

Проблемы обозначения и кросс-идентификации кратных объектов в астрономии

© Д.А. Ковалева

dana@inasan.ru

© П.В. Кайгородов

Институт астрономии РАН, Москва

pasha@inasan.ru

© О.Ю. Малков

malkov@inasan.ru

© Л.А.Калиниченко

Институт проблем информатики РАН, Москва

leonidandk@gmail.com

© Н.А.Скворцов

nskv@mail.ru

Аннотация

Работа продолжает и развивает направление исследований, ориентированных на решение задач всестороннего анализа массивов разнородных данных имеющимся арсеналом научных методов и инструментов для выявления полезной информации и получения новых знаний. Здесь рассмотрены проблемы обозначения и кросс-идентификации кратных объектов в астрономии, с акцентом на упорядочение обозначений и интеграцию многомерных наблюдательных данных различных типов. В работе кратко обзревается существующие методики обозначения одиночных и кратных астрономических объектов (с акцентом на применяемую в Базе данных двойных звезд BDB схему обозначений BSDB), а также принципы кросс-идентификации.

1 Идентификация небесных объектов

Астрономические объекты стали каталогизироваться еще во II в. н. э. Астрономические каталоги, созданные до начала XVII в., насчитывали до полутора тысяч объектов, а затем, с изобретением телескопа, число объектов стало стремительно расти. Каталоги 70-х годов прошлого века, когда стали образовываться центры астрономических данных и создаваться астрономические базы данных, насчитывали до 2 млн объектов, а современные каталоги включают млрд объектов, и в ближайшем будущем это число увеличится на 1-2 порядка.

Необходимо заметить, что постоянно растет не только число объектов, подлежащих каталогизации, но и количество каталогизируемых параметров.

Труды 16-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL-2013, Дубна, Россия, 13–16 октября 2014 г.

Если первые каталоги включали 3-4 параметра на объект (две координаты, визуальная оценка блеска и грубый классификатор уровня «звезда–туманность»), то современные каталоги характеризуют объект десятками и сотнями параметров, не считая различных функций (распределение энергии в спектре, функция изменения блеска со временем и пр.).

При наименовании небесных объектов (необходимом для каталогизации) астрономы не следовали какому-то одному правилу, точнее, применяли различные подходы. Хронологически первыми нужно считать методы, опирающиеся на систему созвездий – поименованных участков небесной сферы, содержащих группы звезд. Были введены в обращение, в частности, система Байера, именующая звезды в созвездии греческими буквами в порядке убывания яркости; система Флемстида, просто нумерующая звезды в созвездии с увеличением их прямого восхождения (т.е., для северного полушария Земли, слева направо); система Аргеландера-Хартвига (для переменных звезд, в порядке их открытия), использующая сложную комбинацию двухбуквенных обозначений и, в случае их нехватки, порядковых номеров. Нехватка букв в греческом алфавите (особенно для протяженных созвездий) стала очевидной довольно быстро, а система Флемстида потеряла свою стройность, если в созвездии обнаруживали новую, более слабую звезду, расположенную на небесной сфере среди уже перенумерованных.

С введением систем небесных координат исследователи получили возможность создавать имена (идентификаторы), базирующиеся на значениях координат. Идентификаторы компилировались из округленных (или, иногда, сокращенных) значений координат, при этом для объектов Галактики применялась экваториальная, а для внегалактических объектов – галактическая система координат. Такие системы уже не были связаны с границами созвездий, но тоже оказались не совсем устойчивыми: из-за прецессии и нутации земной оси, а также из-за собственных движений

координаты небесных тел изменяются (у некоторых – весьма значительно). Таким образом, координатно-ориентированные имена звезд оказывались, вообще говоря, разными в разных системах, если введение этих систем в строй отстояло друг от друга на более чем 10 лет (это число зависит, конечно, от позиционной точности каталога).

