

УДК 52-1; 520.8:519.68

## КОНЦЕПТУАЛЬНЫЙ ПОДХОД К РЕШЕНИЮ ЗАДАЧ В АСТРОНОМИИ

© 2016 Н. А. Скворцов<sup>1\*</sup>, Е. А. Аввакумова<sup>2</sup>, Д. О. Брюхов<sup>1</sup>, А. Е. Вовченко<sup>1</sup>,  
А. А. Вольнова<sup>3</sup>, О. Б. Длужневская<sup>4</sup>, П. В. Кайгородов<sup>4</sup>, Л. А. Калиниченко<sup>1</sup>,  
А. Ю. Князев<sup>5,6,7</sup>, Д. А. Ковалева<sup>4</sup>, О. Ю. Малков<sup>4</sup>, А. С. Позаненко<sup>3,8</sup>, С. А. Ступников<sup>1</sup>

<sup>1</sup>Институт проблем информатики РАН, Москва, 119333 Россия

<sup>2</sup>Уральский федеральный университет, Екатеринбург, 620002 Россия

<sup>3</sup>Институт космических исследований РАН, Москва, 117997 Россия

<sup>4</sup>Институт астрономии РАН, Москва, 119017 Россия

<sup>5</sup>Южноафриканская астрономическая обсерватория, Кейптаун, 7925 ЮАР

<sup>6</sup>Большой южноафриканский телескоп, Сазерленд, 6920 ЮАР

<sup>7</sup>Государственный астрономический институт им. П. К. Штернберга Московского  
государственного университета им. М. В. Ломоносова, Москва, 119991 Россия

<sup>8</sup>Национальный исследовательский ядерный университет «МИФИ», Москва, 115409 Россия

Поступила в редакцию 26 марта 2014 года; принята в печать 30 ноября 2015 года

Появление новых технических возможностей привело к стремительному росту объема информации, получаемой астрономами из наблюдений в разных спектральных диапазонах на различных инструментах. Для результативной работы с этими данными представляется перспективным концептуальный подход. Он использует знания предметной области, позволяя формулировать решение задач и разрабатывать методы анализа информации в терминах понятий предметной области безотносительно конкретных источников данных и, таким образом, обеспечивая решение определенных задач в общем виде. Достоинства применения концептуального подхода продемонстрированы при решении задач, связанных с поиском кандидатов во вторичные фотометрические стандарты, поиском фотометрического красного смещения галактик, созданием хранилища информации о двойных и кратных звездах из неоднородных баз данных, классификацией затменных двойных. Эти задачи сформулированы и решаются над спецификациями используемых разделов знаний в астрономии, таких как фотометрические системы, астрономические объекты, кратные звездные системы и другие, и определены в терминах соответствующих областей независимо от существующих ресурсов данных.

Ключевые слова: *методы: анализ данных — астрономические базы данных — каталоги — инструменты виртуальной обсерватории*

### 1. ВВЕДЕНИЕ

С началом «цифровой эры» большинство естественных наук столкнулись с необходимостью не просто обрабатывать резко возросший объем информации при решении задач, но и использовать для всестороннего исследования всю совокупность накопленных и поступающих фактов. В последнее время в науках с интенсивным использованием данных происходит смена парадигмы исследований [1]. Во главу угла ставятся развитые средства и специализированные методы всестороннего изучения данных, а также среды поддержки исследований, обеспечивающие доступность информации

и методов ее анализа, составление сценариев обработки данных и экспериментов и возможность повторного использования реализаций методов и результатов исследований.

Астрономия переживает сегодня лавинообразный рост количества и качества анализируемой информации. Однако применяемые до сих пор подходы к исследованиям отражают, скорее, достигнутые успехи в организации открытых ресурсов и слабо используют семантику и автоматизацию обработки данных. На протяжении ряда лет достаточно удобным приемом оказывалось применение простых средств программирования (таких как языки сценариев), обращающихся к файлам оригинальных ресурсов, содержащим структурирован-

\*E-mail: nskv@mail.ru

ные записи. Очевидным недостатком такого подхода является зависимость реализации решения задач от структур и специфики конкретных ресурсов, что делает невозможным повторное использование реализованных методов другими коллективами и с другими ресурсами данных.

В последние годы все более очевидными становятся признаки перехода к новой парадигме. Для некоторых приложений характерные для неё подходы востребованы давно. Необходимость создания унифицированных общедоступных методов обработки данных появилась в астрономии одновременно с началом широкой компьютеризации. В первую очередь, такие системы развивались в тех областях, где вся информация подвергается достаточно типичной обработке. Примером могут служить пакеты редукции изображений MIDAS,<sup>1</sup> IRAF,<sup>2</sup> AIPS++.<sup>3</sup> Изначально они разрабатывались коллективами для обработки, в первую очередь, собственных архивов данных, но активно используются астрономами до сих пор и включают стандартные наборы операций: визуализация изображения, выделение его фрагмента, вычитание изображений и пр.

Широко используются программные сервисы, в которых есть возможность реализовать *потоки работ*<sup>4</sup> в виде сценариев на внутренних языках систем. По такому принципу реализованы онлайн-сервис Aladin,<sup>5</sup> интерактивный атлас неба для работы с данными Страсбургского центра данных и наложения графики на снимки участков неба, а также свободно распространяемый продукт Topcat,<sup>6</sup> предназначенный для работы астрономов с каталогами и таблицами.

