

Задачи в области звездной астрономии и поддержка их данными

© О.Ю.Малков, Н.А.Скворцов, Е.А.Аввакумова, О.Б.Длужневская, А.А.Исаева,
П.В.Кайгородов, Л.А.Калиниченко, С.В.Карпов, Е.Ю.Кильпио, Д.А.Ковалева,
А.В.Миронов, С.Г.Сичевский, Д.А.Чулков

Институт астрономии РАН
Физический факультет МГУ им. Ломоносова
Кафедра астрономии и геодезии УрФУ
Астрокосмический центр ФИАН
Специальная астрофизическая обсерватория РАН
Институт проблем информатики РАН
Государственный астрономический институт им. Штернберга МГУ им. Ломоносова

malkov@inasan.ru

Аннотация

Звездная астрономия (называемая также галактической астрономией) – раздел астрономии, исследующий нашу Галактику и, в частности, строение, динамику образование и эволюцию входящих в нее звездных систем. Традиционно для решения большинства задач звездной астрономии привлекаются большие массивы наблюдательных и теоретических данных. В работе представлены примеры таких задач. Здесь описывается также Международная виртуальная обсерватория – коллекция интегрированных архивов астрономических данных и инструментов их обработки, которые создают среду для проведения научных исследований. Мы также обсуждаем вопросы, связанные с доступностью данных, их объемом, качеством, неоднородностью и достоверностью

1 Введение

Четвёртая парадигма научных исследований [3] сегодня прорисовывается всё отчетливее. В разные времена в основе развития науки лежали, во-первых, натуральный эксперимент и установление фактов,

во-вторых, развитие теорий и их проверка, в-третьих, применение вычислительных методов и моделирования. Четвёртая парадигма предполагает исследование открытых массивов научных данных больших объемов.

Происходит переход от поиска данных для решения задач к всестороннему анализу массивов данных имеющимся арсеналом научных методов и инструментов для выявления полезной и новой информации и знаний. Исследование начинается со сбора данных от инструментов наблюдения или моделирования, возможно не связанных с исследователем.

Добытые данные в процессе курирования подвергаются первичной обработке, структурированию и классификации, семантическому анализу в соответствии со знаниями о предметных областях исследования. Они снабжаются информацией об их происхождении, обрабатываются для получения вторичной информации методами, имеющимися в предметной области, и представляются в форме, эффективной для долговременного хранения, открытого доступа в исследовательских средах и дальнейшего анализа.

В условиях возрастания объёмов данных и требуемых манипуляций с ними стоимость и время разработки средств анализа данных становятся критически важными для самой возможности исследований. Для средств анализа становится важна доступность и простота применения всевозможных методов, алгоритмов и инструментов, автоматических и интерактивных, специфических для предметной области

исследования и широкой области применения, таких как реализации математических и статистических методов, средства визуального анализа. Взаимодействие научных коллективов определяется повторным использованием результатов, разработанных методов и средств анализа данных, что также подчёркивает необходимость их доступности. Для многопользовательского, многоэтапного анализа и анализа производных данных применяются магистральные средства и средства потоков работ. С их помощью организуется всесторонний анализ доступных и поступающих данных для выявления всей возможной полезной информации и знаний. Анализ больших объёмов данных или большое количество вычислений связано с необходимостью использования масштабируемых архитектур, обеспечивающих распределённость и параллелизм обработки.

Новая парадигма развития связывается с науками с интенсивным использованием данных, типичным представителем которых является астрономия. Здесь признаки смены парадигмы исследований демонстрирует всестороннее использование Виртуальной обсерватории, средства которой последовательно развиваются в течение последнего десятилетия. В статье проанализированы подходы к обеспечению научных исследований в области звёздной астрономии данными и средствами анализа, обеспечиваемые Виртуальной обсерваторией. В разделах 2 и 3 описаны тенденции, приведшие к необходимости создания средств Виртуальной обсерватории и международного альянса, сложившегося для её развития и эксплуатации в научных исследованиях. Раздел 4 описывает показательные задачи, решённые в области звёздной астрономии с использованием средств Виртуальной обсерватории, а также подход к решению общих задач, возникающих у большинства исследователей. Раздел 5 посвящён перспективам развития и актуальным проблемам Виртуальной обсерватории в связи с требованиями, предъявляемыми к ней быстро растущими потребностями в обработке колоссальных массивов научных данных и получения научных результатов при их использовании.

2 Астрономия – наука с интенсивным использованием данных

В связи со значительными достижениями в наблюдательных, регистрирующих и вычислительных технологиях астрономия в конце прошлого века столкнулась с лавинообразным увеличением количества получаемых данных. Эти наборы данных покрывают небо в различных диапазонах длин волн – от гамма- и рентгеновского диапазона, через ультрафиолетовый, оптический и инфракрасный до радиодиапазона. Для решения многих астрофизических задач накопленного сейчас

наблюдательного материала уже достаточно. С учётом появления недорогих технологий хранения данных и наличия высокоскоростных сетей концепция мульти-терабайтных, бесшовных интероперабельных баз данных уже не считается надуманной. Всё большее количество астрономических каталогов становятся взаимосвязанными, поисковые машины всё более и более усложняются, а результаты анализа таких данных оказываются столь же богатыми, как и для данных, полученных с реальных телескопов. Предназначенные для обзоров телескопы наземного и космического базирования могут поставлять изображение всего неба за несколько дней и производить, таким образом, данные в объёмах, измеряемых в петабайтах. Эти технологические достижения фундаментально изменяют характер астрономических исследований, а изменения, в свою очередь, радикально влияют на социологию самой астрономии.

