

DOI:10.19751/j.cnki.61-1149/p.2021.02.017

# 数据自动采集在煤层气生产中的应用

## ——以沁水盆地柿庄南区块煤层气井为例

陈慧<sup>1</sup>, 郭晖<sup>2</sup>

(1. 北京中煤矿山工程有限公司 北京 100013; 2. 中联煤层气有限责任公司 北京 100015)

**摘要:** 为了提高煤层气生产数据采集效率, 笔者建立了数据自动采集系统。该系统可以自动识别、采集并处理不同格式、不同结构和不同类别的生产数据, 形成具有统一结构的数据格式。系统可分别针对 Word 文件和 Excel 文件查找表格标题和分析表格结构, 并利用整表识别或者关键字识别对单个单元格、连续多个单元格、单元格内的键值对和按行/列增长的数据表进行表格识别, 从而实现数据自动采集。柿庄南煤层气生产数据自动采集结果表明, 该系统可以准确和高效地提取和采集生产数据, 为提高煤层气田生效效率提供有效技术支持。

**关键词:** 数据自动采集系统; 查找表格标题; 分析表格结构; 提取表格数据; 煤层气

中图分类号: TD712

文献标志码: A

文章编号: 1009-6248(2021)02-0203-09

### Application of Automatic Data Acquisition System in Coalbed Methane Production: A Case Study of Coalbed Methane Wells of South Shizhuang Block in Qinshui Basin

CHEN Hui<sup>1</sup>, GUO Hui<sup>2</sup>

(1. Beijing China Coal Mine Engineering Co., Ltd., Beijing 100013, China;

2. China United Coalbed Methane Company, Beijing 100015, China)

**Abstract:** The automatic data acquisition system identifies, collects and processes production data automatically in different formats, structures and categories, which helps improve the data acquisition efficiency of coalbed methane production. The system search table title and analyze table structure of Word file and Excel file, and use table recognition or keyword recognition to identify single cell, continuous multiple cells, key value pairs and data table. Results from coalbed methane wells of south Shizhuang block show that the system can extract and collect production data accurately and efficiently, providing an effective technical approach for the coalbed methane fields.

**Keywords:** automatic data acquisition system; searching table title; analyze table structure; extract table data; coalbed methane

中国煤层气资源丰富, 总资源量约为 36.8 万亿 m<sup>3</sup>, 居世界前列(白振瑞等, 2015; 朱庆忠等, 2017)。目前, 中国煤层气生产也进入了快速发展阶段, 并在

钻完井、产量计量、天然气集输、气水处理等环节产生了大量数据(秦勇等, 2015; 宋岩等, 2012)。因此, 高效、准确的采集和分析数据对于保障煤层气生产计划

收稿日期: 2020-05-27; 修回日期: 2021-01-12

基金项目: 中海石油(中国)有限公司信息化建设项目“中联公司生产操控中心试点及配套基础设施改造”(中海油科信【2020】426 号)。

作者简介: 陈慧(1982-), 女, 硕士, 助理研究员, 主要从事地面预注浆及信息化管理工作。E-mail: 272838115@qq.com。

顺利实施、提高煤层气产量具有重要的意义。

近年来,大数据技术在中国能源企业已广泛应用,其在油田生产中的应用为实现油气智能化、自动化生产提供了有效支撑(张仙伟等,2020;高伟华,2018;刘宏伟,2010)。运用大数据技术对煤层气生产数据进行多维度分析也是煤层气企业降本增效的有效途径。其中,一线生产数据快速、高效的提取和采集是保障大数据技术在煤层气生产过程中顺利应用的重要前提。目前煤层气生产数据采集多沿用传统的工作模式,主要存在以下几个方面的问题:①生产一线缺少标准数据库和专业数据管理软件,大部分数据以静态报表分散存储,数据管理效率较低,给后期数据统计和应用带来不便。②生产一线数据以静态报表的方式以电子邮件的形式传送,依靠人工收集数据的工作步骤繁琐,且不能充分发挥网络资源的作用。③数据利用效率较低,各类数据须通过人工方式录入 Excel 表格中,降低了工作效率并容易出现错误。这种传统的生产数据采集模式制约了数据筛选、数据分析及结果可视化及验证,从而严重影响了煤层气生产运行的稳定性及生产决策的准确性。因此,笔者针对煤层气生产数据建立了一套自动数据采集系统,实现煤层气生产全过程数据准确、快速提取,以期为后续数据分析和应用奠定良好基础,提高中国煤层气开采效率。

## 1 数据自动采集系统

数据自动采集系统是将煤层气生产作业区在生产过程中产生的不同类别、不同格式和不同结构的生产数据自动识别、采集并处理成具有统一结构的数据格式。数据自动采集系统主要包括以下3个子系统:查找表格标题、分析表格结构、提取表格数据(图1)。该系统可从邮件中自动识别并判断文件类型,提取表格标题,分析表格结构,根据指定关键字或表格形式进行识别提取数据。

