

基于 HL7V2.x 协议的消息构建与存储模型

范 潇,黄青松

(昆明理工大学 信息工程与自动化学院,昆明 650500)

(fanxiao2012@sohu.com)

摘 要:针对我国医疗信息系统不能很好地按照统一数据标准进行通信而阻碍医疗信息共享的问题,基于现有医疗信息通信标准 HL7V2.x 和消息解析理论,设计了将已有数据库中的字段构建成标准 HL7V2.x 消息和解析后的消息存储到数据库的方法。在此基础上,提出了一种优化消息构建和消息存储的模型,通过使用映射文件节省重复手工配置时间;并利用哈希表作为数据结构,从而提高消息构建的效率。最后通过仿真实验研究验证了该模型的可行性。

关键词:HL7V2.x 协议;映射文件;哈希表;医疗信息系统

中图分类号:TP315;TN915.04 **文献标志码:**A

Message generation and store model based on HL7V2.x protocol

FAN Xiao, HUANG Qing-song

(School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming Yunnan 650500, China)

Abstract: Data can not be communicated well according to the unified data standards among the health information systems in our country so that information sharing is impeded. On the basis of the HL7 V2.x (health level seven Version2.x) protocol of health information communication standard and the theory of HL7V2.x message parsing, the method was introduced in detail in which data in the database fields was extracted to generate HL7V2.x messages and parsed message information was stored in the specific database. Then an optimized model of the message store and the message generation was brought forward. In this model, mapping files were designed to save much time in redundant manual configurations and the hash table was used as a data structure during the process of message generation to make it more efficient. Finally, the simulation results verify the feasibility of the model.

Key words: HL7V2.x protocol; mapping file; hash table; health information system

0 引言

HL7(Health Level Seven)是一个符合 ANSI 的医疗数据交换标准,是由 Health Level Seven 标准组织制定的,相关操作位于国际标准规划组织(International Standards Organization, ISO)制定的开放式系统互联(Open System Interconnect, OSI)参考模型的第七层^[1]。HL7 标准的推行有助于在医疗信息系统中实现电子数据交换,能够消除或极大地减少用户接口编程和相应软件的维护需求。

目前,很多医疗机构拥有自己的医疗信息系统,但是由于这些系统没有采用统一的数据标准,从而阻碍了不同机构间医疗信息的共享与传递。HL7 组织发行的 HL7V2.x 协议相对成熟,该协议可以解决不同医疗机构之间信息互用与通信的难题,从而提高医疗服务效率和质量。文献[2-5]介绍了 HL7V2.x 协议标准及其应用背景和意义;文献[6]提出了 HL7 消息解析模型,部分医疗信息系统也已经实现了 HL7 消息的构建和存储,但是需要大量重复的手工配置工作,且由于 HL7 消息字段繁多,采用顺序遍历 HL7 消息字段来查找所要插入数据的位置效率低下。

为了减少大量重复的手工配置工作,本文研究了如何在消息构建或存储同时建立和存储映射文件,从而可以直接读

取映射文件中的映射关系完成后续类似消息的构建或存储。同时,在构建消息的过程中,本文采用了哈希存储结构,从而可以使用哈希查找代替顺序查找,实现了快速定位及数据插入。

1 HL7V2.x 消息构建模型研究

HL7 消息构建是在解析的基础上进行的^[6]。简单地说,消息构建是将数据源中的数据填充到模板消息中恰当的位置。以数据库作为数据源为例,首先需要找到数据库中表的字段名与模板消息中可以存储数据一级的节点的对应关系,存储数据一级的节点可能是消息中的字段、组件或者子组件的名字^[1],下文统称为消息字段名。

消息构建模型是基于 XML 映射文件和哈希存储结构设计的,从而可以避免重复手工配置和快速查找消息字段,最终提高消息的构建效率。模型数据流程图如图 1 所示。

1.1 基于 XML 映射文件的消息构建模型

首次配置映射关系是通过手工配置的,即在 HL7 消息字段名与数据库中表的字段形成映射关系。但是消息及消息字段数目巨大,每次消息构建都通过手工配置是一件十分耗时的事情。可以将映射关系存储在物理介质中,在 HL7 模板文