За прошедшие годы концепция виртуальной обсерватории, призванной удовлетворять существующим требованиям к управлению данными, их анализу и распространению, завоевала широкую популярность. Это система, в которой распределённые по всему миру гигантские астрономические архивы и базы данных интегрированы в единую среду вместе с инструментами анализа и вычислительным сервисом. Современная астрономия стоит у границ новых открытий, возможности которых предоставляются современными информационными технологиями, а также политической и технической международной кооперацией.

Необходимо отметить, что виртуальная обсерватория является эффективным инструментом «демократизации» астрономии, поскольку даже небольшие научные, а также образовательные астрономические учреждения получают, по существу, те же возможности для проведения научных исследований на современном наблюдательном материале, что и ведущие астрономические организации мира.

Виртуальная обсерватория представляет собой механизм реализации концепции электронной науки в астрономии. Говоря упрощённо, увеличение научного выхода данных означает получение большего количества научных результатов (например, опубликованных статей или докладов, представленных на конференциях) с каждого гигабайта данных, приходящих с данного конкретного инструмента. В конечном счёте это означает увеличение влияния гигабайта данных на общее развитие науки – точно так же, как публикация научной работы, помимо журнала ещё и на сервере препринтов увеличивает её влияние и роль в отрасли.

В традиционном подходе к научной работе львиная доля времени и усилий исследователя затрачивается на разнообразные низкоуровневые операции: конвертацию форматов, редукцию

данных и подготовку их к научному анализу, поиск информации, написание сценариев автоматизированной обработки для многократных операций и другие действия. Их повторное использование зачастую ограничено или вовсе не представляется возможным. При этом собственно интеллектуальные усилия, анализ и интерпретация подготовленного научного материала, имеющие решающее значение во всём процессе, занимают несравнимо меньшее время. Миссия виртуальной обсерватории в широком смысле заключается в разделении творческого, интеллектуального, высокоуровневого процесса от низкоуровневых операций, которые должны либо происходить прозрачно (незаметно) для исследователя, либо быть максимально пригодными для повторного использования в ходе работы над последующими научными задачами. Таким образом, на первое место выходят абстрагированные от конкретных инструментов и наблюдателей данные, готовые к научному использованию, сервисы для доступа к ним и широкий набор соглашений для обеспечения прозрачности рутинных операций. Всем этим в той или иной степени занимается виртуальная обсерватория. Её успехи в данный момент и текущее положение дел уже несут в себе новые возможности для эффективных научных исследований, обходящихся без низкоуровневой составляющей.

3 Международная виртуальная обсерватория

К настоящему времени в мире функционируют 20 национальных и международных проектов виртуальных обсерваторий. Их осуществление не лимитируется национальными границами. В проектах используются наблюдательные данные в широком диапазоне длин волн, полученные с космических и наземных инструментов, используемых международным астрономическим сообществом. Цель каждого проекта – вооружить астрономов мощными механизмами современного исследования наблюдательных данных, механизмами, которые могут быть использованы и нашими коллегами из смежных наук.

Для того чтобы Международная виртуальная обсерватория (МВО) стала реальностью, необходимо определить её задачи и требующиеся для их достижения шаги как координируемые международные усилия. Каждый из существующих национальных проектов имеет собственные научные и технологические задачи. В то время как это разнообразие представляется достаточно благотворным для успеха МВО, имеются также некоторые общие элементы, которые необходимо согласовывать для эффективного конструирования МВО. Большинство таких элементов связано со стандартами данных и интерфейсов. Другой вид общих (или используемых совместно) элементов – это пакеты программ и библиотеки исходных текстов. В целесообразности международного

сотрудничества можно убедиться и на примере решения проблем безопасности, финансирования, политических вопросов. На достаточно ранних этапах создания МВО должны быть решены вопросы стандартов интероперабельности используемых источников данных. Чрезвычайно целесообразными для приобретения доверия широких слоёв астрономического сообщества представляются демонстрации новых возможностей виртуальных обсерваторий на различных уровнях сложности. Необходим открытый обмен информацией и опытом между различными проектами виртуальных обсерваторий. Некоторые цели на пути достижения прозрачности для сообщества создания МВО уже реализованы, другие – дело ближайшего будущего.

Виртуальную обсерваторию (ВО) иногда называют всемирной сетью (World Wide Web – WWW) для астрономов. Действительно, между всемирной сетью и сетью для астрономов существует значительное сходство. Неотделимая часть обеих концепций – ресурсы. В случае всемирной сети в качестве ресурсов выступают веб-сайты, порталы и директории, веб-сервисы. В Виртуальной обсерватории на их месте стоят архивы данных, сервисы доступа к каталогам (например, SDSS), астрономические веб-сервисы.

Ещё одним краеугольным камнем в основании Виртуальной обсерватории и всемирной сети являются инструменты доступа к ресурсам: 1) в WWW мы имеем веб-браузеры (например, internet explorer, firefox, safari), тогда как в ВО вместо них используются браузеры данных (например, astrogrid vo desktop, cds aladin, nvo datascope); 2) искушённые пользователи достаточно часто используют средства для работы с ресурсами из командной строки – curl или wget в WWW и аналогично stilts или astro-runtime в ВО; 3) существуют специализированные клиенты, которые используют WWW/ВО-протоколы в качестве инфраструктуры и/или передачи данных (например, picasa и google earth) и их аналоги в ВО (например, visivo).