## 2 数据采集步骤

### 2.1 查找表格标题

数据自动提取技术主要针对2种常用的数据载体文件:一种是 Word 文件;另一种是 Excel 文件。针对 Word 文件,首先获取文档所有顶层节点(Node)并过滤出所有“段落(Paragraph)”、“表格

(Table)”节点。对于每一个表格节点,获取其节点下标,找出段落节点中下标分别大于、小于表格下标并最为接近的2个段落节点。任何一个段落节点中的文字如果以“表”开头,则将其作为表格名称。若均符合要求或均不符合要求,则以先出现的段落节点的文字作为表格标题。针对 Excel 文件,则直接使用工作表(Worksheet)名称作为表格标题。

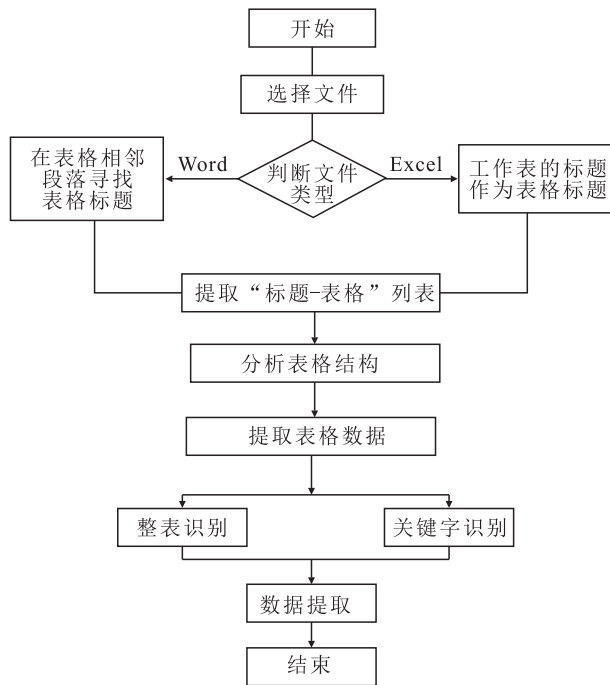


图1 数据自动采集系统结构示意图

Fig. 1 Structure of automatic data acquisition system

### 2.2 分析表格结构

#### 2.2.1 Word 文件

Word 中表格的存储结构与 HTML 类似,主要采用“先行后列”的树状结构。对于常见的列合并单元格,实际存储时并没有存储列合并信息,所以直接读取表格会出现“错位”的情况。分析 Word 表格内容需要将原始的树状结构标准化为二维结构。由于 Word 表格在每个单元格内存有列宽信息(图2a),所以需要遍历每一行的每个单元格,对列宽进行累加,以获取单元格的 X 坐标。在遍历完成所有单元格后,对 X 坐标进行去重和排序,生成一个有序数列  $X_i(i=0,1,2,3,\dots)$ ,其中  $i$  即为所对应的原始单元格在二维结构表格中的列号(即列下标)(图2b)。此步骤同时可以整理出所有单元格横向合并的列跨距,并使用“{←}”标识被横向合并的单元格(图2c、图

2d)。对于存在纵向合并的单元格,被合并的单元格可以读出“CellMerge. Previous”标记,所以纵向合并处

理比较简单,直接处理标记即可(使用“{↑}”进行标识),此步骤兼容同时被横纵合并的单元格。

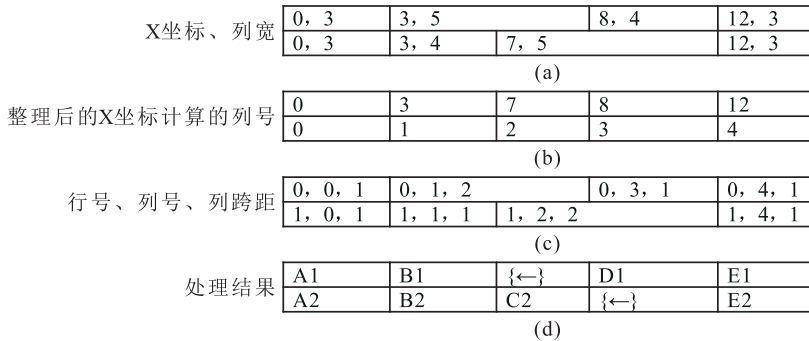


图 2 Word 中横向合并单元格结构分析图

Fig. 2 Structural analysis of horizontal merged cells in Word

### 2.2.2 Excel 文件

针对 Excel 表格,可以读取到每一个非空单元格(图 3a),通过单元格的行列号可以将数据解析至对应位置(A1、B1、B2……E2 代表数据)。解析单元格的“是否被合并(IsMerged)”标记可以获取到当前单元格的合并状态,若为被合并的单元格,可以通过“获取合并区间(GetMergedRange)”找到其合并主体(图 3b 中 A1、B1 和 C2 为合并主体)。若与合并主体处在同一列,则当前单元格视为向上合并(使用“{↑}”进行标识),否则视为向左合并(使用“{←}”进行标识)(图 3c)。

