

基于特征串匹配的 X 射线荧光光谱仪数据采集

吴 旭¹, 贡 华², 李汪根¹, 方 群^{1,3}

(1. 安徽师范大学数学计算机科学学院, 安徽 芜湖 241003; 2. 安徽海螺信息技术工程有限责任公司, 安徽 芜湖 241003;
3. 东南大学计算机科学与工程学院, 南京 210096)

摘 要: 针对水泥企业在异构环境下采集 X 射线荧光光谱仪数据的问题, 提出一种基于特征串匹配的数据采集算法, 在此基础上构建一个具有通用性的水泥质量管理系统, 用于自动实时地采集、存储和上传样品质量数据。应用结果证明, 该系统运行情况良好, 适应能力较强, 较大地提高了水泥企业质量管理和生产控制能力。

关键词: 特征串匹配; X 射线荧光光谱仪; 数据采集; 水泥质量管理系统

X-ray Fluorescence Spectrometer Data Collecting Based on Characteristic String Matching

WU Xu¹, GONG Hua², LI Wang-gen¹, FANG Qun^{1,3}

(1. School of Mathematics & Computer Science, Anhui Normal University, Wuhu 241003, China;
2. Anhui Conch Information Technology Engineering Co., Ltd., Wuhu 241003, China;
3. School of Computer Science & Engineering, Southeast University, Nanjing 210096, China)

【Abstract】 Oriented to X-ray fluorescence spectrometer data collecting in heterogeneous architecture of great cement factories, this paper proposes a data collecting algorithm based on characteristic string matching, and develops a general Cement Quality Management System(CQMS) to automatically collect, store and upload real-time quality data. Application result proves that the system is running steady and has robust adaptive ability. It can help cement firm to improve ability of quality management and product controlling.

【Key words】 characteristic string matching; X-ray fluorescence spectrometer; data collecting; Cement Quality Management System(CQMS)

DOI: 10.3969/j.issn.1000-3428.2012.02.084

1 概述

目前, 我国水泥企业主要采用荷兰的 PanaLytical、日本的 Rigaku 以及德国的 Bruker 等著名品牌的 X 射线荧光光谱仪(简称 X-荧光仪)完成水泥原材料、生料、熟料、水泥等物料的成分分析与检测, 它们的工作原理基本相同, 但是系统结构不同, 数据格式不统一, 对运行环境要求也不同, 给质量数据采集带来一定困难。国际上一些大型水泥企业如拉法基自 20 世纪 80 年代就已普遍采用在线质量检测, 但设备投入大且维护代价高, 不适合我国国情。国内水泥生产企业大多仍用传统手段对质量数据进行人工输入和手工统计分析, 远没有实现检验数据的自动采集和自动分析, 导致质量数据的实时性和准确度不高, 严重影响了水泥质量。本文根据我国国情, 在不增加硬件设备的前提下, 利用纯软件的数据采集 Agent, 提出一种基于特征字符串匹配的 X-荧光数据采集算法, 以实现数据的自动匹配、采集和上传。

2 问题描述

PanaLytical、Rigaku 和 Bruker 等荧光仪一般都提供对外数据接口, 接口方式主要有串口方式、TCP/IP 方式和文件传输方式 3 种^[1], 其中, 串口和 TCP/IP 方式需要详细了解各类荧光仪的数据格式和传输协议, 且由于通信同步困难, 复杂性相对较高, 因此本文选择文件传输方式获得荧光仪输出数据。文件传输方式首先需要配置传输站即文件传输的目的位置(如 D:\DATA), 然后选择需要传输的数据类别并与传输程序关联。当荧光仪生成样品检验数据后, 管理程序向传输站输出一个包含样品成分列表的文本文件, 如表 1 所示。