Набор однородных стандартов и технологий, необходимый для успешного функционирования виртуальной обсерватории, разрабатывает, внедряет и поддерживает Альянс "Международная Виртуальная обсерватория" (International Virtual Observatory Alliance – IVOA, <http://ivoa.net>) – международный союз национальных виртуальных обсерваторий, созданный в 2001 г. и призванный координировать деятельность по стандартизации описания, поиска, доступа и публикации данных. Одним из основателей и активным участником Альянса является Российская виртуальная обсерватория [23].

Альянс «Международная всемирная обсерватория» играет ту же роль для ВО, что и Консорциум всемирной сети (World Wide Web Consortium) для WWW: это административные организации, ответственные за разработку и внедрение стандартов обмена различными видами

информации между сторонами-участниками соответствующих процессов. В качестве примеров в данном случае можно привести спецификации HTML/XHTML, разработанные W3C, и стандарт VOTable, разработанный IVOA.

В последние годы Виртуальная обсерватория достигла существенного прогресса. Рабочими группами альянса разработан достаточно широкий набор стандартов, которые охватывают форматы данных (VOTable), описание метаданных ресурсов (Resource Metadata), модель данных для одномерных спектров (Spectrum Data Model) и значительно более сложную и ёмкую Characterisation Data Model, язык запросов к данным (ADQL), протоколы доступа к спектрам и изображениям (SIAP, SSAP), протокол коммуникаций между различными приложениями на рабочей станции пользователя (SAMP), механизмы аутентификации и авторизации и др. На разных стадиях разработки находится ещё ряд стандартов. Сейчас в Виртуальной обсерватории стало возможным работать даже с чрезвычайно сложными наборами данных, например, 3D-спектроскопией и результатами N-body симуляций. В то же время разработчики приложений создали впечатляющий набор ВО-инструментов, начиная с самых общих и заканчивая узкоспециализированными клиентскими средствами.

Поставщики данных и сервисов внесли свой вклад в Виртуальную обсерваторию, предоставляя доступ к огромным коллекциям и архивам данных от радио- до гамма-диапазона длин волн. Совсем недавно стали появляться сервисы доступа к теоретическим моделям (например, коллекция теоретических спектров звёздных атмосфер в Испанской виртуальной обсерватории или синтетические модели звёздных популяций PEGASE во Французской виртуальной обсерватории, доступ к результатам космологических симуляций в Итальянской виртуальной обсерватории). Нельзя не упомянуть и первые прототипы сервисов для анализа данных и научно значимых сервисов, ассоциированных с архивами данных, такой как моделирование спектрофотометрических свойств взаимодействующих галактик в базе данных GalMer.

После нескольких лет интенсивного технологического развития ресурсы МВО достигли уровня зрелости, достаточного для их каждодневного использования в научных исследованиях. МВО находится на той стадии развития, когда астрономы начинают использовать её в повседневной научной работе. В ближайшее время появятся работы, эксплуатирующие МВО на качественно более сложном уровне, с применением тонких и общих моделей данных и технологических достижений.

Таким образом, международная виртуальная обсерватория представляет собой мощную виртуальную среду, предназначенную для увеличения возможностей астрономических

исследований и научного выхода данных. Виртуальная обсерватория интегрирует в единую систему гигантские астрономические архивы и базы данных, распределенные по всему миру, а также инструменты анализа данных и вычислительный сервис, используя при этом набор однородных стандартов и технологий. Международная виртуальная обсерватория включает в себя все значительные национальные и международные проекты по созданию виртуальных обсерваторий, основная цель которых – объединить существующие архивы наземных и космических инструментов и обеспечить исследователям и общественности удобный доступ к ним. Эта задача представляется чрезвычайно важной не только из-за колоссального объема астрономических данных, но и их спектрального разнообразия (от рентгена до радио). Каждый спектральный диапазон предоставляет свою, уникальную информацию о небесном объекте или явлении; при этом требуется специализированная экспертиза для правильной интерпретации. Вся эта информация также интегрируется в Международной виртуальной обсерватории и позволяет синтезировать данные, чтобы использовать их в конкретных научных приложениях.

4 Примеры задач звездной астрономии

Проект "Международная виртуальная обсерватория" появился более десяти лет назад, и за это время успешно проэволюционировал от чисто демонстрационного уровня к набору серьезных инструментов для анализа астрономических данных и извлечения реальных научных результатов. Первыми научными успехами Виртуальной обсерватории были поиск коричневых карликов в обзорах 2MASS и SDSS [5] и идентификация большого числа оптически слабых квазаров второго типа [17]. Для решения этих задач применялись инструменты кросс-отождествления объектов в больших обзорах по позиционной и фотометрической информации. Некоторые более современные достижения Виртуальной обсерватории в области звездной астрономии обсуждаются ниже. В конце раздела даются также некоторые практические рекомендации по использованию и созданию инструментов ВО.