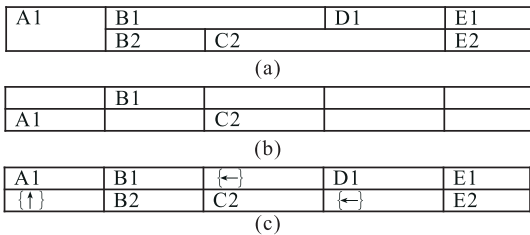
要有 2 种:①为整表导出。②为模板匹配。整表导出使用预设的正则表达式列表,当表格标题与正则表达式列表中的任何一项匹配成功时,将当前表格输出为一个工作表,添加(若文件已存在则覆盖)输出目标的 Excel 文件中,并生成修改记录。模板匹配主要根据用户定义模板中所包含的正则表达式遍历单元格,寻找每一个簇与之相匹配的模板。模板匹配主要对表格进行识别。该系统主要有 2 种表格识别的方式:①为整表识别。②为关键字识别。笔者重点介绍模板匹配的相关识别算法。

#### 2.3.1 整表识别

整表识别时,先判断被识别的表格是否属于标准的二维表结构(分析后的表格不存在任何合并的结构)(李目海,1999)。若不是二维表结构,终止整表识别,忽略所有整表识别的模板,使用关键字识别。若当前表格属于二维表结构(图 4a),则根据定义的模板,在第一行(列)中进行匹配(A1、B1、C1、D1)。匹配完成后若存在有效的匹配结果(A1、B1、D1),则从第二行(列)开始顺次向下(右)遍历每一个行(列)(A2、B2、D2……A5、B5、D5)。提取出匹配到的标题所在列(行)的单元格内的数据,最终形成一个一维的标题列表和一个二维(外层为增长方向,内层为标题对应的数据项)的数据集列表(图 4b),结束识别,输出数据表格。

#### 2.3.2 关键字识别

关键字识别是假设标题簇与数据区成对出现且数据区紧贴于标题簇的右侧(下方),纵(横)向的起止位置相同。该识别方法可进一步根据数据区域划



a. 原始表格;b. 合并主体;c. 处理结果

图 3 Excel 单元格结构分析图

Fig. 3 Structural analysis of cells in excel

### 2.3 提取表格数据

表格结构分析完成之后,Word 文件与 Excel 文件的表格内容已经展开并统一为同样类型的二维表结构。在此基础上可以进行数据提取。提取数据使用合并后的单元格(以下简称“簇”)作为逻辑判断单位,以匹配表格的实际结构,由于簇与簇之间可能不完全对齐,所以同时要记录与相邻簇之间的关系以便后续判断。笔者设计的常用提取表格数据方法主

分为(王瑞通等,2017):单个单元格(Single)、连续多个单元格(Array)、单元格内的键值对(KeyValuePairs)、按行/列增长的数据表(DataTable)。

关键字识别首先根据一系列识别规则查找数据区域标题,按识别结果的位置拓扑关系构建数据识别模型,对数据区域周边单元格的形状排除冲突与歧义,对模型中的数据区域进行筛选匹配,并最终完成识别。

(1)单个单元格(Single)。获取所有与模板匹配成功的标题簇,对于每一个标题簇:获取4条边界

的位置、右侧簇、下方簇(表格边界在判断上等效于标题簇);右侧簇或下方簇的跨度均与本簇相同的情况下,若均为标题簇或均为非标题簇(图5a、图5b),则标记当前簇为无效簇,跳过当前簇,否则将并不是标题簇的作为数据簇(有效数据区)(图5c)。若右侧簇或下方簇的跨度与当前簇跨度相同,且右侧簇或下方簇不是标题簇,则标记为数据簇(有效数据区)(图5d),否则标记当前簇为无效簇,跳过当前簇(图5e)。若右侧簇或下方簇的跨度与当前簇跨度均不同,则标记当前簇为无效簇,跳过当前簇(图5f)。

