助于软件开发中项目组的沟通; 提高软件文档质量和效率。基于此, 本文分析了软件开发文档数据化方法以供参考。

关键词: 软件文档数据化; 范式结构; 数据化文本

引言

软件文档是开展软件开发、设计、测试、运维的基础和依据, 它的质量会对软件开发、设计、测试和维护产生重大影响。在现在的软件工程管理中, 对软件代码、软件测试管理十分重视, 而对文档的质量评价、度量管理常常忽略。用数据化的方法构建软件开发文档, 使文档具有代码化、数据化特征。

一、软件开发文本类型

1. 自然语言文本

自然语言文本自由格式, 文本的表达没有规则和约束。这种表达方式更容易造成人理解的歧义性。客户需求具有模糊性, 而软件需求是建立在客户需求上的, 导致软件需求也有模糊性, 双重不稳定性, 不能定义软件清晰的功能, 最后软件功能的开发也不能得到保障。

2. 结构化文本

结构化文本 (Structure Data) 是指具有高度规则化格式和语法结构的文本, 如数据库表格中的数据, 它们遵循明确的格式和规则。这使得人们可以更容易地理解和使用。结构化的文档主要对文件的编制结构作了规则化要求, 譬如政府公文、论文、法律法规文件等。这类文件内容表达采用自然文本。譬如法律条款中的某一条款, 不同的人站在不同的角度, 往往有不同的解释, 针对不同的解释又要给一个司法解释。所以自然文本的表述方式, 在软件文本的开发中, 具有一定的模糊性、不确定下, 软件功能表述很难清晰的定义和质量量化。

3. 半结构化文本

半结构化文本 (Semi-Structure Data) 是指没有完全遵循任何预定义的结构或规则的文本, 但它们包含部分

规则化格式或语法结构。OEM (Object Exchange Model) 是斯坦福大学 (Stanford University) Papakonstantinou 等人提出的用来描述半结构化数据的数据模型, 它的主要特征是自描述性^[1]。Serge Abiteboul 定义半结构化数据为: 半结构化数据是指那些既不是完全无结构的, 也不是传统数据库系统中那样有严格结构的数据^[2]。我们常见半结构化数据有 HTML、XML、JSON、电子邮件、电子病历等都属于半结构化文档或者半结构化数据。半结构化数据是根据数据的主要特征来定义的, 即数据结构是隐含给出的, 数据之间只有松散的约束, 数据不是不规则、不完整的, 数据和结构交织在一起, 事先没有固定的模式可遵从, 数据结构是动态变化的需要从数据中进行抽取^[3]。半结构化数据在某种程度上可以说是兼顾了自由式文本和结构化文本两者的优点, 一方面它具有比较好的灵活性, 另外一方面通过协议本身的结构化的约束。

二、软件开发文档文本数据化模型

1. 软件开发文档文本巴科斯范式模型

软件功能是工作中业务流程的电子化, 把人工事件固化为电子事件。软件功能描述一般可以抽象模型为某人在某个条件下做某个事件; 做完后产生的一个结果。功能的描述一般可以描述为功能的名称, 该功能的使用人员、此功能要完成的事件、该事件的执行条件和该事件完成后的输出结果。

BNF (巴科斯范式 Backus-Naur Form, BNF) 主要通过定义元符号来形式化语法^[5]。BNF 具有简单易学、开放、易于扩展、分析、效率高等优点。BNF 提供形式化的语法描述, 具有一定的规则, 满足形式化的语法和语义, 避免描述的二义性; 同时又可以采用自然语言文本为补充, 能充分的表达和定义软件功能。