4.1 Рассеянные скопления

Большие многоцветные фотометрические обзоры, такие как Two Micron All Sky Survey (2MASS), UKIRT Infrared Deep Sky Survey (UKIDSS), INT Photometrical H-alpha Survey (IPHAS), в изобилии содержат данные, которые могут быть использованы для изучения подсистемы рассеянных скоплений Галактики. Типичный набор инструментов исследования включает в себя диаграммы цвет-светимость, диаграммы звездной плотности, а также их комбинации в

многопараметрическом пространстве координат. Виртуальная обсерватория предоставляет удобное средство для работы с такими данными, а именно протокол ConeSearch, который по своему замыслу абстрагирован от деталей организации конкретного каталога. Он позволяет использовать приложения для однородного и одновременного анализа нескольких каталогов, или, например, неоднократно осуществлять произвольные запросы на выборку по координатам из одного (в том числе, всенебесного) каталога.

Одновременно аппроксимируя изохроны и анализируя пространственное распределение звезд, отобранных по их цветам и блескам (см. детали метода в [22][11]), в работе [7] был сконструирован крупнейший каталог, содержащий около 200 новых рассеянных скоплений с однородными параметрами, полученными из 2MASS и UKIDSS. Каталог был опубликован как ВО-ресурс {<http://ocl.sai.msu.ru>} [21], с помощью которого любой пользователь может скомпилировать собственный каталог, содержащий, например, собственные движения или более глубокую фотометрию, с использованием собственных данных.

Мощным инструментом для открытия новых рассеянных скоплений служит также оптическая и инфракрасная фотометрия и позиционная информация из фотометрико-кинематических оптических обзоров Hipparcos, ASCC-2.5, PPMXL. Так, в [8] были представлены данные о 109 новых скоплениях.

4.2 Кросс-матчинг больших обзоров и параметризация звезд

Цветовые диаграммы могут служить в качестве мощного инструмента для определения распределения энергии в спектрах звезд и, следовательно, для параметризации звезд. Эта задача неразрывно связана с проблемой определения межзвездного поглощения: распределение энергии должно быть скорректировано за покраснение света. Недавно сконструированные большие астрофотометрические обзоры (2MASS, GALEX, SDSS, DENIS, UKIDSS и др.) и ВО-инструменты для кросс-матчинга объектов в этих обзорах дают возможность получить многоцветную (до 10-15 показателей цвета) фотометрию для миллионов объектов [9]. На основании этой информации могут быть построены цветовые диаграммы, позволяющие классифицировать и параметризовать объекты. Для корректной параметризации используются, помимо каталогизированных данных, кривые реакции исследуемых обзоров, закон межзвездного поглощения и эмпирические распределения энергии в спектрах звезд различных спектральных типов. Проблемы и предварительные результаты конструирования многоцветных диаграмм и параметризации объектов обсуждаются в [13][19].

4.3 Затменные двойные и База данных двойных звезд

Инструменты Виртуальной обсерватории использовались для исследования затменных двойных. Новый каталог затменных двойных был создан в [16]. В работе [14] каталог использовался для создания наиболее полного набора правил для классификации затменных двойных и, с использованием ВО-инструментов, для создания процедуры классификации даже для случая неполного набора данных. Разработанная процедура была применена к большим спискам затменных двойных, в том числе, полученным из наблюдений эффектов микролинзирования. Около 5300 систем были классифицированы впервые. Эти данные могут быть использованы для последующего определения астрофизических параметров компонентов этих двойных. Новая версия каталога [4] содержит параметры и морфологические типы кривых блеска для 7200 систем, также, по возможности, дана спектральная классификация. В частности, в каталог включены недавно опубликованная информация о классификации 1352 систем. Таким образом, каталог представляет собой крупнейший список затменных двойных, классифицированных из наблюдений.

Этот, а также другие основные каталоги двойных звезд различных типов загружаются в Базу данных двойных звезд (Binary star DataBase, BDB). Создание каталогов / баз данных двойных звезд имеет длинную историю, но публиковались преимущественно, каталоги двойных одного определенного типа: визуальные, интерферометрические, спектроскопические и пр. Однако, детальное исследование звездных популяций не может быть ограничено одним типом двойных: так, для определения истории звездообразования необходимо анализировать данные для всех типов двойных – от самых широких (пары с общим собственным движением) до самых тесных (спектроскопические, затменные, рентгеновские с вырожденными компонентами). Также, параметризация компонентов и орбит двойных систем не может ограничиваться одним типом двойных. Таким образом, существующие и распределенные по различным центрам данных сведения о разных типах двойных должны быть объединены. Решением этой проблемы призвана служить BDB, создаваемая сначала в Обсерватории Безансона (Франция) и переданная в последствие для модернизации в Институт астрономии РАН. Основные цели и идеи BDB обсуждаются в [15], текущая версия базы данных описана в [10].

4.4 Молодые звезды и коричневые карлики

Исследование собственных движений объектов в облаках созвездия Lyrus с помощью инструментов ВО были проведены в работе [12]. Используя доступные кинематические данные, авторы протестировали на членство в объекты облаков Lyrus звезды, найденные в программе наблюдений телескопа Spitzer и ряда оптических обзоров. Были также исследованы соотношения между собственными движениями и другими свойствами объектов с целью получения информации об их образовании и ранней эволюции. Авторы скомпилировали список возможных членов облаков Lyrus, содержащий всю доступную информацию об их спектральной классификации и физических параметрах. ВО-инструменты использовались для кросс-матчинга объектов этого списка с доступной астрометрической каталогизированной информацией с целью вычисления собственных движений. Финальный набор содержит источники с блеском $I < 16$ mag и массой (по оценкам) более 0.1 масс Солнца.

