

文章编号: 1671-8585(2010)04-0294-04

Omega 结构化地震数据文件格式研究

陈有明¹, 武永山¹, 秦绪英¹, 伍燕丹²

(1. 中国石油化工股份有限公司石油物探技术研究院, 江苏南京 210014; 2. 中国石油化工股份有限公司华东分公司资料处理解释中心, 江苏南京 210014)

摘要: 为了使许多自主研发的地震资料处理软件能够直接访问 Omega 地震资料处理系统的数据文件, 对 Omega 地震资料处理系统采用的结构化文件格式 (Structured File, 即 SF 格式) 进行了研究。Omega 结构化文件格式由单一的文件头和一个 (或多个) 表构成, 参数个数可随意增减, 灵活的不定长逻辑记录方式节省存储空间, 数据块式记录存取速度快。通过研究自主开发了直接存取 Omega 结构化地震数据文件的输入输出软件, 并已将其集成到中国石化石油物探技术研究院开发的 iCluster 地震成像软件系统中。

关键词: 结构化文件; 地震数据; Omega; iCluster; 物理记录; 逻辑记录

中图分类号: P631.4

文献标识码: A

国内许多自主研发的地震资料处理软件采用 SEG Y 或 CGG 数据格式, 其文件头和道头长度、参数位置和参数类型固定, 数据记录定长, 浪费存储资源, 不能随意增减参数。西方地球物理公司研制的 Omega 地震资料处理系统等广泛采用结构化文件格式存储地震数据, 文件头和道头长度可变, 参数个数可随意增减, 数据类型和数据个数可自行定义, 灵活的不定长逻辑记录方式节省存储空间, 数据块式记录存取速度快。本文叙述了 Omega 结构化文件的构成、物理记录和逻辑记录结构, 介绍了文件头、表标识、表数据定义和表数据的格式及相互之间的关系。

1 Omega 结构化文件的基本构成

Omega 结构化文件^[1]由单一的文件头和 1 个 (或多个) 表构成, 1 个文件中至少有 1 个表存在。图 1 为 Omega 结构化文件的结构。



图 1 Omega 结构化文件的结构

2 Omega 结构化文件的物理和逻辑记录结构

Omega 结构化文件的数据存/取以物理块为单

位, 在磁盘和磁带上用 VBS 格式进行存储。VBS 的每个字符定义了数据在物理介质上存储的特性:

V——不定长逻辑记录;

B——数据块式记录;

S——跨一个以上物理记录的逻辑记录。

图 2 描述了物理和逻辑 VBS 记录在磁盘或磁带介质上记录的具体情况: ①逻辑记录的长度不是固定的; ②物理记录按数据块存储, 且每一个数据块具有相同的字节数; ③如果一个逻辑记录在物理记录 1 中没有记录完, 可在物理记录 2 中继续记录, 即一个逻辑记录可以跨多个物理记录。

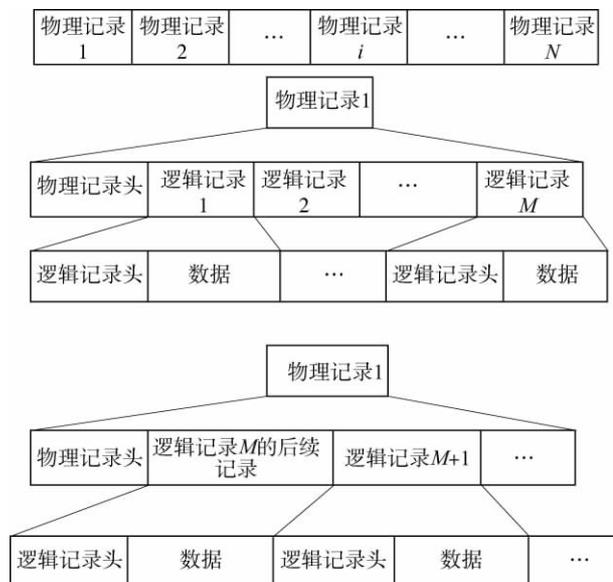


图 2 Omega 结构化文件的物理和逻辑 VBS 记录

收稿日期: 2010-03-20; 改回日期: 2010-05-09。

第一作者简介: 陈友朋 (1951—), 男, 高级工程师, 现从事测井、软件开发工作。

3 Omega 结构化文件的文件头

Omega 结构化文件的第一个逻辑记录是文件头,每个 Omega 结构化文件只有一个文件头。文件头含有对文件进行读取所必须的信息,包括文件的标识符、文件中每个表的定义、表号和表长度。