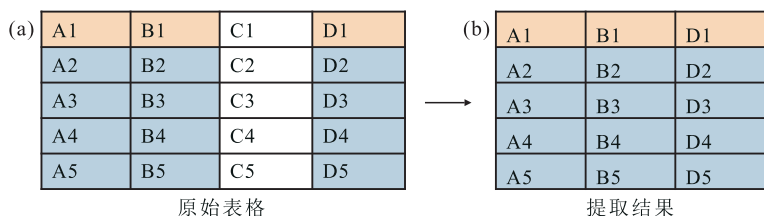


图4 整表识别示意图

Fig. 4 Schematic diagram of whole table identification

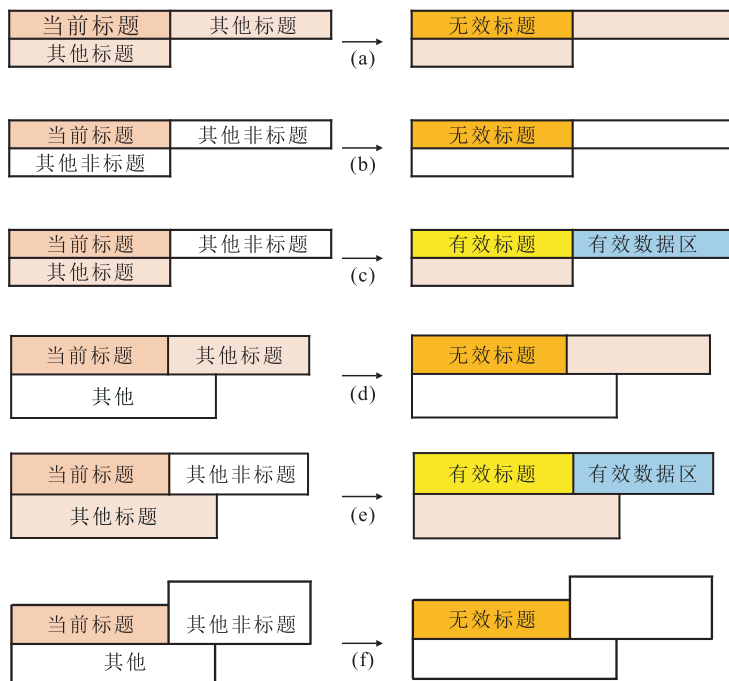


图5 具有跨度相同的单个单元格识别示意图

Fig. 5 Schematic diagram of identifying single cells with the same span

(2)连续多个单元格(Array)。连续多个单元格是指一个合并的标题单元格对应多个数据单元格的结构。若标题单元格为列合并,则数据区位于下方单元格;若为行合并,则数据区位于右侧单元格;

若同时存在行列合并,且除了标题单元格和其他单元格以外,其他方向存在非标题单元格,则提取非标题单元格为数据(图6a、图6b)。若右侧或者下方都是标题单元格或者其他单元格,则当前单元格为无

效单元格(图 6c、图 6d)。若不存在合并,则按照单个单元格的规则处理。

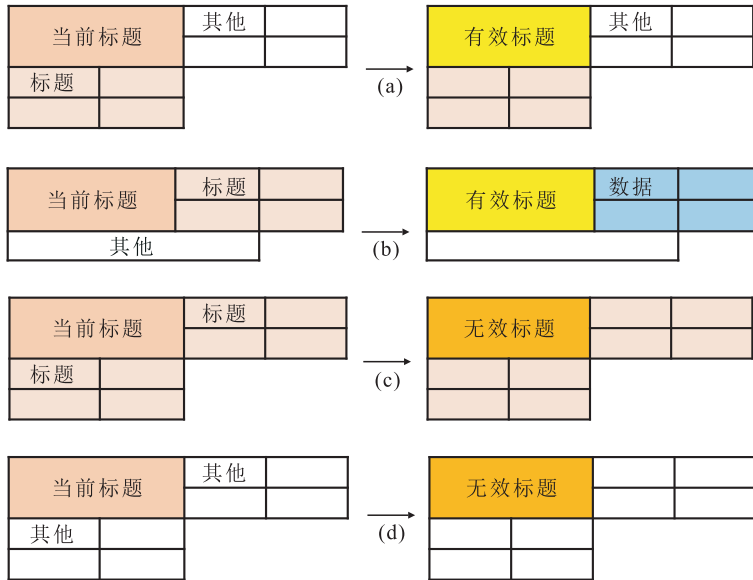


图 6 不同类型的连续多个单元格识别示意图

Fig. 6 Schematic diagram of identifying different types of consecutive cells

(3) 单元格内的键值对(KeyValuePairs)。单元格内的键值对是指单个单元格内既有标题又有数据的数据结构。不向外查找数据单元格,直接提取整个单元格内的文字,具体提取内容按照定义的提取模式进行处理。