表 1 X-荧光仪传输数据样式

PanaLytical			Rigaku		
指标名称	数据	单位	类别	指标名称	数据
样品名	4RM12:00		2	K-14:00	A-1 2010-1-2614:28171
分析程序	RAW-2		3	SiO ₂	mass% 23.02
测量时间	2010-1-13 12:01:29		3	Al ₂ O ₃	mass% 4.97
初始重量	10.00	g	3	Fe ₂ O ₃	mass% 3.08
最终重量	10.00	g	3	CaO	mass% 65.41
Loss	35.34	%	3	MgO	mass% 2.29
SiO ₂	12.46	%	3	K ₂ O	mass% 1.04
Al ₂ O ₃	3.11	%	3	Na ₂ O	mass% 0.72
Fe ₂ O ₃	1.77	%	3	SO ₃	mass% 0.28
CaO	44.78	%	3	Sum	mass% 100.81
MgO	0.13	%	3	KH	0.871
K ₂ O	0.27	%	3	SM	2.86
Na ₂ O	0.02	%	3	IM	1.61
SO ₃	0.04	%	3	C ₃ S	mass% 53.58
Sum	97.89	%	3	C ₂ S	mass% 25.56
LSF	112.80		3	C ₃ A	mass% 7.95
KH	1.12		3	C ₄ AF	mass% 9.24
SM	2.56		3	液相量	25.89
IM	1.76				

基金项目: 国家“863”计划基金资助项目(2007A01Z422); 国家自然科学基金资助项目(61070060); 安徽高校省级自然科学基金资助项目(KJ2011A148)

作者简介: 吴 旭(1971—), 女, 实验师, 主研方向: ERP 系统, WSN; 贡 华, 高级工程师; 李汪根、方 群, 副教授、博士

收稿日期: 2011-05-18 **E-mail:** lqf121@mail.ahnu.edu.cn

3 基于特征串匹配的数据采集算法

3.1 样品指标超集 C

通过对 1 000 多个样品成分报告的对比分析, 并综合考虑所有可能的检测需求, 构造出一个完整的样品指标超集 C , 包含所有可能的样品信息和样品成分指标, 其内容如下: $C = \{\text{样品名称, 分析程序, 测量时间, 初始重量, 最终重量, 烧失量, 二氧化硅, 三氧化二铝, 三氧化铁, 氧化钙, 氧化镁, 三氧化硫, 氧化钾, 氧化钠, 氯离子, 合计, 液向量, 饱和比, 硅酸率, 铝氧率, 石灰饱和率, 硅酸三钙, 硅酸二钙, 铝酸三钙, 铁铝酸四钙}\}$ 。这个超集包含了目前水泥行业中最常用到的指标类型, 必要时可进一步扩展。

3.2 样品报告结构

样品报告是一个包含某个样品信息和检测指标的字符串, 可将样品报告 R 分解为若干个指标行 L , 每个指标行 L 的结构可以抽象为: $\{\text{填充信息, 指标名, 填充信息, 指标值, 填充信息}\}$ 。用模式语言描述通用样品报告的数据结构为:

R	L	N	M
-----	-----	-----	-----

其中, R 为样品报告, 包含若干条样品指标行, $R = L^*$; L 为指标行, 包含一条样品指标信息, $L = \omega^* N \omega^* M \omega^*$, ω 为填充字符; N 为样品指标名, 表示指标种类, 常以英文字符开始, $N = (a..z, 0..9)^* \#(0..9)^*$; M 为样品指标值, 表示成分含量的实数, 多以%为单位, $M = (0..9)^*(0..9)^*$ 。

根据特征字符串匹配的方称为串模式。考虑到部分类型的荧光仪中有些重要指标并没有显著的特征表示(如样品名称或检测时间等), 因此, 在样品指标名 N 中设置行号特征“# n ”, n 表示特殊指标所在行号, 用于提高匹配算法的通用性, 这种方案称为行模式。2 种模式综合使用可以解决大部分的匹配问题^[2]。