文件头的前 50 个字节必须按照表 1 所示的格式排列,它标志一个文件为 Omega 结构化文件。如果前 50 个字节的字符串采用 ASCII 码格式,则此文件中的所有文本信息都必须是 ASCII 码格式;如果这 50 个字节的字符串采用 EBCDIC 码格式,则此文件中的所有文本信息都必须是 EBCDIC 码格式。

文件头中表定义部分的信息以“关键字-值”(keyword-value)的格式存储,数据值前面的关键字用于表明该数据值的含义。关键字和值都是字符串形式,每个串至少用一个空格符或缺省分隔符(|)来分隔。一般设定一些特殊字符为分隔符,当

这些特殊字符包含在一个由缺省分隔符所分隔的字符串之内时,就失去分隔符的意义,表 2 为分隔符列表,所能使用的关键字见表 3。

表 1 文件头中的前 50 个字节

字节	内容
1~29	* * * SEGDEF FILE FORMAT VERSION
30	
31~38	01.00 * * *
39	
40~50	空格符

表 2 表定义部分使用的分隔符

字符	ASCII 码 16 进制值	EBCDIC 码 16 进制值
空格	20	40
TAB	09	05
回车	0D	0D
换行	0A	25

表 3 在结构化文件头中使用的关键字

关键字	键值描述和典型值
FILETYPE	结构化文件的类型,典型值有“SEISMIC_DATA”,地震数据
INTCODE	整数格式标识符,典型值有“SLONG”,32 位二进制补码
FPCODE	浮点数格式标识符,典型值有“FSINGL”,32 位 IEEE 码
ACCESSTYPE	该键指出所建立的文件是随机存取或顺序存取,有效值为 RANDOM 和 SEQUENTIAL,有 RANDOM 标识的文件可以随机读写,也可按顺序读写,但是有 SEQUENTIAL 标识的文件不可以随机存取
TIME	建立时间,格式为:HH/MM/SS
DATE	建立日期,格式为:MM/DD/YY
NUMTABLES	文件中表的数目,至少有 1 个
TBIDXXXX	表识别符,这是识别一个表的唯一字符串,关键字内的字符串 XXXX 含一个号码,该号码用于识别文件中所指定的具体的表,号码从 0001 开始,到文件中表的总数为止
TBLNXXXX	表长度,指定表的最大字节数,关键字内的字符串 XXXX 含一个号码,该号码用于识别文件中所指定的具体的表,号码从 0001 开始,到文件中表的总数为止
FHDREND	该关键字的值必须是 END_OF_FILE_HEADER,而且出现在最后一个表识别符——表长度关键字之后,用于识别文件头的结束。

4 Omega 结构化文件中的表

在 Omega 结构化文件中,至少必须有一个表,而多个表则根据需要而定。Omega 结构化文件中的表由表定义和表数据两部分组成,而表定义包含表标识和表数据定义两个部分。表标识部分是一系列文本域,用于在一个表的分层结构中唯一识别

一个表和它的位置。表数据定义部分也是一系列文本域,它定义了表数据部分所存储的信息,表数据部分所包含的信息是依据表数据定义中所给出的布局而确定的。表数据由一个或几个数据项组成,数据项是一个命名实体,它由一个或多个独立的数据元素组成。

4.1 表标识

表标识段使用“关键字-值”结构来存储信息,

如表 4 所示。

在表 4 中列出的仅仅是必须出现在表标识部

分中的关键字。用户还可以自定义关键字,存放在一个标识为 USER_ATTRIBUTES 的块中。

表 4 表标识部分必须出现的关键字

关键字	键值描述和典型值
SECTION	该键值必须为 TABLE_IDENTIFICATION,指出这是一个表标识段
DEFNUM	用于识别该表定义的号码
TABLE_NAME	表的名字
NUMBER_OF_LITERALS	该表所包含数据项的数量
SECTIONEND	该键值必须为 TABLE_IDENTIFICATION,用于标志本段的结束

4.2 表数据定义

表数据定义段使用“关键字-值”结构来存储信息,因此数据定义段是一系列“键-值”对,表 5 是数据定义段使用的键。每一个定义的数据项都需要

DATA_NAME, DATA_TYPE 和 NUMBER_OF_ELEMENTS,其他关键字根据实际情况使用。任何表数据定义的第一个数据项总是 DEFNUM,用于表示是哪一个表的表数据定义部分。