收稿日期:2011-06-07;修回日期:2011-07-29。

基金项目:国家科技部科技型中小企业技术创新基金资助项目(10C26215305099);云南省社会发展科技计划项目(2010CA016)。

作者简介:范潇(1986-),男,山西大同人,硕士研究生,主要研究方向:智能信息系统;黄青松(1962-),男,云南昆明人,教授,主要研究方向:智能信息系统、数据挖掘。

件和选择字段不变的情形下,通过应用系统读取映射文件,构造批量的 HL7 消息,从而提高效率。

存储在物理介质中的映射文件应该具备以下两个条件。
1) 跨平台、跨程序语言。在这一方面,XML 存储优于数据库的存储,数据库的存储要涉及到不同的数据库管理系统,如目前主流的数据库有 Oracle、SQL Server、MySQL、Sybase、PostgreSQL,如果将映射关系存储在上述某一个数据库中,将

无法适应其他数据库的使用,而使用 XML 存储不会出现这样的问题。2) 清晰反映映射关系,便于查看。XML 文件反映了一种树形的数据结构,树节点存储映射关系,结构清晰,便于阅读、维护。

数据流图如图 1 所示,数据流 $P1 \rightarrow P2 \rightarrow P5$, $P3 \rightarrow P4 \rightarrow P5$ 和 $P5 \rightarrow D3 \rightarrow P6$ 用于映射文件的形成和映射文件的查找,其中 D3 是 XML 格式的映射文件。

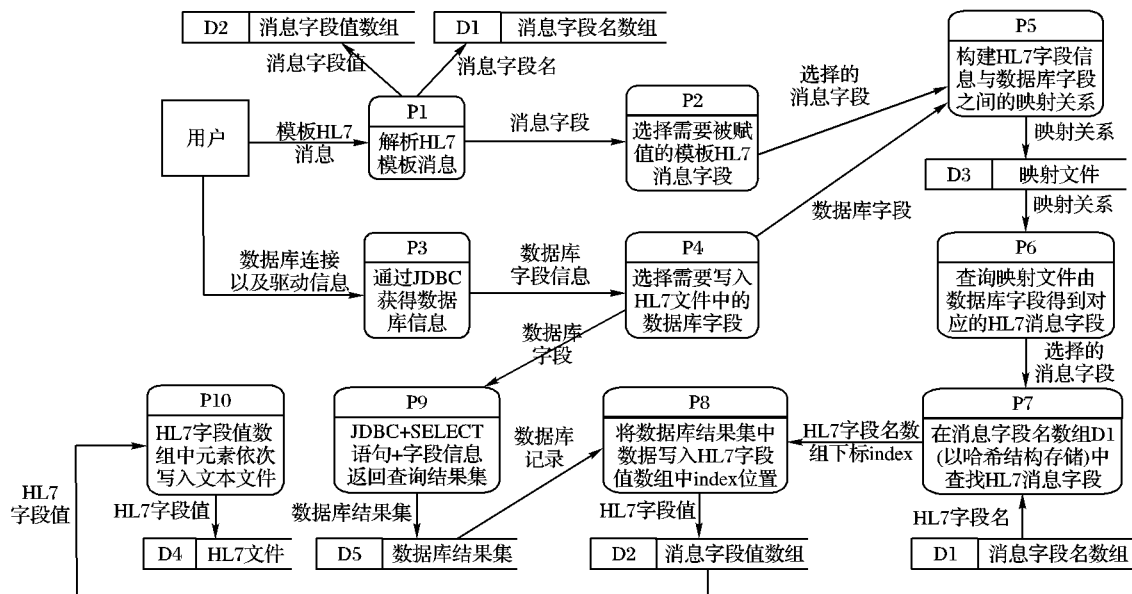


图 1 消息构建模型数据流图

1.2 基于哈希表的消息构建模型

如图 1 所示,数据流 $P6 \rightarrow P7 \rightarrow P8$, $D1 \rightarrow P7 \rightarrow P8$ 和 $P3 \rightarrow P4 \rightarrow P9 \rightarrow D5 \rightarrow P8 \rightarrow D2 \rightarrow P10$ 用于查找 HL7 消息字段名,数据库字段值对消息字段名的赋值和最终构建完成 HL7 消息写