Кроме того, каталоги, базирующиеся на таких системах и традиционно отсортированные по значению прямого восхождения (аналог земной долготы), оказались весьма неудобны для планирования наблюдений, так как в них могли соседствовать объекты с очень разными значениями склонений (аналог земной широты). С увеличением числа каталогизируемых небесных объектов этот недостаток становился все более досадным, и авторы каталогов, содержащих более 100 тыс. объектов в конце концов перешли на зонную систему идентификации. Небесная сфера «нарезалась» на зоны (пояса) по склонению (а позже, для каталогов с числом объектов, превышающим 10 млн, и эти пояса пришлось «порезать» на более мелкие зоны), и объекты в каталоге сортировались по зонам, а внутри зоны – по прямому восхождению.

Наконец, в «тематических» каталогах, содержащих объекты (или параметры объектов) определенного типа, принято просто нумеровать их в порядке открытия. Помимо упомянутых выше переменных звезд можно привести в пример каталог спектроскопических двойных систем и (с некоторыми оговорками) ставший в последнее время популярным, в связи с открытием экзопланет, каталог близких звезд Глизе (V/70A, указана нумерация каталога в Базе данных каталогов VizieR [1]).

Таким образом, небесные объекты оказались поименованы и включены в различные каталоги в соответствии с их:

- положением в созвездии,
- координатами,
- координатами в зонах небесной сферы,
- блеском (разным в различных фотометрических системах),
- очередностью открытия
- или вообще без всякой системы, как, например, объекты в широко известном каталоге туманностей Мессье [2].

Нужно отметить, что практически все эти (даже самые архаичные) системы наименований сохранились, успешно применяются по сей день и продолжают создавать проблемы для успешной кросс-идентификации небесных объектов.

Так, самые яркие (и следовательно, наиболее хорошо изученные и включенные во многие каталоги) звезды имеют 4-5 десятков общеупотребительных наименований.

Данная работа посвящена идентификации звезд, однако практически все вышесказанное справедливо для туманностей и скоплений нашей Галактики и для более удаленных объектов (галактик и квазаров). В статье рассматриваются, с одной стороны, схемы идентификации одиночных и кратных звезд как системы их обозначения в каталогах, а с другой стороны, – методы кросс-идентификации звезд, использующие как их идентификаторы, так и наблюдательные и астрофизические параметры для отождествления звезд в разных каталогах.

2 Кросс-идентификация небесных объектов

Разнообразие методов создания астрономических каталогов поставило астрономов перед задачей выработки схем кросс-идентификации содержащихся в них объектов. Действительно, точная позиционная информация, содержащаяся в каталоге А, будучи проанализированной совместно с высокоточной фотометрией из каталога Б и данными радионаблюдений из каталога В, позволяет получить более полную картину образования, строения и эволюции Галактики, чем данные из каждого из этих каталогов по отдельности. При этом три упомянутых каталога создавались разными коллективами по разным методикам, вообще говоря, в разное время и использовали различные системы идентификации. Базы астрономических данных, объединяющие неоднородную информацию, требуют в первую очередь решения проблемы кросс-идентификации объектов.

Проблема кросс-идентификации (КИ) астрономических объектов состоит в отождествлении одних и тех же объектов среди неоднородных данных из разных каталогов. Она является частным случаем рассматриваемого в информатике направления разрешения сущностей (entity resolution). Комплексный подход к разрешению сущностей обычно представляется в виде определенной последовательности действий, включающей [3, 4]:

- связывание элементов схем данных разных источников, соответствующих по смыслу, предварительную очистку и приведение к единообразному представлению неоднородных данных из связанных атрибутов;
- индексирование данных с целью уменьшения попарного перебора сравниваемых кортежей из разных источников;
- применение методов сравнения данных различных типов, включая строки, числовые типы, даты, пространственные координаты, множества, записи в целом;
- выделение набора данных, однозначно определяющих уникальные объекты, а при невозможности его выделения оценка близости

данных из разных источников по определённым критериям и принятие решения об отождествлении описанных ими объектов.