В соответствии с кинематической информацией источники можно поделить на две основные группы. Первая содержит объекты с большими собственными движениями, согласующимися с данными для других населений Пояса Гулда. Кроме того, их пространственное распределение, а также оптические и ближние-ИК показатели цвета согласуются с данными, имеющимися для объектов, заведомо принадлежащих к облакам Lyrus. Во вторую группу входят объекты с небольшими и случайно ориентированными собственными движениями, расположенные преимущественно за пределами центральных зон облаков Lyrus. В работе был сделан вывод о принадлежности членов этой группы к объектам фона, не связанных с облаками Lyrus. В то же время члены первой, высокоскоростной группы, представляют объекты населения облаков Lyrus.

Результаты исследования суб-звездных объектов инструментами Виртуальной обсерватории были опубликованы в [6]. Здесь исследовались объекты, расположенные в 45' областях, окружающих две молодые массивные звездные системы в Поясе Ориона. Эти области являются достаточно молодыми (моложе 10 миллионов лет), имеют низкое поглощение и расположены недалеко от молодого рассеянного звездного скопления σ Orionis, богатого коричневыми карликами и объектами планетных масс.

Авторы использовали инструменты ВО, астрометрические обзоры Tycho-2, DENIS, 2MASS, 10 контрольных площадок на тех же галактических широтах, а также рентгеновские, инфракрасные и спектроскопические данные из литературы. Был скомпилирован исчерпывающий список известных молодых звезд и новых кандидатов в члены ассоциации Ori OB1b, а также объектов переднего и заднего фона. 136 из исследованных звезд

продемонстрировали свойства чрезвычайно молодых объектов: ранний спектральный тип, наличие лития или избытка излучения в среднем-ИК. Еще два молодых коричневых карлика и 289 звезд-кандидатов были идентифицированы с помощью оптических/инфракрасных диаграмм цвет-светимость. Авторы сделали заключение, что обе исследуемые области являются аналогами скопления σ Orionis, но при этом более массивны, более протяженны, несколько более старые и в меньшей степени радиально концентрированы.

4.5 Внегалактические исследования

Возможности Виртуальной обсерватории также могут быть использованы для внегалактических исследований. Так, в [20] авторы используют возможности ВО-инструментов для исследования AGB-звезд высокой светимости ($M_{bol} < -6.0$ mag) с показателями цвета $J-K_s > 1.5$ mag и $H-K_s > 0.4$ mag расположенных в LMC, SMC, M 31, и M 33. При этом используются данные из обзора 2MASS All-Sky Point Source. Авторы, в частности, скомпилировали каталоги наиболее ярких AGB-звезд и обнаружили, что если в LMC такие звезды расположены преимущественно в центре, то в M31 они концентрируются в кольце активного звездообразования. Количество исследованных звезд в SMC и M31 недостаточно представительно для того, чтобы делать окончательные выводы об их пространственном расположении, хотя тенденция концентрироваться к центру определенно наблюдается. Результирующие функции светимости сходны для всех четырех изученных галактик.

Объекты с указанными показателями цвета соответствуют, преимущественно, углеродным звездам. Хотя среди обнаруженных авторами звезд содержится большое число углеродных, они не были ранее включены в каталоги, являющиеся продуктом оптических исследований, поскольку эти звезды погружены в пылевые оболочки.

Приведенные выше примеры демонстрируют достаточную зрелость ВО-инструментов для решения актуальных задач звездной астрономии и получения новых результатов с помощью использования астрономических данных. При этом применяются как классические методы анализа данных, так и современные интероперабельные ВО-механизмы для идентификации объектов и извлечения «мультиволновой» информации

По данным библиографической базы данных ADS 1900 публикаций (на август 2013 г.) упоминают Virtual Observatory в аннотации, и около 210 из них являются полноценными астрономическими исследованиями, опубликованными в международных рецензируемых журналах. Эту статистику можно уточнить, воспользовавшись сервисом «Some recent

papers about VO-enabled science» по адресу <http://ivoa.net/newsletter/006/>. Первые диссертации по результатам исследований в помощь инструментов ВО были защищены в России в 2009 (кандидатская, И. Золотухин) и 2010 (докторская, И. Чилингарян) г.г.

4.6 Рекомендации по созданию и использованию инструментов Виртуальной обсерватории

В заключении данного раздела дадим сведения о том, как пользоваться инструментами ВО и как принять участие в создании ВО.

Типичной задачей для ВО является получение всех сведений об одном объекте. В общем случае рекомендуется обратиться к сервису Aladin (<http://aladin.u-strasbg.fr/aladin.gml>) и указать название объекта в окне Location. При этом загрузится изображение из DSS (<http://stdu.stsci.edu/dss/>) и объекты из баз данных SIMBAD (<http://simbad.u-strasbg.fr/simbad/>) и NED (<http://ned.ipac.caltech.edu/>). В меню File рекомендуется выбрать опцию Load from the Virtual Observatory, задать радиус и координаты, а также указать желаемые типы данных (спектры, каталоги, и т.п.). Aladin будет последовательно опрашивать сервисы ВО. Для конкретных типов объектов можно также использовать специализированные базы данных: для звезд – SIMBAD, для скоплений – WEBDA (<http://www.univie.ac.at/webda/>), для внегалактических объектов – NED. Для двойных звезд в скором будущем начнет функционировать база данных BDB. Для получения сведений об ансамбле объектов в указанных сервисах есть режимы пакетного доступа. Для всех типов объектов полезным является обращение к библиографической базе данных ADS (<http://adsabs.harvard.edu/>), которая для кросс-идентификации объектов использует системы типа “Name resolver” в SIMBAD и NED.

