Implementation of the Regional Meteorological Database

Marek Wojtylak and Krystian Rorbek

Institute of Meteorology and Water Management, Regional Department of Katowice, Bratkow 10, 40-045 Katowice, Poland, admzmbs@imgw.katowice.pl, Krystian.Rorbek@imgw.pl

(Received 4 November 2002; revised manuscript received 31 January 2003)

Abstract: The special character of meteorological data, especially various measurement times and standard statistics, makes a programmer solve non-standard problems. The Regional Meteorological Database (RMD) was created in the early nineties of the last century. The RMD uses a data module stored in files, a control module (for data reading and recording) based on indexing by the hashing functions, and a module which makes it possible to display data.

Keywords: hashing function, meteorological database, pointer file, portion file

1. Introduction

The special character of meteorological data requires an individual approach to record them and get the best performance. The main obstacle to unite meteorological data has been related to various observation frequencies. For example, observations have been performed every 6 hours by climate stations and every hour by synoptic stations. At the same time, some parameters have been observed once a day, others – every 6 hours, still another – every hour, and some – only during their occurrence.

Another problem has been that the number of meteorological parameters depended on the type of an observation post. For example, some of the rainfall measuring stations observe only one parameter, whereas synoptic stations record over 100 parameters.

Therefore, there has been no possibility to create a satisfactory and user-friendly table structure on the basis of the existing database software. In this situation, a decision has been made to develop a dedicated database structure. Thus, the Regional Meteorological Database (RMD) has been created. The RMD is composed of three modules: a customer module, a control module, and a data module.

The Data Module is composed of three files:

- a *portion file*, where data is stored as a portion of a constant size,
- a *pointer file*, where addresses of the first portion data are stored,
- an *auxiliary file*, where information facilitating data management is stored.

It has been decided that a data set corresponding to one meteorological parameter measured within a month by one post will be used as a logical object. Such an object has been called a MIESIAC, and is created in the Control Module.

A meteorological parameter can be measured with varying frequency and, so, the physical size of a month may vary as well. A month is divided into data portions corresponding to parameters measured with various frequencies, and such data portions are stored in the data file. Such a system makes it possible to store parameters in the database exact to a minute.

2. Implementation of the RMD

A user creates a query by defining such parameters as year, month, a meteorological parameter, and an observation post. Then, an object MIESIAC is generated according to this query and a key is evaluated by means of a hashing function. According to this key, the Pointer File is reviewed. The Pointer File is a constant table with a size corresponding to a prime number. The key designates a position in this table, where an address of the first data portion is recorded. Afterwards, it is checked whether the record of the Pointer File contains data corresponding to the required key or not. If negative, the pointer table is searched by means of the cubic method. After finding the address of the right data portion, this data portion is taken from the Portion File and stored in the table of MIESIAC. The data portion contains its number and an address of the next data portion. If a portion isn't the last one, the next portion is extracted and so on, up to a complete filling.

The Control Module uses standard instructions to open a disk file and store its parts to the memory buffer, so that the software needs no database controllers.

3. Current state

The performance of the RMD depends considerably on the filling degree of the pointer table. The more filled the table, the larger number of collisions occur, which means that the number of searchings within the table increases.

Since the starting point, the table capacity has increased four fold (121 441, 255 023, 400 187, 600 011). The table capacity is presently equal to 600 011 and is filled as much as 63.4%.

The Pointer File presently contains 380 217 indexes, which point at 1 045 153 portions stored in the Data File.

These portions include data related to 85 meteorological parameters measured by 214 observation posts. We have at our disposal complete data sets for most of these posts since 1960. However, some data originate from the fifties of the last century.

4. Maintenance of the RMD

The RMD has now operated continuously for over 10 years. Every month, the database size goes up by 210 indexes, which point at 1 000 data portions.

The maintenance of the RMD relies on its rewriting to obtain as large a number of indexes found without collisions as possible. At the same time, wrong indexes are removed.

In the case of the Data File, the maintenance procedure relies on the arrangement of as large data sequences of one element as possible, and especially on the arrangement in neighbourhood of the data portions used in the composition of the same object MIESIAC. Owing to this, data access time will be minimized.

Structure and Extension Forms of the NPL (News on Forest Literature) Reference Database

Ewa Lewandowska

Forest Research Institute,
Bitwy Warszawskiej 1920 r. 3, 00-973 Warsaw, Poland,
E.Lewandowska@ibles.waw.pl

(Received 30 July 2002; revised manuscript received 28 February 2003)

Abstract: The "News on Forest Literature" database contains general reports from journals and serial publications. The database has up to 75 000 records. The database has been created with MICRO CDS/ISIS software. WWW versions of the database contents are made available with EasyInt and ISISWWW software.

Keywords: forest bibliography, EasyInt, ISIS, ISISWWW

Since 1989 a computerized database for NPL (News on Forest Literature) has been maintained at the Scientific Information Department of the Forest Research Institute in Warsaw. It contains scientific and general reports from journals and serial publications (Polish and foreign) on forestry and related fields. The database is handled by the Department workforce and regularly updated from the documents collected in the Institute's library, which is a crucial element contributing to the value of the database. Over 70 thousand reports have been indexed so far, and this forest database is unique in Poland.

The MICRO CDS/ISIS (free of charge) software has been used to create the database. The CDS/ISIS software is a system for collecting and searching information. It was designed especially for computerized management of structural non-numerical databases, *i.e.* databases in which text is the main component. The NPL reference database contains information on papers from journals and each information unit is composed of elementary data: author, title, publication date *etc.*

A printed (hardcopy) form of the database in also available. However, these are printouts of a part of the database (about 350 records) that are published monthly as a publication of the Forest Research Institute in Warsaw (IBL) entitled "Nowości Piśmiennictwa Leśnego" (News of Forest Literature) – NPL. A WWW version of NPL has been presented on the IBL homepage since 1998 (*http://bazy.ibles.waw.pl/bazy/npl/index.html*). It has been possible thanks to the EasyInt (Easy Presentation on Internet) program. The program has been used to present a part of the textual content of the NPL database on the Internet. EasyInt processes text files prepared earlier by the user, containing suitably sorted database records. Such files are created in the CDS/ISIS package when generating printouts recorded into files. Three printouts are made from the bibliographic content of the NPL database, sorted according to authors, keywords and forest classification. EasyInt makes it possible to present the database in a static way, and this means that the content of the database is made accessible on the Internet in a form which is up-to-date at the moment of making printouts, and any changes introduced later will be seen after generating and processing new printouts.

The appearance of the NPL journal on the Internet has aroused interest and approval of users. However, critical remarks have started to appear as well. Users who wanted to information from several years have complained that they had to search through several, even a dozen or so issues of NPL. Moreover, we are aware that not all documented articles have appeared in NPL, due to selection of materials in the process. Therefore, the user has had no possibility