Естественной основой разрешения объектов применительно к решению задачи кросс-идентификации небесных объектов являются их пространственные координаты. Данные атрибутов, содержащие координаты объектов, приводятся к единообразному представлению с учётом используемых форматов координат и разности эпох наблюдения. Близость координат с определённым допущением решает проблему уменьшения перебора попарного сравнения кортежей в каталогах. Координатное совмещение данных различных каталогов позволяет решать до 80% проблем, связанных с КИ.

Однако чисто координатного подхода оказывается недостаточно, если речь идет о плотных звездных полях (скоплениях или кратных системах, см. ниже), о быстро движущихся и/или переменных объектах, о данных с различающимся угловым разрешением, а соответственно, различной точностью координат и т.п. В таких случаях почти всегда для КИ удастся использовать атрибуты каталогов, содержащие фотометрическую информацию. При приведении фотометрических данных к единообразному представлению приходится принимать во внимание тот факт, что блески (или цвета, т.е., разницы блесков) объектов в различных фотометрических системах, вообще говоря, различны и, хотя и подчиняются неким корреляционным соотношениям, соотношения эти опять-таки различны для объектов различной природы (которая, как правило, при КИ остается неизвестной). Более детальное описание общих принципов кросс-идентификации объектов в астрономии можно найти в [5].

Полезным дополнительным параметром является также классификатор объекта, однако он присутствует в каталогах далеко не всегда и представляет, как правило, достаточно грубую оценку природы объекта (точечный/протяженный). Естественно, используется и вся другая информация (например, спектральный тип объектов), если она присутствует в обоих кросс-идентифицируемых каталогах.

Таким образом, критерии отождествления и оценки близости объектов при кросс-идентификации разрабатываются с учётом позиционных и фотометрических параметров, рассчитанных астрофизических величин, параметров фотометрических систем и углового разрешения оборудования, используемого для обзоров.

Одной их особенностей данных астрономических каталогов является то, что некоторые из них уже включают данные об именах описываемых объектов в соответствии с определёнными системами идентификации, принятыми в других каталогах. Приведённые в каталогах имена идентифицируют соответствующие

объекты в других каталогах, и таким образом, являются результатом уже решённой при их составлении задачи отождествления между конкретными парами каталогов.

Однако кросс-идентификация для разных пар каталогов может производиться различными методами, различные виды идентификаторов имеют разный смысл. Использование разных критериев отождествления объектов, различных систем идентификации и способов связывания имён в каталогах может само по себе рождать конфликты идентификации. В том числе, точность используемых методов отождествления невозможно выяснить из самих идентификаторов. Поэтому для проверки корректности существующих связей идентификаторов объектов и для разрешения конфликтов между идентификаторами при отождествлении объектов более, чем в двух каталогах, приходится заново прибегать к решению задачи КИ на основе наблюдательных и астрофизических параметров.

С учетом изложенных выше соображений астрономам удалось решить (и удастся решить, с появлением новых каталогов и обзоров) большинство проблем КИ, что находит свой результат в публикуемых таблицах КИ и создаваемых базах астрономических данных различных типов. Приведённые выше критерии, однако, не во всех случаях достаточны для решения задачи кросс-идентификации. Нередко возникает необходимость в более специфических методах. Для некоторых типов объектов задача кросс-идентификации далека от окончательного разрешения, и, в первую очередь, это относится к двойным и кратным звездам.

3 Особенности идентификации и кросс-идентификации кратных звезд

Двойные и кратные звезды весьма многочисленны и не исключено, что их доля среди звезд Галактики (если включать в их число и планетные системы) весьма близка к 100 процентам. Столь высокая кратность объясняется особенностями звездообразования, в частности, необходимостью для вращающегося и сжимающегося протозвездного газо-пылевого облака избавиться от осевого момента инерции, что проще всего осуществить за счет фрагментации на компоненты и/или образования планетной системы.

