Hitotsubashi Journal of Arts and Sciences 47 (2006), pp.1-7.

INTEGRATION OF BUSCOMBE'S 6TH GENERAL CATALOGUE OF MK SPECTRAL CLASSIFICATIONS*

KOICHI NAKAJIMA**

I. *Introduction*

The meaning of astronomical catalogues and their archiving has been summarized in a previous study (Nakajima 2004). In the processes of catalogue archiving, there is an important process called "integration" which involves incorporating a catalogue into a database system with a standard form. This report describes the integration process of the catalogue "MK Spectral Classification, 6th General Catalogue" by Buscombe (1984), which was converted from printed form to a machine-readable form in the previous paper.

Integration includes the processes of reforming a published catalogue into a suitable form for the Relational Database System, correcting any formal errors, and providing a well-formed instruction of the catalogue, called a "ReadMe" file. Once a catalogue is well integrated, it can be automatically incorporated into data-retrieving systems such as "VizieR" (Ochsenbein et al. 2000). Thus this procedure is important in present astronomical data services. The integration procedures for various astronomical catalogues or data tables are now mainly carried out by CDS (Centre de Donnees astronomiques de Strasbourg), following a detailed instruction manual (CDS 2001). Even in a future data service system "Virtual Observatory" (Genova 2005), integration will become much important.

As many parts of Buscombe's catalogue are non-standard in the view of the Relational Database System, and as the procedures developed in the present work are useful for many similar works, the procedures are reported in detail in this paper. This study is a project of the astronomical database service group in the Astronomical Data Archives Center (ADAC) of the National Astronomical Observatory, Japan.

II. *Rules for Integration, and Our Procedures for Adjusting the Catalogue*

Besides the normal rules of the Relational Database System, there are many standard rules for integration (CDS 2000). For example, the meaning of each field should be clear and definite following a standard. For this reason, the main procedure in this work is to separate those fields of the catalogue that contain several different kinds of indexes or notes into independent ones. Here we report the procedures in detail.

---

* The contribution of the staff at the Astronomical Data Archive Center, National Astronomical Observatory, Japan is greatly appreciated. The advice of Dr. Shiro Nishimura in compiling the catalogue is also greatly appreciated.

** Emeritus Professor, Hitotsubashi University.

Buscombe's 6th catalogue (Buscombe 1984) collects MK spectral classification data of stars from various publications. It presents the results in a table whose fields are:

1) Identification for the star, with a note of multiplicity
2) Right ascension (2000)
3) Declination (2000)
4) MK spectral type
5) Johnson V magnitude
6) Johnson B-V color
7) Johnson U-B color
8) Alternate identification, usually HD

A detailed explanation of the columns and a sample page of the printed publication are shown in the previous paper.

In order to incorporate the data into the Relational Database System and into the data-retrieving system, at least one field of key index, and a field (or fields) of unique and standard identifier should be provided. Every data field should contain a standard formatted datum, and the main part of the catalogue (i.e. MK spectral type) should be described in a standard form that can be directly used in the statistical analyses. Column misalignments and non-numeric characters in the data columns should be corrected.

The key index field can be simply supplied adding a field of serial number. It is also necessary to add a field for the number of this catalogue series from 1 to 15, in case of joining them in a future work.

The position data (i.e. right ascension and declination) are often used as a unique identifier. However, these data are not necessarily described with sufficient precision in this catalogue, especially for those stars derived from star cluster observations. (We call them "cluster stars" hereafter.) Therefore in this catalogue the fields of identifications (i.e. Field 1 and 8) are used as the unique identifier.

However, descriptions in these fields are not uniform in the catalogue. It contains the note of multiplicity (e.g. A, B, or AB), BD, CD, HD numbers, or serial number with a star cluster name. We need to separate them into a set of several fields. Following "All-sky Compiled Catalogue of 2.5 million stars" (Kharchenko 2001, hereafter ASCC), we adopt two fields; one is for HD number and the other for BD, CD, or CP number, i.e. "DM number" in ASCC.