В настоящей статье дается обоснование использования концептуального подхода к работе с астрономическими данными. В следующем разделе определены основные принципы концептуального подхода и кратко описаны разрабатываемые спецификации разделов астрономии. Раздел 3 посвящён конкретным астрономическим задачам, решение которых осуществляется с применением концептуального подхода. В Заключении сделаны выводы о преимуществах и недостатках концептуального подхода к решению задач в астрономии. Приложение содержит пояснения к ключевым понятиям концептуального подхода.

## 2. О КОНЦЕПТУАЛЬНОМ ПОДХОДЕ К РЕШЕНИЮ ЗАДАЧ В АСТРОНОМИИ

Концептуальный подход заключается в описании *предметных областей* исследований и формулировании решения научных задач над такими абстрактными спецификациями. Это позволяет избежать зависимости выполнения запросов к конкретным ресурсам данных от их структуры, привести в исследования данных семантические подходы к их интеграции, обеспечить возможность повторного использования разработанных методов в задачах, решаемых в предметной области.

Для формализации и систематизации *знаний предметной области* служат *онтологии*, представляющие собой логические формализмы для описания понятий предметной области. В соответствии с принятой в научном сообществе *концептуализацией* предметной области они позволяют разработать достаточно абстрактные и формальные соглашения о понятиях и их связях, принимаемые всем сообществом.

Если онтологии используются для рассуждений на уровне понятий предметной области, то для представления информации в предметной области служат *концептуальные схемы*. Структуры в схемах предназначены для согласованного представления и обработки в них данных об *объектах предметной области*. Помимо структуры в концептуальных схемах специфицируется поведение объектов предметной области. Концептуальные схемы и онтологии предметных областей разрабатываются сообществами таким образом, чтобы научным группам их было достаточно для описания научных задач.

Для корректного учета неоднородности реальных данных необходимо производить отображение схем ресурсов, имеющих отношение к рассматриваемым предметным областям, в концептуальные спецификации этих областей. Поля каталогов снабжаются *семантическими аннотациями* в терминах онтологии предметной области (описывается семантика полей). Эти аннотации используются для соотнесения с соответствующими элементами концептуальной схемы предметной области, также описанными в терминах онтологии. Далее поля каталогов отображаются в элементы концептуальных схем — типы, атрибуты типов, функции — таким образом, чтобы формировать данные в соответствии с типами концептуальной схемы из соответствующих по семантике полей.

Нередко такое отображение схем требует решения достаточно сложных задач, связанных напрямую с преодолением неоднородности между семантикой каталогизированных данных и семантикой схем предметной области. Например, ниже в разделе, описывающем задачу построения предметной области кратных систем звёзд, упомянута

<sup>1</sup><http://www.eso.org/sci/software/esomidias/>

<sup>2</sup><http://iraf.noao.edu/>

<sup>3</sup><http://aips2.nrao.edu/docs/aips++.html>

<sup>4</sup>Здесь и далее курсивом выделены первые упоминания терминов, которые определяются и разъясняются в приложении.

<sup>5</sup><http://aladin.u-strasbg.fr/aladin.gml>

<sup>6</sup><http://www.star.bris.ac.uk/~mbt/topcat/>

проблема неоднозначности идентификации астрономических объектов в разных каталогах. В одном из них идентификатор может описывать звезду как одиночную, в другом — как пару, в третьем — принадлежать одному из компонентов этой пары. Отображение в предметную область в каждом из этих случаев будет отличаться.

Для сравнения, построение выборки в сервисе VizieR<sup>7</sup> начинается с поиска по ключевым словам списка каталогов, содержащих интересующие параметры или поля. Далее выбираются каталоги из списка, отмечаются необходимые для выборки поля, в VizieR загружается список интересующих исследователя астрономических объектов, после чего формируется выборка необходимых данных.

В отличие от выборок в сервисе VizieR, в представленном концептуальном подходе отображение всех полей каталогов в спецификации предметной области производится единожды при регистрации каталогов в системе. Именно на этом этапе разрешаются структурные и семантические конфликты между представлениями данных в каталоге и в рассматриваемой предметной области. После этого задачи любой сложности можно формулировать над структурами и методами спецификаций предметной области, а необходимость обращения к тем или иным каталогам для создания выборки возьмёт на себя исполнительный механизм системы.

Концептуальный подход позволяет упростить и сам процесс решения научных задач, очень часто состоящий в формулировании различных запросов к концептуальной схеме предметной области. В этом случае решение задач не требует использования языков программирования, создания отдельных информационных систем.

Существуют архитектуры виртуальной интеграции информационных ресурсов в концептуальные схемы предметных областей, при которой предполагается не преобразование и загрузка всей информации из интегрируемых каталогов в информационную систему, а задание запросов к концептуальной схеме и автоматическое направление частей запросов к отображённым в неё источникам данных и сервисов. Так работают предметные посредники [2]. В пору существования сети AstroGrid,<sup>8</sup> представлявшей собой среду совместных исследований с общедоступными сервисами, проводилась работа по её интеграции с предметными посредниками, позволяющими формулировать решение задач в грид-среде в терминах спецификаций предметных областей, а не в терминах вызовов доступных сервисов [3]. На примере интеграции социальной сети

исследований myExperiment, собирающей коллекцию доступных сообществу реализаций научных методов [4] и потоков работ [5], с предметными посредниками демонстрировалось применение концептуального подхода для повторного использования потоков работ за счёт обеспечения независимости от конкретных коллекций данных и сервисов [6] и семантического поиска потоков работ [7].