Далеко не все двойные звезды наблюдаются именно как двойные. Для этого паре нужно либо находиться достаточно близко к наблюдателю и быть достаточно широкой (тогда компоненты будут наблюдаться по отдельности), либо демонстрировать доплеровское смещение линий в спектре и/или переменность блеска из-за орбитального движения компонентов, либо проявлять себя как источник рентгеновского излучения (возникающего из-за аккреции вещества на один из компонентов) и т.п. Эти и другие типы

двойных регистрируются с помощью различных методик различными коллективами и им, естественно, присваиваются обозначения в рамках различных схем идентификации. Таким образом, обозначение одиночного объекта в некотором каталоге должно быть приписано двойной системе в другом каталоге, когда эти каталоги имеют разное пространственное разрешение (т.е., речь идет о кросс-идентификации объектов различных категорий). Задача еще сильнее усложняется для объектов большей кратности. В результате к «традиционным» проблемам схем идентификации прибавляется несколько новых, характерных именно для двойных (кратных) систем.

Прежде всего, требуется разработать методику обозначений для идентификации компонентов кратной системы. Для двойных звезд эту проблему решали традиционно, добавляя к идентификатору системы в качестве суффиксов буквы *A* и *B*. Но уже с тройными системами поступали по-разному. В тесных системах, когда оказывалось, что компонент *A* представляет собой на самом деле двойную звезду, новые два компонента получали обозначения *Aa* и *Ab*. Этот принцип, помимо прочего, отражал и тот факт, что кратные системы (кроме самых широких) должны быть иерархическими, иначе они будут динамически нестабильными и просуществуют недолго. Исследователи же широких систем, где компоненты, как правило, наблюдаются по отдельности, а уровни иерархии не очевидны, обозначали вновь открытый компонент буквой *C*. Аналогично эти принципы распространились на системы более высокой кратности.

Эти схемы, естественно, не идеальны. Помимо того, что появляются трудно форматируемые обозначения типа *Aa1*, а в неиерархических системах особенно высокой кратности (которые некоторые исследователи, впрочем, предпочитают называть скорее скоплениями) не хватает букв латинского алфавита. Открытие компонента на промежуточном иерархическом уровне является более редким событием, но также нарушает описанные выше принципы наименования объектов.

Еще одной, дополнительной трудностью, присущей даже двойным системам является порядок присвоения букв *A* (главному) и *B* (вторичному) компонентам, точнее, неоднозначность ответа на вопрос, какой компонент в паре является главным. Для исследователей визуально-двойных звезд это – более яркий компонент (оставим в стороне вопрос о порядке присвоения букв в парах с компонентами одинаковой яркости, а также то обстоятельство, что в разных фильтрах относительная яркость компонентов может быть разной, а в некоторых случаях самым ярким агентом в системе является даже не звезда, а аккреционный диск вокруг одной из звезд), для исследователей переменных звезд – более горячий. При моделировании тесных двойных систем принято считать главным компонентом более массивную на сегодняшний день звезду, а с

точки зрения звездной эволюции главный компонент – изначально более массивная звезда (из-за переноса массы в системе в процессе эволюции это могут быть разные компоненты). С точки зрения кинематики двойной системы главный компонент – меньший по массе. И существуют, наконец, задачи, для которых удобно считать главным компонентом больший по размерам. Присвоение букв (*B*, *C*, ...) компонентам в системах большей кратности также может осуществляться по-разному: в порядке уменьшения блеска, в порядке удаления от главного компонента *A* и т.д.

Все эти обстоятельства приводят к тому, что компоненты (и сами системы) двойных и кратных звезд получают в различных каталогах весьма различные обозначения (присваиваемые в соответствии с различными схемами идентификации), и задача КИ, более-менее решенная для одиночных объектов, становится гораздо более сложной для двойных и кратных систем.