入文件。

其中,数组 D1 中存放了经过解析的 HL7 消息字段名,数组 D2 中存储了消息字段名所对应的消息字段值,如表 1 所示。

表 1 消息字段名和消息字段值对应关系

数组名		数组元素(按数组下标从小到大排列)												
D1	PID	fs	PID.1	fs	PID.2	fs	PID.3.1	cs	PID.3.3	...	PID.3.7	cs	...	st
D2	PID		1				PATID1234	^	M11	...		^	...	<cr>

构建 HL7 消息关键步骤 $P7 \rightarrow P8$, 获得选择的字段名在数组 D1 中的下标 index, 由于数组 D1 中的消息字段名和数组 D2 中的消息字段值是一一对应的, 获得 index 后可以直接将数据库字段值直接写入 $D2[index]$, 最后再依次读取数组 D2 中的元素合成 HL7 消息。

首先,在此过程中,为了防止用户加载错误的映射文件,在 P7 处理前采用了 $D3 \rightarrow P6$ 步骤。该步骤确保了处理 P6 中查找消息字段时映射文件中的消息字段名存在于数组 D1 中,从而将对应的消息字段值正确写入 HL7 文件。

其次,HL7 字典中定义的段共 110 个,数据类型共有 52 个,其中的基本数据类型有 7 个^[1]。按照这样的组合,图 1 中,D1 消息字段名数组中的元素数目可能十分庞大。假设按照字段、组件、子组件三层之间数据类型的等级关系(有的情况不止 3 层,例如 CQ 数据类型是处于字段层次的,但是 CQ 却可以作为和它处于同一级的 TQ 数据类型的组件存在,这样的数据类型等级就达到了 4 层^[1]),最坏情况下将会有 $110 \times 45 \times 7 = 34\,650$ 个数据存入到 D1 数组中,所以如果每次顺序查找这个数组将会十分耗时。为了提高查询效率,P7 处理中使用了哈希表作为数据结构来存储 D1。由于数组元素不是数字而是字符串,哈希化之前需要对数组 D1 做以下步

骤的处理:

1) 数组 D1 中存放了一些无法描述 HL7 数据类型的元素,即不能存储值的元素,例如段 MSH、EVN 和大量的分隔符^[1],将这些无用的元素去掉。

2) 字符串的点号“.”没有提供任何有用的信息,对哈希化来说是多余的,没有必要进行哈希化,可以把数组中的点号去掉。

经过上述两个步骤处理后的数组 D1 在表 2 中列出。

表 2 三个步骤处理后的数组 D1

数组名	数组元素(按数组下标从小到大排列)						
D1	PID1	PID2	PID31	PID33	...	PID37	...

3) 哈希化的对象是字符串,应该把字符串映射成数字。此时,数组中元素形如 $\bigcirc \bigcirc \bigcirc \times \times$ 或 $\bigcirc \bigcirc \bigcirc \times$ (其中 \bigcirc 代表大写字母表示的范围是 A~Z, \times 代表的是数字范围是 1~9),共有 35 个字符。为了将字符串转换为数字,制定了数组中字母、数字与自然数的对应关系,即编码表(字符 1~9 表示自然数 1~9,字符 A~Z 代表自然数 10~35)。

在程序中,可以通过其 ASCII 码值减去固定值来实现将

1~9 和 A~Z 转换为表3中的值。由于数组中的字符串转换后的数值必须唯一,若出现重复的数据项填充到哈希表后将无法验证找到的字符串是否为目标字符串。因此,为了得到字符串对应的独一无二的数字,采取以下步骤实现,以○○○××格式为例进行说明。