Другой часто встречающейся задачей является получение всех сведений о наблюдательных архивах. Для этого рекомендуется обратиться к регистрам (реестрам) архивов, самыми полными из которых являются Euro-VO registry (<http://registry.euro-vo.org/search.jsp>) и NVO directory (<http://nvo.stsci.edu/vor10/index.aspx>).

Для того, чтобы предоставить международному астрономическому сообществу вновь созданный каталог (т.е., чтобы он был включен в ресурсы ВО), от автора требуется опубликовать каталог или его описание в рецензируемом журнале, подготовить в формате, пригодном для импорта в Vizier и отправить его администраторам этого сервиса (<http://vizier.u-strasbg.fr/>).

Общего ответа на вопрос о том, как предоставить международному астрономическому сообществу новый архив наблюдений или базу данных пока нет. В качестве одной из возможностей можно использовать инструмент для создания баз данных

Astronomical Database Generator SAADA (<http://saada.u-strasbg.fr/saada/>).

В сложных случаях для решения проблем, связанных с использованием и созданием ВО, рекомендуется пока обращаться к экспертам в области ВО. Полезную информацию можно также найти на сайтах Альянса «Международная виртуальная обсерватория» (<http://www.ivoa.net>) и Российской виртуальной обсерватории (<http://www.inasan.rssi.ru/rus/rvo>).

5 Актуальные проблемы виртуальных обсерваторий

Существенное развитие инструментов наблюдения для получения данных о звёздах, открытость данных и сплочение исследователей в сообщества диктуют новые требования к инфраструктуре исследований в астрономии. Виртуальная обсерватория уже имеет свои плоды, но для решения предстоящих проблем она должна развиваться в определённых направлениях.

Проблемой, которая, как будет видно ниже, несёт за собой также много последствий, является неоднородность накопленных до сих пор данных. Каждый обзор неба и каждый из множества каталогов создавались разными людьми или исследовательскими группами для решения определённых задач, поэтому состав и структура данных в каталогах бывают ограничены теми целями, которые преследовались их создателями. Каталоги имеют неоднородную структуру и различную семантику полей и значений, множество важных, но никак не формализованных комментариев. Совместная обработка данных на стыке каталогов становится не автоматизируемой и требует программирования. Использование в Виртуальной обсерватории каталогов в их первоначальном виде оправдывается целью сохранения уникальных и не теряющих актуальности наблюдений в астрономии, однако в результате приходится решать одни и те же проблемы неоднородности множество раз от задачи к задаче. Неоднородность данных является первейшей проблемой, отнимающей время и ресурсы при разработке научных приложений.

Работа по сопоставлению каталогов в рамках виртуальной обсерватории ведется на уровне создания таблиц кросс-идентификации. Примерами такого подхода могут служить популярная база астрономических данных Simbad (<http://simbad.u-strasbg.fr>) и ее прототипы. Также создаются системы с ограниченными наборами интегрированных каталогов [10]. И в том, и в другом случае установление взаимосвязи является кропотливой работой. Различия в семантике данных порой столь неочевидны, что могут выявляться со временем при обнаружении противоречивости результатов, появлении неверных связей.

С возрастом количества каталогов, исследуемых в них параметров, разновидностей

решаемых задач и интенсивности использования данных различного происхождения при их решении надрезает необходимость в семантических подходах к работе с каталогами и к решению научных задач. Исследовательские группы должны взаимодействовать в рамках сообществ с целью разработки спецификаций предметных областей, в которых они работают. Разрабатываются онтологии предметных областей, знаниям которых берут на себя обязательство следовать члены сообщества. Стандартизируются концептуальные схемы для представления данных в приложениях сначала по общим разделам знаний, используемым практически во всей астрономии, затем более специальным.

В рамках Международной виртуальной обсерватории наработки в области семантического моделирования предметных областей присутствуют, в частности онтология астрономических объектов, ряд схем (называемых моделями данных) для представления разного рода данных, например, Phot DM для представления фотометрических данных, VO Event для описания физических явлений и процессов и другие [2]. Часть этих спецификаций уже активно используется астрономическим сообществом, а часть ещё требует существенного развития.

Детальное описание знаний предметных областей в астрономии позволит определять в их терминах спецификации имеющихся и новых каталогов, отображать их с учётом семантики в стандартизованные представления. Это основной подход для решения проблемы неоднородности астрономических данных, как и многих других проблем. Существующие подходы недостаточны для решения подобных проблем. В частности, семантические дескрипторы UCD [1] обладают слабой выразительной мощностью для исключения неоднозначности описания.

Важной информацией о семантике данных, без которой сложно судить о применимости данных или определённых методов их анализ при решении задач, является качество данных. Нефункциональные требования к данным должны учитывать их точность, полноту, достоверность. О каталогах необходимо учитывать информацию о типах представленных в них объектах, охвате таких объектов, точности оборудования, которое использовалось при обзоре. Данные каталогов и данные на всех этапах обработки должны сопровождаться информацией об их происхождении. При обилии источников данных и применяемых к ним методов, при многоэтапной обработке такая информация становится обязательной. На сегодня в каталогах обычно предоставляются данные о погрешностях измерений и результатов вычислений, флаги уверенности в определённых значениях, а информация о происхождении данных иногда представлена в текстовых комментариях к полям каталогов или к отдельным их записям. Автоматически учитывать

их не представляется возможным. Прежде они требуют структуризации.