4 Поиск информации в Базе данных двойных звезд

В данном разделе описаны принципы поиска и отображения информации в Базе данных двойных звезд (Binary star DataBase, BDB, [6]), которая разрабатывается в настоящий момент в Институте Астрономии РАН и содержит данные о порядка 110 000 звездных систем с кратностью от 2 и выше. Данные, содержащиеся в BDB, получают путем объединения множества каталогов (с разными принципами организации и разными системами идентификации), содержащих информацию о двойных и кратных звездах разных наблюдательных типов: визуальных двойных, спектральных, рентгеновских, астрометрических, интерферометрических, спектроскопических, фотометрических и т.д. Из этих каталогов извлекаются как наблюдательные данные о координатах, собственных движениях, периодах, переменности, звездных величинах, так и астрофизические параметры – эволюционный статус, массы и т.п.

Запрос данных в BDB возможен либо с использованием идентификатора, либо при помощи поиска по параметрам. При поиске по идентификатору пользователь может выбрать систему идентификации из нескольких, включенных в BDB: ADS, Bayer, CCDM, DM, Flamsteed, GCVS, HD, HIP, IDS, IGR, Name, SBC9 или WDS; либо ввести идентификатор в свободной форме. В последнем случае BDB прежде всего попытается найти введенный идентификатор среди имеющихся в базе. Если найти идентификатор не удалось, будет сделан запрос (с использованием протокола SOAP) к системе, связывающей обозначения (name resolver) Sesame [7], и среди выданных ею результатов будет выбран идентификатор, имеющийся в BDB.

Здесь необходимо отметить, что иногда, когда таких идентификаторов несколько, и не все они принадлежат одному и тому же объекту BDB, мы получаем указание либо на наличие кратной системы, состоящей из двух других систем, либо на ошибки, содержащиеся в Sesame (исключая, конечно, тривиальный случай, когда эти идентификаторы принадлежат двум разным компонентам одной системы).

Результаты запросов к Sesame кэшируются. После получения идентификатора:

1. BDB находит все записи (системы, пары или компоненты), ссылающиеся на данный идентификатор. Найденные записи включаются в список (здесь и далее в список включаются только ранее отсутствовавшие в нем элементы).

2. Для каждой найденной записи находятся (и также включаются в список) записи с тем же внутренним идентификатором, что и у нее (каждому объекту в BDB может соответствовать несколько записей, поступивших из разных каталогов, но имеющих один внутренний идентификатор). В качестве внутреннего идентификатора используется имя BSDB (см. ниже).

3. Если найденная запись относится к системам, то в список включаются все пары, ссылающиеся на нее. Для пары включаются в список оба ее компонента, а также система, на которую она ссылается. Для компонента в список включается его родительская пара.

4. Составляется (пополняется) список внешних идентификаторов, на которые ссылаются записи, найденные на предыдущих этапах.

Перечисленные этапы повторяются циклически для всех найденных идентификаторов до тех пор, пока формируемый список не перестанет расти.

5 Система идентификации BSDB

При создании Базы данных двойных звезд BDB авторами разработана схема обозначений BSDB, призванная разрешить существующие проблемы идентификации и кросс-идентификации двойных систем. BSDB должна была удовлетворять (и удовлетворяет) следующим критериям:

- ни один объект не должен носить более одного идентификатора;
- ни один идентификатор не должен быть присвоен более, чем одному объекту;
- открытие новых компонентов в системе не должно нарушать принципы присвоения идентификаторов;
- система должна быть несложной, близкой к традиционным и интуитивно понятной исследователям двойных.

При присвоении идентификатора по системе BSDB мы выделяем три категории объектов: система, пара, компонент. Это подход следует считать пионерским, и диктуется он тем