(4) 按行/列增长的数据表(DataTable)。标题单元格合并,数据区位置判定规则与“连续多个单元格”相同。数据区形状为二维表格,向下增长,第一行为数据区的列标题。将查找指定方向上的所有单元格,直到出现合并方式不同的单元格或到达表格尽头。该类数据表可分为 2 种:①为标题单元格为行合并(图 7a)。②为标题单元格为列合并(图 7b)。前者数据区在标题单元格的右侧,所以“向右”寻找“相同行跨距”的“列”,后者数据区在标题单元格的下方,所以“向下”寻找“相同列跨距”的“行”。

按列增长的数据表与按行增长的数据表相似(数据区行列关系互换),数据区的查找方式也相同,但选择第一列判定数据区内表头。此外,按列增长的数据区,在输出时需要做行列转置,统一为按行增长相同的数据格式。

### 2.3.3 数据提取

每一个识别模板中包含多个提取项,用于提

取坐标信息,从识别出的数据区域内提取需要的数据,并按实际需要映射到数据库的字段上,用以直接生成数据表条目。每个提取项有 2 个正则表达式,提取过程中先使用“识别匹配表达式”抽提单个字段,再使用“解析匹配表达式”抽提字段数据内容。

## 3 数据自动采集在柿庄南煤层气生产中的应用

### 3.1 煤层气生产概况

柿庄南区块处在沁水盆地南部,煤炭储量巨大,具有煤变质程度高、含气量高、煤层埋藏深度适中等特点,为煤层气规模化开发提供了优良的条件(伊永祥等,2019;张松航等,2015;康清清等,2020;宿虎等,2020)。柿庄南区块煤层气井整体投产时间较长,至 2019 年底,该区块已建煤层气井 1 300 余口。据统计,柿庄南区块 5 个产区中每天可产生钻井日报、煤层气井生产日报数据多达 2 300 多份,这些数据信息覆盖面广,数据量也十分庞大,涉及钻井、完井、测试、稳压、采气、运输、储存等多个方面。因此,建立数据自动采集系统对于加快煤层气田智能化发展、降低生产成本具有重要意义。

A	B1	C1	D1	E1		
	B2	C2	D2	E2		
	B3	C3	D3	E3		
	B4	C4	D4	E4		
	B5	C5	D5	E5		
	B6	C6	D6	E6		

(a)

A						
B1	C1	D1	E1	F1	G1	H1
B2	C2	D2	E2	F2	G2	H2
B3	C3	D3	E3	F3	G3	H3
B4	C4	D4	E4	F4	G4	H4

(b)

图7 按行增长的数据表示意图

Fig. 7 Schematic diagram of data table growing by row

### 3.2 数据自动采集在生产中的应用

#### 3.2.1 单个单元格

本示例中, Word 表格所有要提取的字段均为标题、值成对出现的单元格(图 8a), 所以使用“单个单元格”匹配模板, 多数情况下, 也可以用“连续多个单元格”代替。表格结构较为繁杂, 出现了大量列合并情况(图 8b)。其中, “一开日期”、“二开日期”等标题簇的右侧均存在其他标题簇, 所以数据区被认为在下方; “完钻依据”右侧和下方均无标题簇, 在第一轮识别中无法确定数据区位置, 在第二轮识别中由于下方簇已被标记为数据区, 所以将右侧区域标记为数据区。在表格结构识别的基础上根据匹配模板中包含正则表达式(标题模式和内容模式)进行标题识别和抽提数据并得到匹配结果(图 8c、图 8d)。最终完成单个单元格的数据提取(图 8e)。

#### 3.2.2 连续多个单元格

Word 单元格中坐标数据紧贴在标题簇一侧、总跨距与标题簇一致(图 9a)。标题单元格纵向合并可直接使用“{↑}”进行标识(图 9b)。在此基础上使用“连续多个单元格”匹配模板中的正则表达式匹配数据(图 9c、图 9d)。输出结果将二维结构转化为文本结构, 使用“||”分隔每一行, 使用“|”分隔每一列(图 9e)。这样做的目的是为了将所有数据的格式均一化, 方便后续的数据提取。当数据簇的外

侧有其他不同合并数的簇时, 也可以使用“按行增长的数据表”模板得到相同的结果。

#### 3.2.3 单元格内键值对

本示例中, 需要提取的数据为 Word 单元格内部文本中的一部分(图 10a)。“单元格内的键值对”不向外查找数据区, 标题单元格自身即为数据区。在得到表格结构之后(图 10b), 将根据匹配模式进行数据匹配(图 10c、图 10d)。该提取模式是从当前单元格的文本中找到以“固井:”开头、“。”结尾的字符串, 然后提取其“:”和“。”之间的内容。利用该模板可提取“固井”和“山西 11 固井队”(图 10e)