Ещё одной стороной, требующей существенного развития в Виртуальной обсерватории, является накопление общедоступных реализаций методов вычислений и анализа данных, характерных как для астрономии в целом, так и в конкретных областях исследования. Помимо исследуемых данных астрономы нуждаются в средствах их анализа. Необходимо создавать коллекции универсальных настраиваемых сервисов данных, независимых от источников данных, реализующих методы анализа. Общедоступными и легко настраиваемыми на любые доступные данные, на пересечение различных массивов данных и параметров должны быть средства статистического и визуального анализа, в частности, создания диаграмм. Необходимы мощные средства редукции данных, поиска похожих объектов, поиска отличающихся объектов (ошибочных или необычных) и решения других универсальных задач. Необходимы более узконаправленные средства, связанные с конкретными методами, законами в астрономии. Реализуемые методы должны учитывать информацию о качестве, точности, полноте и происхождении данных, и предоставлять подобную информацию о результатах, полученных при работе методов. Для того, чтобы не теряться в коллекции методов ВО, и с целью поддержки независимости их от конкретных источников данных, спецификации методов должны быть связаны со спецификациями предметных областей и курироваться сообществами, работающими в соответствующих областях.

Некоторые системы, предоставляющие доступ к данным каталогов, в частности, упомянутый выше каталог рассеянных скоплений [7], имеют в своих интерфейсах наборы сервисов данных для вычисления различных параметров, и количество их реализаций растёт. Подобные сервисы привязаны не к спецификациям предметной области, а к ограниченному числу каталогов, предоставляемых системой. Применение их к другим данным невозможно. Созданию универсальных реализаций методов в астрономии препятствует также неоднородность представления данных, над которыми они должны выполняться.

Показательной является также проблема кросс-матчинга и кросс-идентификации астрономических объектов в плане реализации методов для её решения. Один и тот же астрономический объект имеет множество разных имён, данных разными сообществами и исследователями. Частично проблема их отождествления решается с помощью сервисов "Name resolver". Множество каталогов имеют свою собственную систему идентификации записей об объектах. Неоднородность идентификаторов рождает ошибки идентификации. Для кросс-идентификации объектов из ограниченного числа каталогов, то есть для отождествления объектов по идентификаторам, присутствующим в каталогах, построены таблицы

кросс-идентификации, не применимые к произвольным каталогам. Например, так построены таблицы кросс-идентификации в Общем каталоге переменных звезд GCVS [18]. Инструменты кросс-матчинга, то есть реализации методов отождествления объектов по астрометрическим, фотометрическим и другим данным о них, также привязаны к определённым парам каталогов. Американский сервис OpenSkyQuery (<http://www.openskyquery.net>), российский сервис SAI CAS Crossmatch (<http://vo.astronet.ru/cas/crossmatch.php>), немецкий проект Multi-Catalogue Multi-Cone Search (<http://www.g-vo.org/mcmcs/>) – все эти средства отождествления объектов работают с ограниченным числом каталогов. Поэтому при решении задач над множеством каталогов исследователь неизбежно сталкивается с проблемой отождествления астрономических объектов [24] и тратит на её решение внушительную часть времени, выделенного на решение задач. Разработка универсальных параметризуемых средств отождествления звёзд осложняется неоднородностью данных каталогов, до сих пор не решена, и имеет смысл, если все каталоги делаются изначально по стандартам ВО.

Помимо прочего, в Виртуальной обсерватории ожидается существенное развитие технологий, связанных со скоростью обработки данных при анализе. Исходные каталоги обычно используют файловое хранение с записями фиксированной длины (это делается традиционно для удобства обработки подручными средствами, и такая структура привычна для астрономов). В таком представлении нет индексации, нет предварительного агрегирования больших массивов данных, хранение данных не оптимизировано. Результатом этого является медленная работа обрабатывающих их приложений. Задача эффективного представления данных требует решения.

Используемые подходы к решению задач часто требуют физического перемещения внушительных массивов данных через интернет или на носителях. При решении некоторых задач требуется множественный прогон алгоритмов и методов. Некоторые каталоги интенсивно пополняются данными от версии к версии, ожидается появление обзоров колоссальных объёмов. Задачи, требующие много ресурсов, попросту не формулируются, отбрасываются изначально как нерешаемые. По этим причинам невозможно не обращать внимания на активно исследуемые и развиваемые сегодня горизонтально масштабируемые архитектуры. В Виртуальной обсерватории необходимо использовать при автоматической обработке и предоставлять в пользование исследователям средства распределённого хранения и параллельной обработки на основе технологий больших данных и облачных технологий. Возможна организация архитектуры виртуальной обсерватории таким

образом, чтобы исследователь, использующий предоставленные ею методы и средства анализа данных, не задумывался о распараллеливании обработки, реализации методов агрегирования в горизонтально масштабируемой среде и других технологиях, напрямую не связанных с его астрономическими исследованиями.