3.3 样品指标特征库

根据样品指标超集和样品指标行结构构建样品指标特征库 S , 用于辅助样品指标匹配算法的执行, 指标特征库 S 结构如下: $S = \{mname, sname, spos, epos, remark\}$, 其中, $mname$ 表示特征串; $sname$ 表示指标名称; $spos$ 表示在指标行中的起始位置; $epos$ 表示结束位置; $remark$ 表示说明信息。

例如, Panalytical 荧光仪的数据“烧失量”(Loss)如下:

指标名称	数据	单位
Loss	35.34	%

则在特征库中与其对应的匹配模式为: (LOSS, Loss, 15, 30, 烧失量)。

上述匹配模式说明烧失量指标应当从含有“LOSS”的行采集, 其指标数据位置在 15 和 30 之间, 因此, 匹配程序先在目标行中搜索特征标识“LOSS”, 匹配后, 再从指定区间提取指标值即 35.34%, 从而获得烧失量指标(Loss, 35.34)。特征库中的初始值可根据不同的荧光仪类型预先设定, 在系统部署后还可以根据实际情况进行添加和修改。

3.4 数据采集算法

数据采集是从荧光仪输出的样品检验报告中提取指定的指标串然后写入本地数据库的过程, 需要反复使用串匹配算法^[3-4], 此处的串匹配是一种精确的匹配, 特别是当指标特征串间存在前缀关系时, 更应当采取最大匹配策略, 以避免采集错误的数据库。样品数据采集程序的输入是一个样品报告字符串 R , 匹配算法的处理对象是一个样品指标行 L , 用类 C 语言表示的算法如下:

算法 1 样品指标行匹配算法

```
String[2] MatchByLine(int ln, String L)
{//输入 待分析的样品指标行字符串 L 及对应行号 ln
//输出 返回<指标名, 指标值>
(1)cdb=OpenCharacterDB(); //打开样品指标特征库, 获得句柄 cdb
(2)for (i=0;i<length(cdb);i++){ //按特征库中逐行进行比对
(3)if islineMode(cdb[i].mName) //行模式匹配成功
(4) return(cdb[i].sName,substr(L,cdb[i].spos,cdb[i].epos));
(5)else if (MetricStrMatch(cdb[i].mName,L)) //串模式匹配成功
(6) return(cdb[i].sName,substr(L,cdb[i].spos,cdb[i].epos));
(7);}
```

算法 2 样品报告分析算法

```
String[] MatchByReport (String R)
{//输入 待分析的样品检验报告文档 R
//输出 返回{<指标名, 指标值>}
(1)open(R,'r'); //打开指定文档 R(一般是一文本文档)
(2)rpt = resoluteDoc(R); //获取文档 R 中所有行
(3)for (i=0;i<length(rpt);i++)
(4)AddReport(mtrc,matchByLine(i+1,rpt[i])); //解析指定行并加入指标集 mtrc
(5)return mtrc;
(6);}
```

分析算法 1 可以看出, 为了获得每一行的匹配结果, 必须对整个特征串库进行顺序扫描, 假设特征库的规模为 m , 则将耗费 $O(m)$ 的时间。分析算法 2 可知, 对于一个有 n 行的样品文档来说, 需要进行 n 次行匹配, 因此, 分析一个样品报告的时间复杂度将为 $O(mn)$ 。考虑到一般情况, m 和 n 均为一常数(通常不超过 20), 由此可以确定本文算法的时间复杂度近似于 $O(1)$, 即在常数时间内完成一个样品数据的采集。与同类算法^[5-6]相比, 本文算法简单灵活, 适用于荧光仪数据简单但存在差异性的场合。