#### 3.2.4 按行增长的数据表

该数据表的坐标信息为“一个标题格+二维数据区”式的组合型数据(图 11a), 所以在表格结构分析的基础上(图 11b)使用“按行增长数据表(Data Table)”模板进行匹配(图 11c、图 11d)。由于每个标题簇的下方均存在标题簇, 所以判定数据区位于右侧。数据区的判定从标题簇右侧第一列开始, 直到行数发生变化或出现其他标题簇、表格边界。所以 2 组数据最终均囊括了所有的 3 行 2 列单元格簇。输出结果将二维结构化为文本结构, 使用“||”分隔每一行, 使用“|”分隔行内列数据(图 11e)。这样做的目的是为了将所有数据的格式均一化, 方便后续的数据提取。

一开日期	二开日期	二开中完日期	三开日期	完钻日期
2009. 10. 31	2009. 11. 2		2009. 11. 9	2009. 12. 13
完井日期	钻井周期	建井周期	中完停待周期 (d)	完钻层位
2009. 12. 14	43. 54	58. 08		P1s
完钻依据	因煤层垮塌埋钻事故完井			

(a)

一开日期	{←}	二开日期	{←}	二开中完日期	{←}	三开日期	{←}	完钻日期	{←}
2009. 10. 31	{←}	2009. 11. 2	{←}		{←}	2009. 11. 9	{←}	2009. 12. 13	{←}
完井日期	{←}	钻井周期	{←}	建井周期	{←}	中完停待周期 (d)	{←}	完钻层位	{←}
2009. 12. 14	{←}	43. 54	{←}	58. 08	{←}		{←}	P1s	{←}
完钻依据	{←}	因煤层垮塌埋钻事故完井				{←}	{←}	{←}	{←}

(b)

名称	标题模式	项目格式	内容类型	内容模式
一开日期	^一开日期	Array	System. String	. *@ *
二开日期	^二开日期	Array	System. String	. *@ *
二开中完日期	^二开中完日期	Array	System. String	. *@ *
三开日期	^三开日期	Array	System. String	. *@ *
完钻日期	^完钻日期	Array	System. String	. *@ *
完井日期	^完井日期	Array	System. String	. *@ *
钻井周期	^钻井周期	Array	System. String	. *@ *
建井周期	^建井周期	Array	System. String	. *@ *
中完停待周期(d)	^中完停待周期(d)	Array	System. String	. *@ *
完钻层位	^完钻层位	Array	System. String	. *@ *
完钻依据	^完钻依据	Array	System. String	. *@ *

(c)

一开日期	{←}	二开日期	{←}	二开中完日期	{←}	三开日期	{←}	完钻日期	{←}
2009. 10. 31	{←}	2009. 11. 2	{←}		{←}	2009. 11. 9	{←}	2009. 12. 13	{←}
完井日期	{←}	钻井周期	{←}	建井周期	{←}	中完停待周期(d)	{←}	完钻层位	{←}
2009. 12. 14	{←}	43. 54	{←}	58. 08	{←}		{←}	P1s	{←}
完钻依据	{←}	因煤层垮塌埋钻事故完井				{←}	{←}	{←}	{←}

(d)

一开日期	二开日期	二开中完日期	三开日期	完钻日期	完井日期	钻井周期	建井周期	中完停待周期(d)	完钻层位	完钻依据
2009. 10. 31	2009. 11. 2		2009. 11. 9	2009. 12. 13	2009. 12. 14	43. 54	58. 08		P1s	因煤层垮塌埋钻事故完井

(e)

a. 原始文档; b. 分析表格结构; c. 匹配模板; d. 匹配结果; e. 识别结果

图 8 单个单元格数据提取示例图

Fig. 8 Example of single cell data extraction

坐标	X: 4992959.550	坐标	X: 4992959.550	{←}	{←}
	Y: 29657261.003	{↑}	Y: 29657261.003		
	H: 2093.874m		H: 2093.874m		

(a)

(b)

名称	标题模式	项目格式	内容类型	内容模式
坐标	^坐标	Array	System.Decimal, System.Decimal	X.*d+^\\d+@\\d+\\.\\d+   Y.*d+^\\d+@\\d+\\.\\d+

(c)

坐标	X: 4992959.550	{←}	{←}
{↑}	Y: 29657261.003	{←}	{←}
{↑}	H: 2093.874m	{←}	{←}

(d)

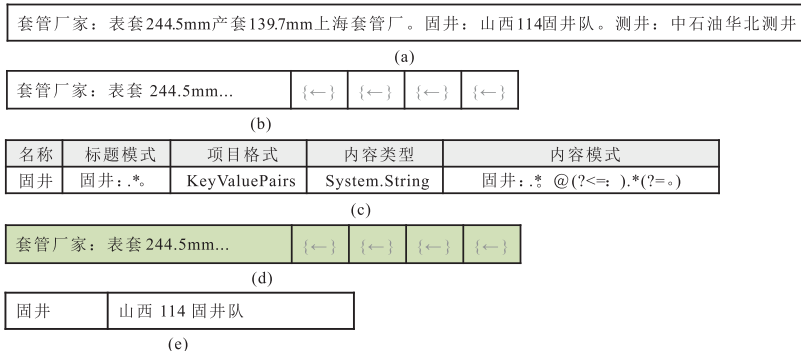
坐标	X: 4992959.550    Y: 29657261.003    H: 2093.874m
----	---