В рабочих группах альянса «Международная виртуальная обсерватория» в некотором виде реализованы описания понятий астрономии. Такой реализацией является, в частности, стандарт UCD (Unified Content Descriptors),<sup>9</sup> используемый для аннотации полей астрономических каталогов и определения их смысла. UCD успешно используется, однако сильно ограничен используемой *моделью данных*, недостаточно выразительной для описания понятий однозначным образом. Во множестве случаев для описания полей каталогов, имеющих совершенно разную семантику, используются одним и тем же UCD.

В качестве альтернативы UCD развиваются словари,<sup>10</sup> содержащие множество терминов, помимо имеющихся в UCD. Они включают большинство понятий, встречаемых в астрономии. Между терминами определены тезаурусные семантические отношения: синонимия, гипер- и гипонимия и ассоциативные отношения, но для использования в семантических подходах они недостаточны, так как не включают однозначного определения терминов.

Другим известным примером систематизации знаний является онтология типов астрономических объектов<sup>11</sup> на языке OWL (Web Ontology Language).<sup>12</sup> Понятия данной онтологии определяют не только названия астрономических объектов и их иерархию, но и выражают некоторые знания об астрономических объектах разного типа, ограничивая возможную интерпретацию понятий. Однако и эта онтология сильно ограничена, она охватывает только виды астрономических объектов, не затрагивая понятия различных проблемных областей.

Таким образом, данное направление требует дальнейшего развития. Исследовательским сообществам следует производить описание онтологий для представляемых ими предметных областей с использованием уже существующих наработок. Онтологии должны содержать достаточно подробные определения понятий, принимаемых сообществами, включая вербальные определения, отношения между понятиями и ограничения.

<sup>9</sup><http://www.ivoa.net/documents/latest/UCD.html>

<sup>10</sup><http://www.ivoa.net/documents/latest/Vocabularies.html>

<sup>11</sup><http://www.ivoa.net/documents/latest/AstrObjectOntology.html>

<sup>12</sup><http://www.w3.org/TR/owl-overview/>

<sup>7</sup><http://vizier.u-strasbg.fr/viz-bin/VizieR>

<sup>8</sup><http://www.astrogrid.org/>

Концептуальные схемы всегда являлись частью разработки больших каталогов. Каталоги, отражающие наблюдения небесных объектов, а также включающие данные разнонаправленных исследований, помимо внушительных объёмов могут иметь сложную структуру для хранения различного рода параметров. Такие каталоги заполняются новой информацией автоматически при разборе обзоров с помощью определённого набора алгоритмов. Одновременно производится расчёт различных параметров, выделение специфических объектов предметных областей. Для организации хранилищ, отвечающих результатам обработки данных, применяются подходы к концептуальному моделированию в теории баз данных. Исходя из состава данных о предметной области и решаемых задач при автоматической обработке, моделируются структуры хранения.

Так, развитие каталога SDSS [8] происходило по мере проведения наблюдений и решения различных задач, использующих поступающие данные. К простому обзору неба добавлялось решение специфических задач, таких как исследование Галактики, поиск сверхновых звёзд, инфракрасные исследования дальних звёзд и спектроскопические наблюдения ближних звёзд, исследование галактик и квазаров с учётом красного смещения, исследование распределения вещества во Вселенной. Структура каталога отражает структуру предметных областей, связанных с этими и другими задачами.

При постоянном увеличении требований к обрабатываемым объёмам информации и к взаимодействию сообществ исследователей любая область исследования будет испытывать потребность концептуализации, принимаемой всем сообществом, и даже стандартизации представления данных в предметной области. Международная виртуальная обсерватория разрабатывает ряд рекомендаций по структуре концептуальных схем (называемых в альянсе термином «модель данных») для общеупотребительных областей астрономии (VOEvent,<sup>13</sup> Spectral Lines Data Model,<sup>14</sup> Photometry Data Model<sup>15</sup> и другие). Разработка концептуальных спецификаций международным астрономическим сообществом обозначает позитивное движение к концептуальному подходу в научных исследованиях.

Разрабатываемые на основе онтологий разделов астрономии концептуальные схемы используются повторно при решении разных задач и претерпевают незначительное развитие при появлении новых задач, затрагивающих ту же область. Это

показано в следующем разделе на примерах решаемых с применением концептуального подхода задач в астрономии.

### 3. РЕШЕНИЕ АСТРОНОМИЧЕСКИХ ЗАДАЧ

#### 3.1. База данных двойных и кратных звёзд

В астрономическом сообществе, исследующем двойные звёзды, используются базы данных по различным типам двойных: визуальным WDS,<sup>16</sup> затменным CEV [9], спектроскопическим SB9,<sup>17</sup> орбитальным ORB6,<sup>18</sup> интерферометрическим INT4<sup>19</sup> и множеству других выборок. Однако до недавнего времени не существовало базы данных, содержащей сведения о всех типах двойных. Ликвидировать этот пробел призвана база данных двойных и кратных звёзд BDB<sup>20</sup> [10].