обстоятельством, что каждая из трех категорий характеризуется своим набором наблюдательных данных. Компонент характеризуется массой, радиусом, температурой, светимостью, и т.п. (то есть, тем набором астрофизических параметров, которым характеризуется, например, одиночная звезда). Пара – это два объекта (каждый из которых, кстати, тоже может оказаться парой), связанных гравитационно. Эта категория характеризуется такими параметрами как относительное положение членов пары на небесной сфере (для визуальных двойных), орбитальными параметрами (период обращения, эксцентриситет орбиты и пр. – для орбитальных и части спектроскопических двойных), интегральным блеском (для фотометрически неразрешенных) и спектром (для спектроскопически неразрешенных). Здесь следует заметить, что наблюдатели имеют дело преимущественно именно с парами, и именно информация о парах, как правило, и включается в каталоги. Наконец, такая категория как система характеризуется общими параметрами: возраст, металличность, расстояние, кинематика в Галактике и пр. Некоторые параметры могут приписываться различным категориям: так координаты характеризуют каждый из компонентов в случае разрешенной двойной и пары – в случае неразрешенной.

Идентификатор BSDB состоит из цифровой части, компилируемой из значений небесных координат и предваряемой символом ‘*J*’ (означающим, что координаты относятся к эпохе 2000.0 года); индикатора «система–пара–компонент» (“*s*”, “*p*”, “*c*”), отделяемого двоеточием; и буквенным обозначением, в общих чертах напоминающим знакомые исследователям двойных звезд схемы обозначений. Так, обозначения объектов некой тройной системы будут выглядеть следующим образом:

*J*000144.48+590527.1:*s*
*J*000144.48+590527.1:*pAa-Ab*
*J*000144.48+590527.1:*cAa*
*J*000144.48+590527.1:*cAb*
*J*000144.48+590527.1:*pA-B*
*J*000144.48+590527.1:*cB*

Отметим, что в списке отсутствует компонент *A*, поскольку в данной кратной системе объект *A* является не звездой, а парой звезд (*Aa-Ab*) и, соответственно, описывается параметрами, характерными для пары (например, период обращения), а не для компонента (например, масса).

Координатная часть обозначения BSDB внутри системы не меняется, несмотря на то, что координаты компонентов, вообще говоря, могут различаться.

Принципы создания идентификатора BSDB удовлетворяют правилам, утвержденным Международным астрономическим союзом.

6 Решение проблем кросс-идентификации двойных и кратных систем

Схема обозначений BSDB должна быть всеобъемлющей, поэтому необходимо позаботиться о том, чтобы кратные системы всех наблюдательных типов могли получить в ней соответствующие обозначения. Для этого коллективом ведется работа по созданию общего каталога идентификаций двойных звезд (предварительное название – Identification List of Binaries, ILB), который должен включать обозначения BSDB для всех каталогизированных в настоящее время двойных систем, а также предоставить такую возможность и для будущих списков/каталогов/обзоров двойных.

К каталогу ILB постепенно подключаются каталоги двойных систем, начиная с самых широких и, одновременно, самых представительных. Каждому объекту, встречающемуся впервые, присваивается уникальное обозначение BSDB. Объекты, уже имеющиеся в ранее исследованных каталогах, дописываются в соответствующие строки ILB. Новые объекты, входящие в уже существующие в ILB звездные системы, приводят к корректировкам соответствующих разделов каталога. При этом приходится решать проблемы КИ, которые возникают даже в том случае, когда кратная система принадлежит только одному наблюдательному типу и, следовательно, ее составляющие поименованы хоть, возможно, и по-разному, но, по крайней мере, в соответствии с одной и той же схемой идентификации. Задача усложняется, когда в системе присутствуют объекты, проявляющие свою двойственность по-разному (т.е., принадлежащие различным наблюдательным типам, изучаемые различными группами исследователей и, в результате, имеющие весьма разные обозначения). Более того, объект, представляющий собой одиночный с точки зрения одной методики наблюдений, может оказаться двойным или кратным с точки зрения другой; это является следствием разницы в позиционной и фотометрической точности используемых каталогов (методик наблюдения), сказывается на присваиваемых идентификаторах и усложняет проблему КИ.