(e)

a. 原始文档; b. 分析表格结构; c. 匹配模板; d. 匹配结果; e. 识别结果

图 9 连续多个单元格数据提取示例图

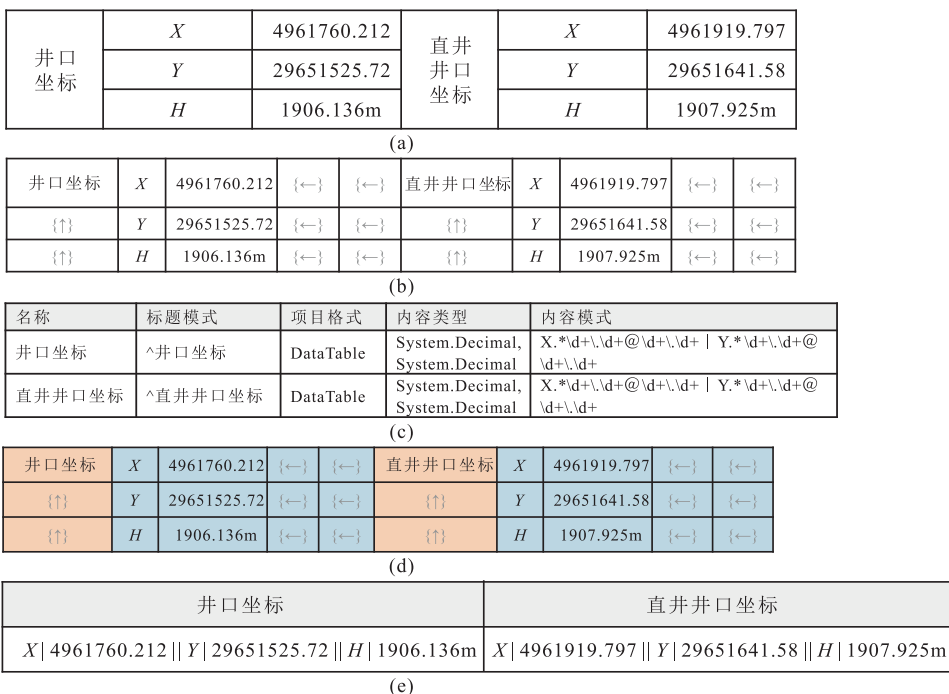
Fig. 9 Example of data extraction from consecutive cells



a. 原始文档; b. 分析表格结构; c. 匹配模板; d. 匹配结果; e. 识别结果

图 10 单元格内键值对数据提取示例图

Fig. 10 Example of data extraction of key value pairs in cells



a. 原始文档; b. 分析表格结构; c. 匹配模板; d. 匹配结果; e. 识别结果

图 11 按行增长的数据表数据提取示例图

Fig. 11 Example of data extraction from data table of row growth

## 4 结论

(1)数据自动采集系统可以将不同格式、不同结构和不同类别的生产数据自动识别、采集并处理成具有统一结构的数据格式。该系统主要包含“查找表格标题、分析表格结构、提取表格数据”3个关键模块。

(2)数据自动采集系统可分别针对 Word 文件

和 Excel 文件查找表格标题和分析表格结构,在此基础上利用整表识别或者关键字识别对单个单元格、连续多个单元格、单元格内的键值对、按行/列增长的数据表进行表格识别,实现数据自动采集。

(3)柿庄南煤层气生产数据自动采集应用表明,该系统可以针对单个单元格、连续多个单元格、单元格内的键值对和按行/列增长的数据表等进行表格分析、模板匹配和数据提取,从而准确和高效地开展数据采集,为加快煤层气田智能化发展提供有效支撑。