Целью создания BDB являлось предоставление пользователю каталогизированных данных о двойных звёздах, возможности выборки звезд, отвечающих требованиям исследователя и доступа к полному набору данных из разных каталогов, включенных в состав базы. Для этого в BDB интегрируются каталоги всех типов двойных, каталоги астрофизических параметров, содержащие данные о массах, радиусах, температурах и другие источники данных.

Перекрыстная идентификация проводится на этапе создания базы по идентификаторам различных видов, присутствующим в разных каталогах. Как уже упоминалось, неоднородность представления информации об одних и тех же астрономических объектах в разных каталогах рождает проблему соотнесения идентификаторов определённым компонентам. Дело в том, что в некоторых каталогах идентификатор может описывать компонент как одиночную звезду, в то время как в других каталогах, использующих более точные подходы к обнаружению двойных, этот же компонент может быть описан как пара, и тот же идентификатор дан обоим компонентам этой пары, что может показаться ошибкой каталога или привести к неверному сопоставлению идентификаторов в других стандартах. Подобные ситуации необходимо обнаруживать и разрешать автоматически, либо с привлечением экспертов. Для исключения риска ложных отождествлений требуется привлечение дополнительных наблюдательных данных по близости координат и светимостей компонентов, собственных движений, близких

<sup>13</sup><http://www.ivoa.net/Documents/VOEvent/>

<sup>14</sup><http://www.ivoa.net/documents/SSLDM/>

<sup>15</sup><http://www.ivoa.net/documents/PHOTDM/>

<sup>16</sup><http://ad.usno.navy.mil/wds/>

<sup>17</sup><http://sb9.astro.ulb.ac.be/>

<sup>18</sup><http://ad.usno.navy.mil/wds/orb6.html>

<sup>19</sup><http://ad.usno.navy.mil/wds/int4.html>

<sup>20</sup><http://bdb.inasan.ru>

для систем в целом. В BDB выбрана собственная система идентификации, учитывающая возможную детализацию информации о системах звёзд и существующие идентификаторы компонентов однозначно соотнесены с идентификаторами BDB.

Схема BDB сформирована на основании состава полей, которые обычно присутствуют в каталогах, с учетом необходимости индексирования по ним. Она содержит типы, включающие описание систем в целом, пар и компонентов. Атрибуты, характеризующие параметры звёзд, связаны с одним из этих типов. Ещё одним существенным типом в BDB является таблица идентификации.

Одновременно разрабатывается предметный посредник с описанием онтологии двойных, концептуальной схемы, близкой к схеме BDB, и возможностью задания разнообразных запросов к предметной области двойных [11]. Онтология вмещает в себя как понятия, специфические для области двойных, так и различные модули, содержащие смежные области, затрагиваемые при работе с двойными и кратными звёздами. Прототип онтологии представлен на языке OWL и доступен в Интернете.<sup>21</sup>

Круг вопросов, рассматриваемых в данной области, достаточно широк. Во-первых, исследование двойных и кратных звёзд связано с наиболее общими разделами астрономии, такими как измерения (понятия, необходимые для описания характеристик объектов предметных областей, единиц и точности измерения), системы координат, фотометрические системы, спектроскопия, астрофизические описания астрономических объектов (в частности, понятия массы, температуры), наиболее общие понятия об астрономических объектах (понятия составных объектов, звёздных объектов), но не затрагивает само множество их разновидностей. Во-вторых, в спецификации представлены более специализированные области понятий: звёзды и их характеристики, кратные системы и их составляющие (пары, компоненты), параметры орбит, параметры кривых блеска, характеристики затменных двойных и другие. Это, в основном, понятия, используемые в исследовании конкретных классов астрономических объектов, в конкретных методиках их наблюдения.

Концептуальная схема предметной области посредника разработана на основе знаний, вошедших в онтологию, с использованием принципов, изложенных в работе [12]. Схема включает описания типов для представления данных об основных объектах предметной области: системах, парах и компонентах (таблица 1), — а также данных смежных

областей знаний (таблица 2), таких как астрометрия, фотометрия, параметры кривых блеска, параметры орбит и другие. Фактически атрибуты типов формируются на основе связей между понятиями различных разделов астрономии.

Правила преобразования структуры данных в соответствии с типами концептуальной схемы посредника из данных, присутствующих в каталогах, задаются единожды для каждого каталога и используются при обработке запросов к концептуальной схеме для автоматического обращения к оригинальным каталогам. Модульность концептуальной схемы позволяет формулировать задачи над спецификациями смежных областей и соответственно автоматически использовать для их решения данные из каталогов, связанных с определёнными модулями.

Разработка концептуальной схемы и применение технологии посредников для области двойных и кратных систем позволяет не просто собрать в один каталог данные из каталогов разных типов двойных, но задавать всевозможные запросы в терминах одной схемы и получать ответы с данными из множества оригинальных каталогов. Такими запросами могут являться поиск всех данных о конкретном астрономическом объекте, всех объектов в участке неба или наблюдений во времени, объектов с определёнными диапазонами значений параметров, с определённой точностью наблюдений, со сходными или зависимыми характеристиками, взаимодействующих определённым образом компонентов и другие. Состав учитываемых в исследованиях каталогов может меняться динамически, не влияя на работоспособность системы в целом и на формулировки решения задач над концептуальной схемой посредника.