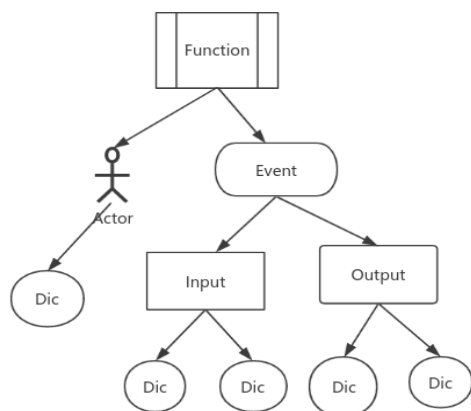


图1 软件功能数据化模型

软件功能文本数据化抽象为BNF可以描述为：

```
<function-name>::= //功能名称
<actor>: //使用人员
<action>: //事件/动作名称
[input-param]*=> //输入条件参数
<output> //输出产生结果
```

软件功能文本数据化方式，用范式方法表述软件功能，具有半规则化、数据化的特征，表述清晰准确、功能场景表述完备，方便沟通理解。

2. 软件开发文档文本数据化关系管理

每个软件都是由多个功能的集合，功能之间存在关联关系。软件功能中有些功能的前置条件是另外一个功能运行的结果。数据化的关系中，采用范式前后逻辑的流程可以function-name、actor、action、input-param和output构建软件功能关系。构成了数据化构建的多参数形式。所以function-name构成了功能函数集、actor构成了功能角色集、action构成了软件事件集。在通过数据化关系把数据集构建建成软件功能。

通过BNF范式中规则来表达软件功能的业务流程、业务逻辑；使用形式化、图形化的方式表述软件功能。软件功能的范式关系组成了软件功能结构体系关系，业务逻辑关系。把软件文档文本用数据化的形式编辑、构建；也可以用数据化的形式构建软件文档所要表达的软件功能。建立软件功能文本表述的标准，从而建立质量管理的量化数据。

3. 软件开发文档数据化模型

软件文档其核心是软件功能的表述，能够清楚的说明软件所有功能、软件的业务流程、软件的数据状态转变等。要求各个功能能够表述的准备、完整、无歧义；文档之间同功能的表述要一致。自由文本方式很难实现上面的要求，但用范式表述能够很好的实现上面的目标。

软件文档的BNF的构成包括如下几个部分：

Dictionary: //软件系统中数据属性（字典）
Actors: //软件中的操作者（角色）
Actions: //软件中操作的方法
Functions: //软件中的功能
Business Flow: //软件中的业务流
Status Flow: //软件中的数据状态流
ER（Entity-relationship） //软件中的实体关系

Actors是描述软件使用人员，通常代表一个对象。在BNF里面我们用角色Role有限集合表述。Actions是描述软件中某个功能的事件行为，我们用事件Event有限集合表述。Functions表述软功能的有限集合。数据集中主要有软件数据字典、用户数据集、事件数据集和功能数据集。

软件文档数据化把从各异的软件文本提升为统一模式，方便了文档的管理；又可以通过工具、算法提取半结构化数据，量化软件功能数据，用工具、算法实现文本的信息化审查；提高数据处理效率，也提高管理的质量和效率。

结语

综上所述采用数据化的软件文档，用结构化的数据描述软件的属性数据；用半结构化的数据描述软件的功能。通过一定的规则表述软件功能，使每一个功能的角色、事件、条件和结果都表述清晰、准确；无二义性。半结构化数据表述方式也有利于统一软件文档功能表达，根据模型编制软件功能文档，即使是新入行员工也能理解软件功能。结构化数据和半结构化数据更利于文档的自动化处理，更有利于自动化数据提取、自动化数据筛查和自动化软件文档功能审查。

参考文献

- [1] Papakonstantinou 等. Object exchange across heterogeneous information sources[C]. In Proc of the IEEE ICDF IEEE Computer Society Press, 1995: 251~260
- [2] Serge Abiteboul. Querying semi-structured data. Computer Science 1186, Database Theory. New York. Springer-Verlag, 1997. 1-18.
- [3] 曹文仙等. 半结构化数据OEM图应用[J]. 西安工程科技学院学报, 2007, 21(1): 92-95
- [4] Zhong J, Feng Y. Attribute-based access control policy specification language[J]. Journal of Southeast University, 2008, 24(3): 260-263