Because the cluster stars do not have such identifiers, and because the positions of these stars are not shown with sufficient precision, we add two more fields, one of which is the name of the cluster data source and the other is its serial number.

The notes of multiplicity should be separated into a new field.

Also there are several other notes, such as "VAR", "LMC", etc., and another new field is created for them.

Several non-numeric characters have been included in the fields of magnitude and colors. Because these fields are only for reference as Buscombe described, these characters are omitted without leaving any note.

The most important field of this catalogue is the MK spectral type description (i.e. Field 4). It contains mainly three types of data, i.e. spectral type (A-Z characters), its subdivision (floating point data), and luminosity class (Roman numbers). In order to utilize the catalogue for various statistical analyses, these categories should be separated as independent fields.

However, the field in Buscombe's catalogue contains many notes and comments besides the three types, such as variability, multiplicity, or those notes included in the sources of the data. Because these notes are so many and so complicated, and because the sources of the data are not indicated in the catalogue, we do not separate these categories, in this work. Neither of the catalogues of this series (i.e. 1-15) nor that of Skiff's (Skiff 2005) yet separate the categories. This must be taken into account when we combine all these catalogues in future.

In these kind of compiled catalogues, it is important to indicate the source(s) of data. However, no such indication is given in Buscombe's catalogue. Because he is deceased, it seems impossible to find them in Buscombe's works. It may possibly be re-compiled in the near future, if all important astronomical data and tables are incorporated in the database of the Virtual Observatory.

Thus the fields in our newly processed catalogue are as follows:

1) Serial number, in the same order in the original (Buscombe's) catalogue
2) HD number
3) BD, CD, CP identification, i.e. DM number, as in ASCC catalogue
4) Cluster catalogue name, if any
5) Cluster catalogue number, if any
6) Notes in Fields 1 and 8 of the original catalogue
7) Multiplicity
8) Right ascension, described in the original catalogue
9) Declination, described in the original catalogue
10) Magnitude, described in the original catalogue
11) B-V color, described in the original catalogue
12) U-B color, described in the original catalogue
13) MK spectral type, described in the original catalogue.

III. *Actual Procedures*

After the OCR works described in the previous paper, we acquired a machine-readable table of Buscombe's 6th catalogue. In the table we acquired, such errors in the original catalogue as column misalignments or position errors, which are clearly identifiable, were corrected in the course of the consistency check of the OCR results. Also non-numeric characters in the numeric fields were omitted. Omitting a line that was canceled in the original catalogue, we obtained 18773 records of data.

Although the fields of main data (i.e. Fields 2-7 of the original catalogue) were checked carefully in the previous work, fields of identifications (i.e. 1 and 8) were not fully checked because they contained many misidentified characters due to the difficulty of OCR identifications in these fields. Taking account of this problem, we carried out the following procedures.

1. Identify stars in ASCC catalogue and find HD and/or DM numbers

Using position data in Buscombe's catalogue, we selected stars from the ASCC catalogue (Kharchenko 2001) within 2 arcmin from the position. This procedure was performed by

VizieR service, with an input file made from cut out position data. Here we did not use position data with insufficient precision, which are those of cluster stars

In some cases, 2 or 3 stars were selected around the target object. In this case, we selected only one candidate star by eye-inspection, comparing positions and magnitudes with the target. In some cases no correspondent star was selected, or different stars were selected. In this case we considered it as “not-selected”, even though HD or DM numbers are written in Buscombe’s catalogue.

Those stars whose position data was insufficient were considered as not-selected stars for HD and DM numbers.

2. Select stars that have coincident HD numbers

First, we selected computationally those stars that have the same HD numbers in both catalogues. 8147 stars were selected. Then, we selected those starts that have slightly different HD numbers, and inspected them one by one to correct OCR identification errors. We dealt with 1705 stars.