Таким образом, преимущества применения подхода заключаются в формализации знаний о предметной области двойных в одной концептуальной схеме, в возможности решения многих задач предметной области над ней, в универсализации спецификаций решения задач и независимости их от сегодняшнего набора источников данных. Особенно полезен подобный подход для сообщества исследователей двойных звёзд, включая специалистов в более узких областях, а также для взаимодействия с сообществами в смежных специализациях.

Формулирование алгоритмов решения задач требует в такой инфраструктуре большей универсальности, однако в большинстве случаев это нельзя назвать недостатком. Во-первых, задачи формулируются единожды над различными наборами данных. Во-вторых, в целом универсальный подход к разработке методов решения задач предпочтителен и характерен для научных исследований.

<sup>21</sup><http://ontology.ipi.ac.ru/ontologies/astront>

**Таблица 1.** Специальные типы концептуальной схемы предметной области двойных и кратных звёзд

<i>System</i>	
pairs: Pair*	
objects: Component*	
refObject: Component	
<i>Pair</i>	
system: System (inverse: System.pairs)	
primary: Component	
secondary: Component	
epoch: Integer	
plx: Measurement	
curve: LightCurve	
V0: Measurement	
rel: RelPosition	
orbit: Orbit	
oType: ObsType*	
eType: EvolStatus	
massRatio: Measurement	
rocheFilling: Measurement	
<i>Component</i>	
system: System (inverse: System.objects)	
primaryOf: Pair (inverse: Pair.primary)	
secondaryOf: Pair (inverse: Pair.secondary)	
refObjectOf: System (inverse: System.refObject)	
epoch: Integer	
crd: Coordinate	
plx: Measurement	
pm: ProperMotion	
mag: Magnitude*	
spType: SpectralType	
lumClass: LumClass	
Rv: Measurement	
K: Measurement	
mass: Measurement	
t: Measurement	

\* Множественное значение атрибута.

**Таблица 2.** Описание схемы для данных из смежных областей

<i>Coordinate</i>	
object: Component	
crdRA: Float	
crdDE: Float	
<i>Magnitude</i>	
object: Component	
value: Measurement	
band: Passband	
<i>ProperMotion</i>	
object: Component;	
pmRA: Float	
pmDE: Float	
<i>RelPosition</i>	
pair: Pair	
theta: Float	
rho: Float	
<i>Orbit</i>	
P: Measurement	
A: Measurement	
I: Measurement	
E: Measurement	
T: Measurement	
w: Measurement	
W: Measurement	
<i>LightCurve</i>	
P: Measurement	
curveType: lightCurveMorphology	
magMax: Magnitude	
depthPrimary: Magnitude	
depthSecondary: Magnitude	
durPrimary: Measurement	
durSecondary: Measurement	
<i>Measurement</i>	
value: Float	
error: Float	

### 3.2. Поиск кандидатов во вторичные стандарты

Исследование транзиентных явлений, например, оптических послесвечений космических гамма-всплесков, требует быстрого реагирования на сообщение о произошедшем явлении. Когда такая информация поступает, необходимо безотлагательно навестись на область бокса ошибок послесвечения и начинать наблюдение. Как правило, при этом возникает необходимость калибровки наблюдений с помощью звезд-стандартов, находящихся в той же части неба. В области послесвечения может не оказаться известных фотометрических стандартов (например, из списка Ландольта [13]), поэтому возникает задача поиска в данной области наиболее «стабильных» звёзд, кандидатов во вторичные стандарты. Эта задача была решена<sup>22</sup> с применением описанного концептуального подхода и предметного посредника, к концептуальной схеме которого при появлении информации о координатах произошедшего всплеска задаются запросы [14].

Схемы каталогов SDSS [8], USNOB-1.0 [15], 2MASS [16], GSC [17], UCAC [18] отображены в концептуальную для виртуальной интеграции каталогов в посреднике. При помощи запросов к посреднику выбираются стабильные звёзды, которые должны удовлетворять следующим критериям:

- принадлежать указанной площадке в области послесвечения всплеска;
- быть изолированными;
- не обнаруживать собственного движения;
- иметь «надёжную» фотометрию;
- иметь определённые пределы блеска;
- не быть переменными.

Все условия, кроме последнего, разрешаются над указанными каталогами, после чего проводится перекрёстная идентификация выявленных астрономических объектов, чтобы исключить дубликаты из разных каталогов. Для исключения переменных звезд система обращается к ещё нескольким каталогам, а именно VSX,<sup>23</sup> ASAS [19], GCVS [20], NSVS [21], которые также интегрированы в посреднике. Если выбранные по критериям звёзды идентифицируются с каким-либо объектом из этих каталогов, они рассматриваются как переменные и отбрасываются.

<sup>22</sup>Подписку на рассылку по электронной почте результатов поиска кандидатов во вторичные стандарты в поле послесвечения гамма-всплесков можно согласовать с А. С. Позаненко (apozanen@iki.rssi.ru).