## 参考文献(References):

- 白振瑞,张抗. 中国煤层气现状分析及对策探讨[J]. 中国石油勘探, 2015, 20(5):73-80.
- BAI Zhenrui, ZHANG Kang. Analysis of China's CBM Conditions [J]. China Petroleum Exploration, 2015, 20(5):73-80.
- 朱庆忠, 杨延辉, 陈龙伟, 等. 我国高煤阶煤层气开发中存在的问题及解决对策[J]. 中国煤层气, 2017(01):5-8.
- ZHU Qingzhong, YANG Yanhui, CHEN Longwei, et al. Problems and countermeasures in the development of high rank CBM in China [J]. China Coaebei Methane, 2017(01): 5-8.
- 秦勇, 袁亮, 胡千庭, 等. 我国煤层气勘探与开发技术现状及发展方向[J]. 煤炭科学技术, 2012, 40(10):1-6.
- QIN Yong, YUAN Liang, HU Qianting, et al. Status and Development Orientation of Coal Bed Methane Exploration and Development Technology in China [J]. Coal Science and Technology, 2012, 40(10):1-6.
- 宋岩, 张新民, 柳少波. 中国煤层气基础研究和勘探开发技术新进展[J]. 天然气工业, 2005, 25(01):1-7.
- SONG Yan, ZHANG Xinmin, LIU Shaobo. Progress in the Basic Studies and Exploration & Development Techniques of Coalbed Methane in China [J]. Natural Gas Industry, 2005, 25(01):1-7.
- 张仙伟, 尚俊倩, 张鑫, 等. 基于物联网的抽油机参数分析与故障诊断[J]. 西安石油大学学报(自然科学版), 2020, 35(5):47-53.
- ZHANG Xianwei, SHANG Junqian, ZHANG Xin, et al. Parameter analysis and fault diagnosis of pumping unit based on internet of things [J]. Journal of Xi'an Shiyu University (Natural Science Edition), 2020, 35(5):47-53.
- 高伟华. 基于混合智能技术的抽油机电力故障诊断方法的研究[J]. 计算机与数字工程, 2018, 46(9):1905-1910.
- GAO Weihua. Research on power fault diagnosis method of pumping unit based on hybrid intelligent technology [J]. Computer and Digital Engineering, 2018, 46(9): 1905-1910.
- 刘宏伟. 油田生产综合数据统计管理系统[D]. 长春: 吉林大学, 2010:1-2.
- LIU Hongwei. The composite data of oil production management system for statistics [D]. Changchun: Jilin University, 2010:1-2.
- 李目海. 表格的整体识别及C语言实现[J]. 山东师大学报(自然科学版), 1999, 14(4):451-455.
- LI Muhai. The whole recognition of forms and its realization in C language [J]. Journal of Shandong Normal University (Natural Science Edition), 1999, 14(4):451-455.
- 宋华, 戴一奇. 一种用于内容过滤和检测的快速多关键词识别算法[J]. 计算机研究与发展, 2004, 41(06): 940-945.
- SONG Hua, DAI Yiqi. A New Fast String Matching Algorithm for Content Filtering and Detection [J]. Journal of Computer Research and Development, 2004, 41(06): 940-945.
- 王瑞通, 李炜春. 大数据基础存储系统技术研究[J]. 计算机技术与发展, 2017, 27(08):66-72.
- WANG Ruitong, LI Weichun. Research on Technology of Basic Large Data Storage System [J]. Computer Technology and Development, 2017, 27(8): 66-72.
- 伊永祥, 唐书恒, 张松航, 等. 沁水盆地柿庄南区块煤层气井储层压降类型及排采控制分析[J]. 煤田地质与勘探, 2019, 47(05):118-126.
- YI Yongxiang, TANG Shuheng, ZHANG Songhang, et al. Analysis on the type of reservoir pressure drop and drainage control of coalbed methane well in the southern block of Shizhuang [J]. Coal Geology & Exploration, 2019, 47(05):118-126.
- 张松航, 唐书恒, 李忠城, 等. 煤层气井产出水化学特征及变化规律——以沁水盆地柿庄南区块为例[J]. 中国矿业大学学报, 2015, 44(2):292-299.
- ZHANG Songhang, TANG Shuheng, LI Zhongcheng, et al. The hydrochemical characteristics and ion changes of the coproduced water: taking Shizhuangnan block, south of the Qinshui basin as an example [J]. Journal of China University of Mining & Technology, 2015, 44(2): 292-299.
- 康清清, 张熊猫, 孟华. 小秦岭西段稀土矿特征及找矿远景浅析[J]. 西北地质, 2020, 53(1):107-121.
- KANG Qingqing, ZHANG Xiongmao, MENG Hua. Analysis on the Characteristics and Prospecting of Rare Earth Ore in the Western Section of Xiaoqinling [J]. Northwestern Geology, 2020, 53(1):107-121.
- 宿虎, 陈美媛, 张丹青, 等. 高植被覆盖区遥感矿化蚀变信息提取方法研究——以甘肃省西河县大桥—石峡地区为例[J]. 西北地质, 2020, 53(1):146-161.
- SU Hu, CHEN Meiyuan, ZHANG Danqing, et al. Study on the Method of Extracting Information of Mineralization Alteration by using Remote Sensing in High Vegetation Coverage Area——Taking Daqiao-Shixia Area of Xihe County, Gansu Province For Example [J]. Northwestern Geology, 2020, 53(1):146-161.