步骤1 使用 Horner 方法^[7]建立式(1),其中 a_0, a_1, a_2, a_3, a_4 分别为形如 ○○○××字符串从左到右每个字符经过编码后对应的自然数, $n = 35$ 。

$$(((a_4 \times n + a_3) \times n + a_2) \times n + a_1) \times n + a_0 \quad (1)$$

程序实现式(1)时,从字符串最左边的字符开始,这样每次循环中只有一次乘法和加法,相对于公式展开式的幂运算提高了效率。

步骤2 设计哈希表时使用哈希函数,如式(2)。

$$\text{index} = \text{key} \% \text{hashSize} \quad (2)$$

其中:key为表2中的字符串所对应的数字,hashSize为哈希表的容量,%表示取余数。

步骤3 哈希化时可能带来冲突,解决冲突的方法有链地址法和开放地址法。装填因子 μ 是数据项数 N 和哈希表容量 hashSize 的比值,见式(3):

$$\mu = N / \text{hashSize} \quad (3)$$

其中,对于链地址法,如果数据项中采用的是有序链表,查找成功或失败时探测序列 p 和装填因子 μ 的关系是 $p = 1 + \mu/2$ 。

而对于开放地址法,在线性探测的成功查找中,探测序列 p 和装填因子 μ 的关系是 $p = (1 + \frac{1}{(1-\mu)^2})/2$,不成功查找

中 p 和 μ 的关系是 $p = (1 + \frac{1}{(1-\mu)})/2$;二次探测和再哈希法的性能相当,它们的性能比线性探测略好,对成功的搜索, $p = -(\ln(1-\mu))/\mu$,对于不成功的查找, $p = 1/(1-\mu)$ 。

图2分别显示了链地址法,开放地址法—线性探测、开放地址法—二次探测和再哈希法三种策略随着装填因子的增大,效率下降趋势^[7]

现实医疗工作中,HL7消息种类繁多,数据项数 N 的值不确定,而且可能会很大,从而造成装填因子的增大,从图2

比较来看,随着装填因子变大,开放地址法—线性探测的性能、开放地址法—二次探测和再哈希法的性能会下降很快;但是链地址法的性能只能线性地下降。所以本模型中采用了链地址法来解决冲突。

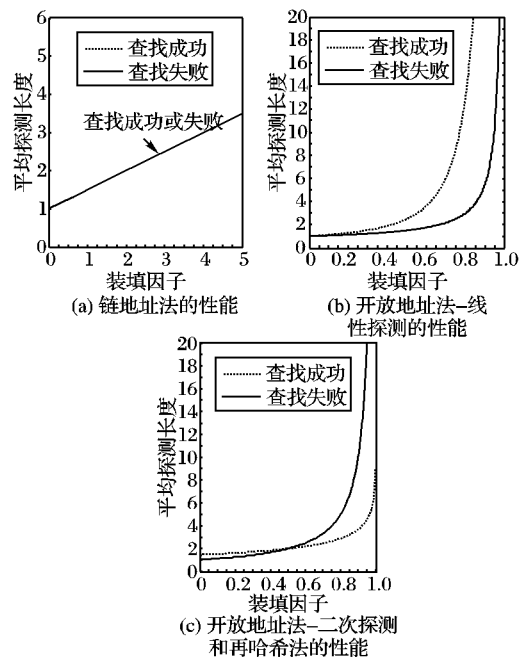


图2 三种策略随装填因子增大效率下降趋势

2 HL7V2.x 消息存储模型

将 HL7 消息存储到目标数据库是消息构建的逆过程,消息存储过程和消息构建过程原理相似,消息存储模型如数据流图3所示。主要流程如下:步骤{P3→P4, P9→P10}→P5→D3生成映射文件D3,步骤P1→P2→{P6, P7}→P8完成映射文件的读取、查询和将消息字段名插入到数据库字段中。映射文件D3也是基于XML格式设计的,与1.1节中消息构建中的映射文件一样可以节省大量手工配置映射关系的时间,提高效率。