Заключение

Средства Виртуальной обсерватории в области звёздной астрономии как науки с интенсивным использованием данных стали её неотъемлемой частью и двигателем развития. Об этом говорит их использование при решении множества научных задач в этой области. В приведённых примерах задач, решённых в среде Виртуальной обсерватории, использовались совместно данные из разных астрономических каталогов. Открытость и обилие научных данных наблюдения, колоссальный рост объёмов данных и их обработки предъявляют новые требования к данным и к Виртуальной обсерватории, являющейся их источником для научных исследований.

Благодарности

Авторы благодарят И.Золотухина за помощь в подготовке материала. Работа выполнена при поддержке грантов РФФИ 12-02-31904, 12-07-00528, при поддержке Федерального агентства по науке и инновациям (грант 02.740.11.0247), Программы Президиума РАН Поддержка ведущих научных школ (грант НШ-3602.2012.2), а также в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009 - 2013 годы.

Литература

- [1] An IVOA Standard for Unified Content Descriptors. Version 1.10. IVOA Recommendation 19 August 2005. – IVOA Documents. – 2005. – URL: <http://www.ivoa.net/documents/latest/UCD.html>
- [2] Introduction to VO Concepts. – URL: http://www.ivoa.net/deployers/intro_to_vo_concepts.html
- [3] The 4th Paradigm Book. Data-Intensive Scientific Discovery. – Microsoft Research, 2009. – URL: <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>
- [4] Avvakumova E.A., Malkov O.Yu., Kniazev A.Yu., 2013, *Baltic Astronomy*, accepted
- [5] Berriman B., Kirkpatrick D., Hanisch R. et al. 2003, in *Large Telescopes and Virtual Observatory: Visions for the Future*, 25th meeting of the IAU, Joint Discussion 8, Sydney, Jul 2003, p. 60
- [6] Caballero J. A., Solano E. 2008, *A&A* 485, 931
- [7] Glushkova E., Koposov S., Zolotukhin I. et al. 2010, *AstL*, 36, 75

- [8] Kharchenko, N.V., Piskunov, A.E., Roser, S., Schilbach, E., and Scholz, R.-D. 2005, *A&A* 440, 403
- [9] Karpov S.V., Mironov A.V., Malkov O.Yu. 2012, *Astrophysical Bulletin*, 67, 82
- [10] Kaygorodov P., Debray B., Kolesnikov N. et al. 2012, *Baltic Astronomy*, 21, 309
- [11] Kopusov S., Glushkova E., Zolotukhin I. 2008, *A&A* 486, 771
- [12] Lopez Marti B., Jimenez-Esteban F., Solano, E. 2011, *A&A* 529A, 108
- [13] Malkov O.Yu., Mironov A.V., Sichevskij S.G. 2011, *Astrophys. Space Science*, 335, 105
- [14] Malkov O., Oblak E., Avvakumova E.A., Torra J. 2007, *A&A* 465, 549
- [15] Malkov O., Oblak E., Debray B. 2009, in *ADASS XVIII Conference*, eds. David Bohlender, Daniel Durand, Patrick Dowler, Quebec, Nov 2008, ASP Conf. Ser., 411, 442
- [16] Malkov O., Oblak E., Snegireva E.A., Torra J. 2006, *A&A* 446, 785
- [17] Padovani P., Allen M. G., Rosati P., Walton N. A. 2004, *A&A* 424, 545
- [18] Samus N., Durlevich O. et al. General Catalog of Variable Stars. – VizieR Online Data Catalog, 1, 2025. – 2009.
- [19] Sichevskij S. G. 2012, *Astronomy Reports*, 56, 710
- [20] Tsalmanza P., Kontizas E., Cambresy L. et al. 2006, *A&A* 447, 89
- [21] Zolotukhin I. 2010, in *JENAM 2010*, Lisbon, Sep 2010, 229
- [22] Zolotukhin I., Koposov S., Glushkova E. 2006, in *ADASS XV Conference*, eds. Carlos Gabriel, Christophe Arviset, Daniel Ponz, Enrique Solano, San Lorenzo de El Escorial, Oct 2005, ASP Conf. Ser., 351, 240, arXiv:astro-ph/0601691
- [23] Длужневская О. Б., Малков О. Ю. *Астрономический журнал*. 2005. Т. 82. 1148
- [24] Карпов С. В., Малков О. Ю., Миронов А. В. Кросс-отождествление больших обзоров для определения межзвёздного поглощения. // *Астрофизический бюллетень*. Том 67, №1. – 2012. – С. 87–94.

Provision of Data for Solution of Problems in Stellar Astronomy

Oleg Yu. Malkov, Nikolay A. Skvortsov, Ekaterina A. Avvakumova, Olga B. Dluzhnevskaya, Anastasia A. Isaeva, Pavel V. Kaygorodov, Leonid A. Kalinichenko, Sergej V. Karpov, Elena Yu. Kilpio, Dana A. Kovaleva, Aleksej V. Mirnov, Sergej G. Sichevskij, Dmitry A. Chulkov

Stellar astronomy (or galactic astronomy) is the study of our Galaxy and all its contents. In particular, stellar astronomy investigates structure, dynamics, formation and evolution of stellar systems. Traditionally, for solutions of the most problems in stellar astronomy a large amount of observational and theoretical data should be provided. Examples of such problems are presented here. International Virtual Observatory, a collection of integrated astronomical data archives and software tools that utilize computer networks to create an environment in which research can be conducted, is described in the paper. We also discuss issues related to data availability, their volume, quality, heterogeneity and reliability.