<sup>23</sup><http://www.aavso.org/vsx/index.php>

Концептуальная схема посредника построена на основе знаний онтологии предметной области, связанной с многоцветными фотометрическими наблюдениями. Описываемая система работает в автоматическом режиме и предоставляет список кандидатов во вторичные стандарты для каждого поступающего извещения о гамма-всплеске. При появлении новых каталогов, которые могли бы улучшить результаты поиска кандидатов в стандарты или позволили бы выявлять больше переменных звёзд с определённой долей точности, достаточно отобразить схемы таких каталогов в концептуальную схему посредника. Это не повлияет на работоспособность системы и не потребует её переписывания.

Таким образом, концептуальный подход к решению данной задачи позволил решить её над концептуальной схемой, характерной для решения многих задач в области многоцветной фотометрии. Разработанная система может повторно использоваться для решения целого класса задач в этой области. Реализация алгоритма поиска вторичных стандартов имеет вид запроса к схеме, и не зависит от изменения со временем состава используемых астрономических каталогов.

### 3.3. Поиск фотометрического красного смещения галактик

С использованием наработок спецификаций предметной области, полученных в процессе решения предыдущей задачи, при незначительном их развитии, возможно решать задачу определения космологического красного смещения галактик. Существуют следующие подходы к определению расстояния до галактик: спектроскопический, учитывающий смещение спектральных линий, и фотометрический, основанный на сравнении синтетических спектров галактик со спектральными распределениями энергии (SEDs), полученными по величинам блеска галактик в широкополосных или узкополосных фильтрах. Фотометрический метод оценки красного смещения галактик бывает эффективен, когда нет возможности получить спектр галактики (например, если галактика имеет слабый блеск), или же в спектре отсутствуют заметные линии. В некоторых случаях фотометрический способ определения красного смещения является единственным доступным методом оценки расстояния до объекта и определения его основных свойств (см., напр., [22]).

Для решения задачи используются фотометрические данные из известных каталогов с максимальным набором полос пропускания. Например, используются полосы *ugriz* из каталога SDSS, дополненные полосами JHK из каталога 2MASS.

При генерации наблюдаемого спектра учитывается поглощение в Галактике. Сформированный спектр сравнивается с тестовыми шаблонными синтетическими спектрами, при этом подбираются параметры тестовых спектров (красное смещение, тип, возраст и другие), обеспечивающие наилучшее совпадение с исследуемым спектром. Библиотеки синтетических спектров учитывают в себе как звёздную эволюцию, так и различные модели эволюции галактик различных типов, а также могут включать в себя дополнительные эмиссионные или абсорбционные линии. Одной из наиболее используемых библиотек синтетических спектров является библиотека PEGASE2 [23]. Также есть возможность учесть внутреннее поглощение в исследуемой галактике, выбирая определённый закон поглощения (например, закон поглощения для Млечного Пути или для Магеллановых Облаков).

Таким образом, для решения данной задачи можно использовать схемы и методы, накопленные при решении задачи поиска фотометрических стандартов, дополнительно используются определения предметной области, связанной с исследованием галактик.

Примечательно, что задача параметризации звёзд Галактики на основе широкополосной фотометрии из различных каталогов, решаемая коллективом из ИНАСАН [24] и, в том числе, авторами настоящей работы, также, в основном, укладывается в разработанные спецификации предметной области, используемой в задачах поиска вторичных стандартов и определения фотометрического красного смещения галактик. При её решении используется аналогичный подход, в котором строится наблюдаемый спектр звезд, учитывается межзвёздное поглощение в Галактике и результирующий спектр сравнивается с базой известных звездных спектров.

В настоящее время существуют коды, разработанные специально для оценок красного смещения фотометрическим методом, например, HyperZ [25], Le Phare [26, 27], EAZY [28], Evolution of Radio Galaxies [29] и другие. Они основаны на поиске минимума функционала в многомерном пространстве параметров и подразумевают наличие входных параметрических файлов, содержащих необходимые наблюдаемые величины. Применение концептуального подхода позволяет повторно использовать наработки концептуальных спецификаций в области фотометрии с уже интегрированными в них каталогами и свести решение задачи к составлению запросов к существующей системе при незначительном развитии спецификаций в смежной предметной области и в простоте добавления новых каталогов.

### 3.4. Классификация затменных

Ещё одна задача, решаемая авторами в рамках настоящего исследования, — это задача классификации затменных двойных систем. Ранее она была решена над данными каталога CEV<sup>24</sup> средствами классификации на основе машинного обучения [30]. С тех пор была разработана новая версия каталога, содержащая параметры и морфологические типы кривых блеска для 7200 систем [9], поэтому решение задачи потребовалось в новых условиях. Часть систем в новом каталоге имеют известный эволюционный класс, и на основе существующих данных необходимо классифицировать остальные двойные системы.

Задача решается независимо несколькими коллективами с применением разных подходов с целью последующего сравнения результатов. В первом подходе разрабатывается набор критериев, по которым можно отнести звёзды с определёнными параметрами к одному эволюционному классу. Во втором подходе используются методы классификации с машинным обучением. Методы, основанные на деревьях решений, позволяют также сделать выводы о влиянии тех или иных параметров звёзд на эволюционный класс, что может повлиять на решения, принимаемые в первом подходе. Следует заметить, что при решении данной задачи используются не только спецификации предметных областей, но и методы общего назначения, средства реализации которых должны быть доступны исследователям в различных областях знаний.