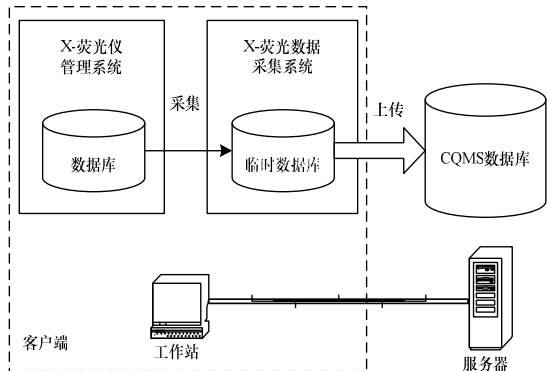
4 X-荧光仪数据采集系统设计与实现

荧光仪数据采集程序是水泥质量信息管理系统(Cement Quality Management System, CQMS)的基本组成部分, 它将荧光仪数据采集并上传至 CQMS 数据库服务器, 结合从其他位置采集到的自动压力机数据、生产控制数据、销售发货系统数据等, 为整个质量管理提供了一个数据量大、更新迅速、关联性好的数据平台, 也为质量管理专家系统提供了帮助。

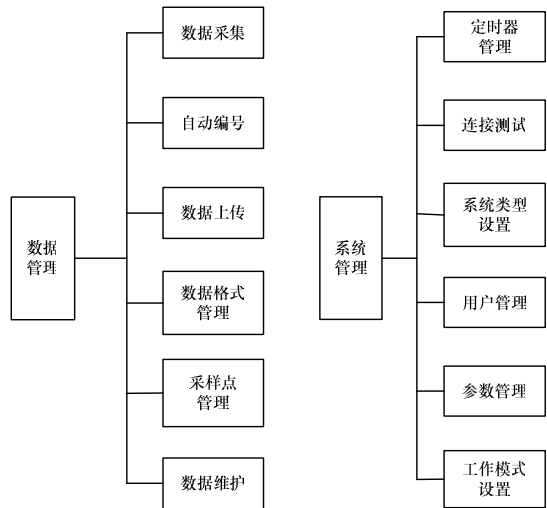
通常, X-荧光仪都配备专用的管理 PC, 不同类型的荧光仪与其管理 PC 的通信方式有所差异, 如 Panalytical 型的荧光仪借助安装在管理 PC 中的控制接口卡, 通过并行数据端口进行连接, Rigaku 型的荧光仪则通过网络接口卡与管理 PC 连接。为了保证荧光仪的正常运转, 可以将传输站设置在管理 PC 的硬盘上, 然后将数据采集 Agent 部署在其他位置的 PC 或者服务器中, 系统运行时, 数据采集 Agent 定时地(通常间隔为 1 min~60 min)从管理 PC 传输站中读取样品检验文档, 取回到宿主 PC 中进行分析处理, 这样就不会影响到荧光仪本身的控制。

基于第 3 节的特征串匹配算法, 结合公司质量管理需求, 构建的 CQMS 如图 1 所示。如图 1(a)所示, X-荧光仪数据采集系统从荧光仪管理 PC 中定时读取检验数据并上传至 CQMS 数据库。图 1(b)显示了采集系统的内部结构, 大致包括系统管理和数据管理两大部分。系统管理主要负责用户管理、系统参数管理、系统类型设置、工作模式设置以及定时器管理等, 管理员用户可以通过系统管理针对不同的荧光仪

类型、检验指标集、采集频率等进行设置，以配置满足子公司需要的工作环境；数据管理提供给普通用户进行数据采集、管理和维护的良好工具，正常运行时系统处于自动值守状态，基本上不需要人为的干预。



(a)系统工作原理



(b)系统功能结构

图1 X-荧光仪数据采集系统原理及结构

除数据采集模块外，数据上传也是整个系统的关键组

编辑 张帆

(上接第 249 页)

本文处理器在输入随机数列的算法仿真下，输入数据量化为 10 bit、输出为 13 bit，SQNR 达到 47.2 dB，在 SMIC 0.13 μm 1.2 V 电压供应的 1P8M 工艺下综合，最高时钟频率可达 125 MHz，面积达到 1.800×1.500 μm²。