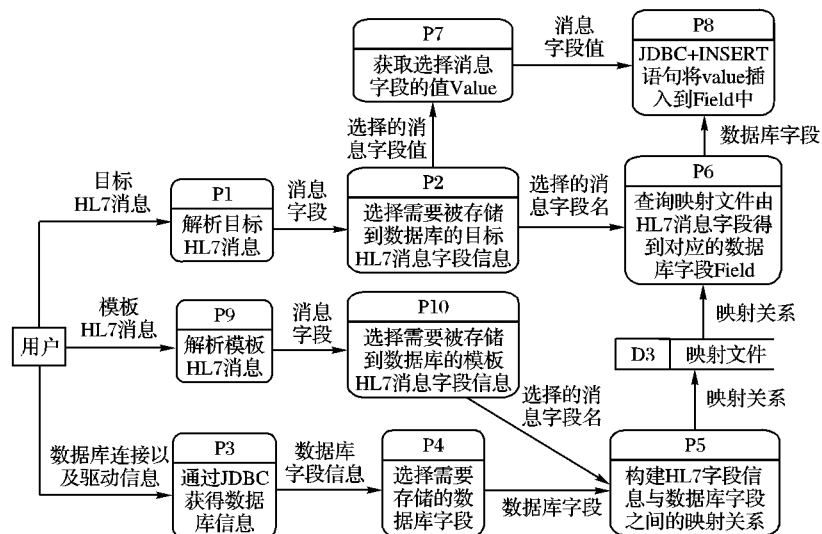


图3 HL7 消息存储模型数据流图

3 仿真实验验证及优化效果

为了验证模型是否具有优势,实验测试了8条不同的消息在相同的测试环境下使用哈希表和数组顺序查找使用的平

均查找时间。

实验的测试环境是 PC,各项配置列举如下,处理器: Intel Core(TM) i3 2.53 GHz,内存: 2.8 GB,硬盘: 300 GB,操作系统: Windows 7 家庭普通版,关系数据库系统: Oracle 10g,开发

工具:MyEclipse6.5。

基于上文分析,消息构建和存储实验中均使用映射文件存储映射关系,节省了手工重复配置对应关系的时间。消息构建实验中,使用了哈希表优化消息构建模型,哈希表的容量固定 $hashSize = 1000$ 。

从表 3 中可以发现使用链地址法来解决哈希表的冲突与数组顺序查找相比具有明显的优势,而且与使用开放地址法—线性探测、开放地址法—二次探测和再哈希法相比,随着装填因子的增大,尤其当装填因子大于 0.6 时,具有显著的优势。

表 3 哈希表与数组顺序查找的平均查找时间比较

测试消息类型	消息字段名数目 N	装填因子 μ	链地址法	开放地址法—线性探测	开放地址法—二次探测和再哈希法	数组顺序查找	
			平均查找时间/s	平均查找时间/s	平均查找时间/s	平均查找时间/s	与链地址法时间比值
A01-入院/访问通知	188	0.188	0.640	0.638	1.0	0.650	1.0
A02-转移患者	280	0.280	0.938	1.034	1.1	1.008	1.1
A12-取消转移	302	0.302	1.028	1.542	1.5	1.439	1.4
A03-清除患者	408	0.408	1.400	2.090	1.5	2.110	1.5
A04-登记患者	523	0.523	1.801	3.710	2.1	2.968	1.6
A05-入院前通知	626	0.626	2.144	5.244	2.4	4.074	2.1
A08-更新患者信息	723	0.723	2.471	9.640	3.9	4.900	2.0
A19-患者查询	826	0.826	2.834	13.120	4.6	6.221	2.2

4 结语

为了在不同的医疗机构之间更好地实现医疗信息共享,本文提出了一种基于 HL7V2.x 协议标准的 HL7 消息构建与存储模型。仿真实验结果证明,基于 XML 映射文件的构建和存储模型能够避免大量重复的手工配置工作,基于哈希表的消息构建模型可以提高消息字段名的查找速度,快速完成消息构建。