Задача решается над единственным каталогом небольшого объёма. Тем не менее, концептуальный подход к её решению предпочтителен. Явное концептуальное описание предметной области двойных позволяет реализовать решение задачи таким образом, чтобы реализацию можно было использовать повторно другими исследователями, в других задачах или над другими данными. С другой стороны, при решении этой задачи используются уже формализованные ранее в онтологии знания о предметной области, включая параметры звёзд, такие как светимость и спектральный тип, параметры кривых блеска, данные о разновидностях эволюционных статусов звёзд (таблица 3). Они все использовались при создании спецификаций предметной области двойных и кратных систем.

## 4. ЗАКЛЮЧЕНИЕ

Изложенный концептуальный подход к решению задач в науках с интенсивным использованием данных, состоящий в описании разрабатываемых научным сообществом общедоступных спецификаций предметных областей и их использовании

<sup>24</sup><http://www.inasan.ru/~malkov/CEV/>



**Таблица 3.** Эволюционный статус двойных звёзд

D: Detached systems
DM: Detached main sequence systems
DR: Detached sub-giant systems
DG: Detached giant or supergiant systems
DW: Detached systems with a white dwarf
D2S: Detached symbiotic systems
S: Semi-detached systems
SA: Classical Algols
SC: Cool semi-detached systems
SH: Hot semi-detached systems
S2: Late stage semi-detached systems
S2C: Cataclysmic systems
S2H: High-mass X-ray binaries
S2L: Low-mass X-ray binaries
C: Contact systems
CB: Near-contact systems
CE: Early-type contact systems
CW: Late-type contact systems
CWA: Late-type contact A systems, primary is larger
CWW: Late-type contact W systems, secondary is larger
CG: Giant contact systems

сообществом для формулирования задач и методов их решения, обладает рядом преимуществ перед непосредственным использованием источников информации. Во-первых, при разработке решения задач уменьшаются накладные расходы на разрешение технической и семантической неоднородности данных, эти проблемы решаются единожды для определённых источников информации при отображении их схем в концептуальные спецификации предметной области. Во-вторых, появляется возможность накопления, систематизации, семантического поиска и повторного использования произведённых различными научными коллективами работ, включая реализации методов и полученные результаты. В-третьих, исчезает зависимость реализаций научных методов от конкретных источников информации, используемых при решении задач.

Таким образом, представленный подход создаёт предпосылки к взаимодействию исследователей внутри сообществ, работающих в близких или

смежных предметных областях, и сосредоточению на совместное развитие научных методов в исследуемых областях знаний, что ведёт к развитию науки и эффективности исследований. В условиях быстрого роста объёмов обрабатываемой информации, количества и разнообразия источников данных, появления потоковых источников, расширения направлений исследований, для повышения эффективности исследований требуются совместные усилия специалистов в области астрономии и информатики по реализации представленного в работе подхода.

#### ПРИЛОЖЕНИЕ. ИСПОЛЬЗУЕМЫЕ ТЕРМИНЫ КОНЦЕПТУАЛЬНОГО ПОДХОДА

Ниже определены некоторые термины, используемые в данной работе. Они также разъясняются на примерах исследования двойных и кратных звёзд.

*Предметная область* — круг научных интересов, характеризующийся определённым множеством объектов исследования, специфическими методами и инструментами исследования, определёнными классами решаемых научных задач и сформированным научным сообществом, занимающимся исследованием.

Предметная область двойных и кратных звёзд изучает широкий класс систем звёзд, гравитационно связанных между собой или визуально близких, и коррелирует с множеством областей, в которых предметами исследования являются системы с определёнными разновидностями взаимодействия (например, системы с аккрецией вещества) или определённым образом проявляющие себя для наблюдателя (например, затменные двойные). Для данной области характерен широкий круг решаемых задач, таких как классификация, параметризация взаимодействующих пар звёзд в таких системах, а также различные методы их наблюдения и решения задач с использованием фотометрии, спектроскопии и других подходов.

*Объект предметной области* — вещь, явление или процесс, на которые направлена исследовательская деятельность в данной предметной области, либо фигурирующая в исследованиях предметной области.

В области двойных и кратных звёзд объектами предметной области являются сами кратные системы звёзд, различные типы компонентов и пар звёзд в составе систем, а также явления, наблюдаемые в них.

*Знания предметной области* — зафиксированная и проверенная практикой информация об объектах предметной области, которая использовалась и может многократно использоваться для

решения задач и принятия решений в ней. Знания включают как факты об объектах предметной области, так и закономерности их поведения.

В области двойных звёзд можно привести следующие примеры знаний: информация о наиболее изученных кратных системах и параметрах их взаимодействия; законы Кеплера в применении к взаимодействию компонентов двойных систем; эмпирически установленная близость значений собственного движения и возрастов всех компонентов системы; этапы эволюции тесных систем; методы обнаружения двойных систем, определяющие их наблюдательный тип; и другие факты и закономерности, известные и постоянно используемые в предметной области.