Для решения проблем КИ привлекается вся имеющаяся в каталогах информация, в первую очередь – позиционная и фотометрическая, а также уже содержащаяся в некоторых каталогах кросс-идентификация. При этом попутно решается (сама по себе весьма актуальная [8], в частности, в астрономии [9]) проблема достоверности (согласованности, непротиворечивости) каталогизируемых многомерных числовых измерительных данных, поступающих из разных источников. В частности, в процессе КИ нами обнаруживаются ошибки как в оригинальных каталогах, так и в базах данных общего назначения, о чем мы сообщаем их авторам.

Около 90 процентов всех проблем КИ удается разрешить автоматически, и это делает проблему создания унифицированных методов КИ в принципе решаемой. Для оставшихся 10 процентов все же требуется ручной подход – опять-таки, в сотрудничестве с авторами оригинальных каталогов.

Каталог ILB будет постоянно пополняться, станет основой для базы данных BDB, а также может служить для других приложений. Методика КИ компонентов, пар и систем двойных и кратных звезд также должна считаться оригинальной (как и система обозначений BSDB); она постоянно модифицируется и станет полезной для будущих астрономических обзоров.

Заключение

Проблема обозначения небесных объектов появилась в астрономии давно и окончательного решения не нашла до сих пор. Параллельное существование и интенсивное использование десятков систем обозначений приводит, кроме всего прочего, к необходимости постоянно решать проблемы кросс-идентификации. Особенно остро эти вопросы стоят для двойных и кратных систем – объектов, выглядящих, обозначаемых и каталогизируемых различными группами исследователей по-разному.

В работе дан обзор систем и стандартов идентификации астрономических объектов, описаны сложности и особенности для кратных объектов. Обсуждаются существующие методы и средства кросс-идентификации объектов. Описаны методика обозначений BSDB и общий каталог идентификаций двойных звезд ILB, применяемые в Базе данных двойных звезд BDB, а также иллюстрируются примеры разрешения конфликтов идентификации.

Благодарности

Мы благодарны анонимным рецензентам за ценные замечания, которые позволили улучшить текст статьи. Работа выполнена при поддержке грантов РФФИ 12-02-31904, 12-07-00528, и Программы Президиума РАН Поддержка ведущих научных школ (грант НШ-3620.2014.2)

Литература

- [1] VizieR database: <http://vizier.u-strasbg.fr/viz-bin/VizieR>
- [2] O'Meara S. J. The Messier objects. – Cambridge University Press, 1998. – P. 3. – 304 p.
- [3] Christen P. Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection. – Springer, 2012.
- [4] Christen P. A survey of indexing techniques for scalable record linkage and deduplication // IEEE Transactions on Knowledge and Data Engineering. – 2011.

- [5] Karpov S.V., Malkov O.Yu., Mironov A.V. 2012, *Astrophysical Bulletin*, 67, 82.
- [6] Malkov O.Yu., Kaygorodov P.V., Kovaleva D.A., Oblak E., Debray E. 2014, *Astronomical and Astrophysical Transactions*, 28, 235.
- [7] Sesame: <http://cds.u-strasbg.fr/cgi-bin/Sesame>
- [8] Ежела В.В., *данный сборник*
- [9] Авраменко А.Е. Концептуальные модели данных в отождествлении физических свойств пульсаров на вековом масштабе // Труды RCDL 2012. – CEUR, 2012. – Т. 934. – С. 245–251.

Problems of Designation and Cross-Identification of Multiple Objects in Astronomy

Dana A. Kovaleva, Pavel V. Kaygorodov,
Leonid A. Kalinichenko, Oleg Yu. Malkov,
Nikolay A. Skvortsov

In this work we continue and develop research focused on solving of problems arising in the comprehensive analysis of heterogeneous data sets by available arsenal of scientific methods and tools to identify useful information and to gain new knowledge. Here the problems of designation and cross-identification of multiple objects in astronomy are discussed, with a focus on streamlining of designation schemes and integrating of multidimensional observational data of different types. We shortly review existing methods of designation of single and multiple astronomical objects, describe BSDB schemes, implemented in Binary star database (BDB) and discuss problems and solutions of cross-identification.