表 5 数据定义段使用的键

关键字	键值描述和典型值
SECTION	必须为 DATA_DEFINITION,说明这是一个数据定义段开始
DEFNUM	用于表示此数据定义段的号码,此值与表标识段中的设置是一样的,由此将表标识和数据定义段建立连接
DATA_NAME	数据记录内数据项的名字,该名字必须来自有效的数据词典
DICTIONARY	本项描述的数据词典,必须含数据词典的名字和版本
DATA_TYPE	一个 FORTRAN 风格的说明(比如 INTEGER * 4, REAL * 8)
UNITS_TYPE	该值是此数据项的单位类别,可取的有效值包括 DISTANCE(距离), TIME(时间)和 UNITLESS(无单位),使用指定单位的情况下通过关键字 UNITS_USED 引用
NUMBER_OF_ELEMENTS	存储在本数据项下元素的个数
SECTIONEND	此值必须为 DATA_DEFINITION,说明这是一个数据定义段结束

在数据定义段,用户可以自定义简单的数据结构^[2],通过使用组(GROUP)结构将几个数据项组合在一起。一个组简单地将一组数据项看作一个单位,用 GROUPSTART 和 GROUPEND 组合在一起,组内的数据项定义的数目与包含在数据记录中的组元素对应。在 Omega 格式 1.0 版本中没有使用组。

4.3 表数据

表数据由一系列数据记录构成,数据记录的内容在表数据定义段中描述,各数据记录要按照表数据定义中的顺序排列,否则就会失去对应关系。如果定义的表多于一个,每个表描述的不同数据记录可以混在一起。每一个数据记录的第一个字节都含有一个识别符,用以表明该数据记录对应哪个表

数据定义的数据项。因此,对任何给定的数据记录都能进行解码。

5 Omega 结构化地震数据文件的格式

一个 Omega 结构化地震数据文件由 3 个必须的表构成,它们分别为 IDHEADER 表、TRACE 表和 HISTORY 表,标准 Omega 数据词典定义了这些表中可以使用的关键字及含义。

5.1 IDHEADER 表

IDHEADER 表用来说明整个文件的特性参数, IDHEADER 表数据定义的最小子集使用的关键字如下。

CLIENT_NAME	UNIT_OF_DISTANCE	SAMP_INT
LINE_NAME	DATA_DESC	MAX_REFLECT_TIME
MAX_TRACE_NUM	AREA_NAME	SORT_LITERAL
EARLIEST_TIME		

5.2 TRACE 表

TRACE 表包含每一个道记录所需的信息,其中 SEISFLOAT 是每道采样点元素的最多个数,

SEISFLOAT	CMP
LTRSAM	TIME_SHIFT_ALIGNMENT
IDENT_NUM	STATION_NUM_SOURCE
TRACE_NUM	STATION_NUM_DETECT

5.3 HISTORY 表

HISTORY 表用于保存有关地震处理过程的文本信息,它只包含一个唯一的数据项,它使用的关键字是 HISTORY,数据类型是字符型。

LTRSAM 说明每道的长度,即每道的实际样点数。以下列出的是 TRACE 表数据定义的最小子集。

STACK_WORD	YCORD_DETECT
XCORD_SOURCE	FIELD_CHANNEL_NUM
XCORD_DETECT	SHOTPOINT_NUM
YCORD_SOURCE	

构化文件格式(1.0 版本)研究的基础上研制了直接存取 Omega 结构化文件的输入输出软件,并已集成在中国石化石油物探技术研究院自主开发的 iCluster 地震偏移成像处理系统中。

6 结束语

本文对 Omega 结构化地震数据文件的结构、物理和逻辑记录作了基本描述。我们在 Omega 结

参 考 文 献

- 1 王宏琳.地震软件技术[M].北京:石油工业出版社,2005.70~76
- 2 严蔚敏,吴伟民.数据结构[M].北京:清华大学出版社,2001.306~322

(编辑:戴春秋)

《石油物探》2011 年征订启事

《石油物探》创刊于 1962 年,面向国内外公开发行人。《石油物探》为中文核心期刊,办刊宗旨为:介绍科研成果,推广先进技术,开展学术讨论,交流工作经验,直接为发展我国油气地球物理勘探事业服务。开设的栏目有:基础理论,方法技术,经验交流,学术论坛,问题讨论等。

《石油物探》全彩色印刷,装帧精美。《石油物探》每册定价为 40 元,全年定价为 240 元。热忱欢迎广大读者订阅。

发行地址:南京市卫岗 21 号石油物探研究所信息中心发行组

邮 编:210014

帐 户:中国石油化工股份有限公司南京石油物探研究所

银行帐号:4301010609100157066 开 户 行:南京市工商银行孝陵卫支行

电 话:(025)84287600 (025)84287718