4 结束语

本文设计了一种 FFT/IFFT 处理器，适用于 IEEE802.16e 标准的 MIMO-OFDM 系统，具有高效、可配置的优点。IEEE802.16e 标准中的子载波间隔为 10.94 KHz，本文设计的处理器在相应的工作频率下可满足系统需求。下一步将扩展处理器的处理点数，使之适用于更多的 MIMO-OFDM 系统。

参考文献

[1] Cooley J W, Tukey J W. An Algorithm for the Machine Calculation of Complex Fourier Series[J]. Mathematics of Computation, 1965, 5(5): 87-109.
 [2] Lin Yuwei, Lee Chenyi. Design of an FFT/IFFT Processor for

件，可以采取调用 SQL Server 存储过程的方法，将原始数据上传至 CQMS 数据库。使用存储过程的好处是可以根据子公司情况进行远程修改和维护，而不必到现场更新采集 Agent 程序，省去了大量的时间和费用，从而大幅降低了维护成本。

5 结束语

X-荧光仪数据采集是水泥企业和相关行业在信息化进程中必须要实现的基础任务，本文提出的基于特征串的荧光仪数据采集算法实现了数据采集的自动化、系统化，目前已经部署在 50 多个节点上，通过近两年的运行，X-荧光仪数据采集系统运行情况良好，适应能力较强，明显提高了水泥企业的质量管理和生产控制能力，为提高信息化水平、构建 ERP 系统提供了支持。

参考文献

[1] Prasad R, Agarwal S. Parameterized Shift and String Matching Algorithm Using Super Alphabet[C]//Proc. of International Conference on Computer and Communication Engineering. Kuala Lumpur, Malaysia: [s. n.], 2008.
 [2] Murty V S, Reghu R P C, Raman S. Design of a Language-independent Parallel String Matching Unit for NLP[C]//Proc. of IEEE International Workshop on Computer Architectures for Machine Perception. [S. l.]: IEEE Press, 2003.
 [3] Martinek T, Kofenek J, Fucik O, et al. A Flexible Technique for the Automatic Design of Approximate String Matching Architectures[C]//Proc. of Design and Diagnostics of Electronic Circuits and Systems. [S. l.]: IEEE Press, 2006.
 [4] 王静帆, 郭晓钧, 夏云庆. 中文信息检索系统的模糊匹配算法研究和实现[J]. 中文信息学报, 2007, 21(6): 59-64.
 [5] 刘萍, 郭莉, 方滨兴, 等. 串匹配算法中模式串与文本之间关系的研究[J]. 软件学报, 2010, 21(7): 1503-1514.
 [6] 马莉, 唐素勤, 陈立娜, 等. 改进的基于句模匹配算法的问句理解方法[J]. 计算机工程, 2009, 35(20): 50-52.

MIMO OFDM Systems[J]. IEEE Trans. on Circuits and Systems-I: Regular Papers, 2007, 54(4): 807-815.

[3] Sansaloni T, Perez-Pascual A, Torres V, et al. Efficient Pipeline FFT Processor for WLAN-OFDM Systems[J]. Electronics Letters, 2005, 41(19): 1043-1044.
 [4] Li Simeng, Chen Yun, Zeng Xiaoyang, et al. Paris, Circuits and Systems[C]//Proc. of ISCAS'10. Paris, France: [s. n.], 2010.
 [5] Schwoerer L, Zielinski E. Optimized FFT Architecture for MIMO Applications[C]//Proc. of EUSIPCO'05. Antalya, Turkey: [s. n.], 2005.
 [6] Yang Gi-Jung, Jung Yun-Ho. Design of Low-complexity FFT Processor for MIMO-OFDM Based SDR Systems[C]//Proc. of ICACT'10. Gangwon-Do, South Korea: [s. n.], 2010.
 [7] 汪文义, 王琳凯, 周金元, 等. 改进的多路基-24 FFT 处理器设计[J]. 计算机工程, 2011, 37(7): 262-264.

编辑 顾姣健