另外,模型针对的 HL7 协议是 HL7V2.x 版本的,HL7V3 协议已经发布。两者的主要区别是 HL7V3 消息在结构上采用 XML 形式描述,语义上更加直观。不同医疗机构信息化水平不同,HL7V2.x 和 HL7V3 消息可能在很长时间内共存于医疗信息管理系统中,HL7V2.x 和 HL7V3 消息的相互转化成为后续的重点工作。深入研究 HL7V2.x 和 HL7V3 消息解析、存储、构建、两者之间的相互转换以及整个流程的效率问题具有比较重要的意义。

参考文献:

- [1] Health Level Seven International. Health level seven version 2.4 [EB/OL]. [2011-03-20]. <http://www.hl7.org/>.
- [2] 杨琳,李伟鹏,陆波. HL7 在 HIS 标准化建设中的应用[J]. 中国医学物理学杂志,2005,22(1):412-416.
- [3] 马雪英. 采用 HL7 设计管理信息系统中的通信模块[J]. 计算机应用,2003,23(1):76-78.
- [4] 张翠英,张尚军,徐晨勇,等. 医疗信息交换标准—HL7[J]. 生物医学工程学杂志,2001,18(1):111-115.
- [5] 杲耀景,冯大眷,王潜平. 基于 HL7 和 XML 技术实现 HIS 间信息交换与集成[J]. 计算机应用,2003,23(1):109-110.

- [6] 郭冰洁,韩旭,赵媛元. 基于 HL7 V2.x 协议的消息解析模型设计与应用分析[J]. 计算机工程与设计,2010,31(3):667-683.
- [7] LAFORE R. Java 数据结构和算法[M]. 计晓云,赵研,曾希,等译. 北京:中国电力出版社,2003.
- [8] 李兴华. Java Web 开发实战经典[M]. 北京:清华大学出版社,2010.
- [9] 萨师煊,王珊. 数据库系统概论[M]. 3 版. 北京:高等教育出版社,2000.
- [10] HUANG E W, LION D M. Performance analysis of a medical record exchanges model[J]. IEEE Transactions on Information Technology in Biomedicine,2007,11(2):153-160.
- [11] Health Level Seven International. HL7 Version 3 Guide [EB/OL]. [2011-03-26]. <http://www.hl7.org/>.
- [12] IHE International. Patient Identifier Cross-Reference (PIX) and Patient Demographic Query (PDQ) HL7 v3. 2009 [EB/OL]. [2011-03-05]. http://www.ihe.net/technical/frameworks/load/IHE_ITI_TF_SupplementPIX_PDQ_HL7v3_TL2009-08-10.pdf/.
- [13] Apache log4j. log4j [EB/OL]. [2010-12-05]. <http://logging.apache.org/log4j/index.html>.
- [14] Apache Tomcat. Tomcat [EB/OL]. [2010-12-05]. <http://tomcat.apache.org/>.
- [15] Health Level Seven International. V3 Messaging Standard, Health Level Seven (HL7) Standard [EB/OL]. [2011-04-05]. <http://www.hl7.org/implement/standards/v3messages.cfm/>.
- [16] W3C. Extensible Markup Language (XML) [EB/OL]. [2011-04-05]. <http://www.w3.org/XML/>.

(上接第 3417 页)

- [8] 寇连坡. AIS 船位信息产生误差的原因与对策[J]. 世界海运,2006,29(6):8-9.
- [9] PHELP D, HOUGH G. Digital image monitoring to optimise safe port operations [EB/OL]. [2011-02-15]. http://researchspace.csr.co.za/dspace/bitstream/10204/2536/1/Phelp_2008.pdf.
- [10] 于臣,柳晓鸣,黄上海,等. 基于电子海图 CCTV 智能控制系

- 统[C]//中国航海学会通信导航专业委员会 2006 年学术年会论文集. 大连:大连海事大学出版社,2006:93-98.
- [11] 赵森,许振宇,张秉权. 地理坐标与直角坐标相互转换的算法[J]. 火力与指挥控制,2001,26(4):64-66.
- [12] SINHA S N, POLLEFEYS M. Towards calibrating a pan-tilt-zoom camera network [EB/OL]. [2011-02-10]. <http://www.cs.unc.edu/~marc/pubs/SinhaOMNI04.pdf>.