Finally, we selected 9831 stars with definite HD numbers. In this procedure we found and corrected several input errors in Buscombe’s HD numbers.

3. Select stars that have coincident DM numbers

We checked by eye-inspection 1638 stars that have only DM number in the ASCC catalogue (i.e. no HD number). We selected 921 stars with coincident DM numbers.

4. Select stars that have different HD and/or DM numbers

From the remaining 9070 stars we selected those stars whose HD or DM numbers were slightly different from those of ASCC. Although the possibility of input error cannot be neglected, we adopted original numbers, and put them into each field of our new catalogue. If the position and magnitude were almost the same and if the numbers were different, we added a note for future checking. We selected 290 such stars.

5. Select stars that have HD and/or DM numbers and do not have the numbers in ASCC

This case is almost the same as the above case, and we simply adopted the written numbers. Some of them were checked referring to other catalogues such as ASCC, and notes were added as mentioned above.

6. Select the stars that have cluster catalogue identifications

At first we selected those records that include a name of the cluster catalogue, using the “grep” command. For the remaining records, we checked by eye-inspection one by one. This procedure was the most time-consuming part of this work. Names of catalogues of star clusters and the serial numbers in them were separated into 2 fields.

7. Multiplicity field

The notes for multiplicity, i.e. "A", "B", "C", "AB", or "AC", can be identified computationally. Identifying them, we put them into a new field. We put note "ABC" etc., which is very rare, into the field of other notes. (See below)

8. Field of other notes

Both the first and the last field of Buscombe's table were inspected for this purpose, extracting various notes. Eliminating HD number, cluster catalogue name and number, and multiplicity notes, we extracted notes for this field.

This procedure is performed mostly computationally, but all the results were checked by eye-inspection one by one.

IV. *Discussion and Future Works*

After very complicated, laborious work, we integrated Buscombe's "6th General Catalogue of MK Spectral Classification". A sample is shown in Table 1. All results were archived and made public at ADAC in its catalogue archive service (ADAC 2006). After some technical procedures, it is scheduled to become public at CDS.

The procedures and computational programs developed here are useful for successive works integrating other catalogues in this series. Although several catalogues in this series have already been included in the CDS catalogue archive (e.g. 12 to 14th ones, CDS 1998-2005), the fields of these integrated catalogues are the same as the original one, and cannot be used for computational retrieval or for statistical analysis.

However, it remains impossible to use these catalogues for statistical analyses even though these catalogues are fully integrated by the procedures described in this work. The reasons are:

1) This series of catalogues of MK classification is only a compiled catalogue, and not the survey one.
2) Estimations of classifications are not uniform, i.e. do not follow any finite standard of classification.
3) Notes of sources for the classifications are incomplete, and we cannot check the data further even if there are doubtful descriptions.

As Buscombe wrote in the explanations in the catalogues (Buscombe 1977-2001), these catalogues are useful to quickly find MK classification data for special purposes. However much simpler and more useful services such as "SIMBAD" (CDS 2006) are now available for investigating restricted numbers of stars. Thus catalogues such as this are mainly meaningful in their statistical use.

In order to attain this goal, it is necessary to combine these catalogues with a comprehensive star catalogue such as Hipparcos and Tycho (ESA 1997), or even USNO-B1.0 (Monet et al. 2003). We will take account of this procedure in our future works in which we integrate all catalogues in this series.

After this integration, we will be able to clarify the areas or groups of coverage by this

TABLE 1. A SAMPLE OF THE

No.	HD	DM	CatName	CatNo.	Notes	Mul.
1			T	61		
2	224732	−4015306				
3	224721	3704912				
4	224710	1905184			EP PEG	
5	224769	3005071				
6			EPSTEIN	3		
7			EPSTEIN	26		
8	224792	6102580				
9			CASE	268		
10	224782	−5309468			CP−5310565	
....	....	....	....	....	....	....

For explanations of table columns, see the text.

joined catalogue. Using this result, we can estimate its completeness, and then its incompleteness for which a further survey observation is needed. These further works should be taken over by those researchers whose principal research theme are those accomplished by using these data.

V. *Concluding Remarks*

It is widely accepted among astronomers that as much astronomical data should be integrated as possible, in order to utilize them computationally via the Internet. This is one of the most important purposes of the worldwide Virtual Observatory Project (Quinn et al. 2004). This necessity is emphasized also in Commission 5 of the International Astronomical Union (IAU), which deals with “Documentation and Astronomical Data” (e.g. Norris 2006).

This requirement includes not only archiving newly published astronomical data but also converting old printed data and catalogues into machine-readable form integrating them as a standard format. The previous work and this work meet the latter requirement.

However, the work of archiving old data is extremely laborious as described above, and it may be impossible to carry out such a project systematically in the astronomers consortium.

The recommendations published by Commission 5 state that such a project is important and should be tackled cooperatively (IAU Resolution 2006). However, it is not practical to organize a voluntary team for such a purpose because the work is too laborious to accomplish without any reward.

This work should be undertaken by the users of the data, i.e. by those astronomers who need the data for their own researches. The main role of the Virtual Observatory is to provide a well-organized registry or meta-data that is useful for these researchers. The limited resources of the astronomers’ effort should be oriented first toward works that meet the present demands before the future ones.

HITOTSUBASHI UNIVERSITY

Top 10 Lines of Our Results

R.A.	Dec.	Vmag.	B-V	U-B	MK spectral type
0 00.1	+61 10	9.81			A 0 V
0 00.1	−40 12	8.13	0.43		F 3 V
0 00.2	+38 18	6.54	0.96	0.72	G 8 III
0 00.2	+20 14				M 7 II:
0 00.4	+31 12	8.2			K 2 III
0	−30				F 4 V
0	−30	9.85			G 6 IV
0 00.6	+62 10	7.05	0.49		F 6 V
0 00.6	+59 52				C
0 00.6	−53 06	6.5			G 1: IV
....	....	....	....	....	....

References

ADAC 2006, ftp://dbc.nao.ac.jp/DBC/ADACnew/3/3994/

Buscombe, W. 1977-2001, "MK spectral classifications. Third — Fifteenth general catalogue" (Northwestern Univ., Evanston, Illinois)

Buscombe, W. 1984, "MK spectral classifications. Sixth general catalogue" (Northwestern Univ., Evanston, Illinois)

CDS 1998-2005; 12th (2005) http://cdsweb.u-strasbg.fr/ftp/cats/3/223/
13th (1998) http://cdsweb.u-strasbg.fr/ftp/cats/3/206/
14th (2003) http://cdsweb.u-strasbg.fr/ftp/cats/3/222/

CDS 2000, http://vizier.u-strasbg.fr/doc/catstd.htx

ESA 1997, "The Hipparcos and Tycho Catalogues"(ESA Publications Division)

IAU resolution 2006, (in press)

Kharchenko, N.V. 2001, Kinematics and Physics of Celestial Bodies 17, 409

Monet, D.G. et al. 2003, Astron. J., 125, 984

Nakajima, K. 2004, Hitotsubashi Journal of Arts and Sciences, 45, 1

Norris, R 2006, "Astronomers' Data Manifesto" (Documents of The IAU Working Group on Astronomical Data, in http://www.ivoa.net/twiki/bin/view/Astrodata/Astronomers-Manifesto)

Ochsenbein, F. et al. 2000, Astron. Astrophys. Suppl. Ser. 143, 23

Quinn, P. et al. 2004, A Discussion Paper for the OECD Global Science Forum, http://www.ivoa.net/pub/info/TheIVOA.pdf

Skiff, B.A. 2005, "General Catalogue of Stellar Spectral Classifications" (Lowell Observatory)