*Декларативные знания* — описательная форма представления знаний в виде утверждений об объектах предметной области, их свойствах и отношениях между ними. В отличие от них процедурная форма представления знания описывает пошаговый процесс преобразования объектов предметной области. Однако и преобразование объектов предметной области можно описывать декларативно в виде предусловий, описывающих объекты до преобразования, и пост-условий, описывающих результаты преобразований. Декларативность описаний позволяет использовать над ними формальные рассуждения, распараллеливать анализ данных.

*Концептуализация* — процесс осмысления предметной области и формирования понятий для идентификации и классификации её объектов. Также концептуализацией называют результат этого процесса.

*Онтологии* — декларативные спецификации понятий предметной области и связей между ними. Онтологии выражают в явном виде концептуализацию предметной области. В общем случае они не связаны со структурой данных, а отражают только концептуализацию предметной области, то есть, её разделение на понятия.

Например, в онтологии двойных и кратных звёзд одним из ключевых понятий является «пара». Это понятие наследует отношение «иметь компонент» от понятия составного объекта, отношение «иметь координаты» от понятия астрономического объекта и другие спецификации. Однако в паре множественность отношения «иметь компонент» дополнительно ограничивается ровно двумя компонентами. Понятие пары связано отношениями с понятием кратной системы, которой принадлежит пара, а также с понятием наблюдательного типа. С понятием пары связаны понятия взаимного расположения компонентов, орбитального движения, отношения масс, заполнения полости Роша и другие.

*Концептуальная схема* предметной области — абстрактная спецификация структур для

представления данных об объектах, присущих данной предметной области, поведения объектов предметной области, и ограничений целостности, выполняемых при любых манипуляциях данными. Схема описывается на определённом языке концептуального моделирования, определяющем модель данных.

В концептуальной схеме для представления данных о паре звёзд в составе кратной системы определяется тип «пара» с набором атрибутов, ассоциаций и методов. Атрибуты включают наблюдательный тип, являющийся перечислением или строковым значением, позиционный угол и угловую скорость как вещественные числа, атрибуты отношения масс и заполнения полости, также имеющие вещественные значения, и другие. Ассоциации реализуют представление ссылок на объекты системы, первичного и вторичного компонентов пары. Таким образом, то, что в онтологии представляется отдельными понятиями, в схеме может иметь тип строки или числовой тип. К тому же, в абстрактном типе пары могут быть определены методы, в частности, метод оценки отношения масс на основе значений других параметров.

*Концептуальные спецификации предметных областей* — абстрактное описание на определённом языке обобщённых знаний о предметной области, определяющих свойства объектов предметной области и их взаимосвязь друг с другом. Абстрактность обеспечивает независимость описаний предметной области от реализаций спецификаций и от конкретных источников данных. Для достижения различных целей используются спецификации онтологий или концептуальных схем предметных областей.

*Семантическая аннотация* — информация, сопровождающая данные, документы, спецификации и описывающая их смысл, происхождение и другие аспекты. Обычно аннотации содержат набор полей для определения названия, типа объекта, его авторов, даты создания и другие. Семантическое аннотирование для сопровождения объектов понятиями, определяющими их смысл в предметной области, проводится в терминах онтологий данной предметной области. Примером семантических аннотаций, широко используемых в астрономии, являются UCD, прикрепляемые к полям каталогов.

*Модель данных* — в теории баз данных набор языков, на которых определяются структура данных, методы манипулирования данными и ограничения их целостности. Примерами моделей данных являются реляционная модель данных, являющаяся основой систем управления базами данных и определяющая операции над таблицами, объектно-ориентированная модель данных, используемая в языках программирования, и модель

данных языка онтологий OWL, определяемая специфической логикой, лежащей в его основе. Другими словами, модель данных определяет язык, на котором специфицируются концептуальные схемы, а не сами схемы данных.

*Поток работ* — спецификация многоступенчатого процесса, представляющая собой граф, в узлах которого находятся определённые задачи или сервисы, вызываемые для обработки данных, а направленные дуги определяют направление передачи данных от задачи к задаче.

## БЛАГОДАРНОСТИ

Приносим благодарность анонимному рецензенту статьи за подробные замечания, позволившие существенно улучшить излагаемый материал. Работа выполнена при поддержке Российского фонда фундаментальных исследований (гранты 12-02-01336, 12-07-00528, 13-01-92204, 13-07-00579, 14-07-00548, 15-02-04053, 16-07-01162), программ фундаментальных исследований Президиума РАН No. 16 «Фундаментальные проблемы системного программирования» и No. 41 «Нестационарные явления в объектах Вселенной», госзаказа ИПИ РАН, тема No. 38.25 «Спецификация и решение задач анализа данных в концептуальных терминах предметных областей с интенсивным использованием данных на основе Big Data-ориентированных информационных технологий», программы государственной поддержки ведущих научных школ РФ (грант 3620.2014.2), Национального исследовательского фонда (National Research Foundation) ЮАР.

## СПИСОК ЛИТЕРАТУРЫ

1. *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Ed. by T. Hey, S. Tansley, and K. Tolle (Microsoft Research, Redmond, 2009).
2. D. O. Briukhov, A. E. Vovchenko, V. N. Zakharov, et al., *Informatics and Applications* **2** (1), 2 (2008).
3. A. E. Vovchenko, L. A. Kalinichenko, and S. A. Stupnikov, in *Proc. 4th Int. Conf. on Distributed Computing and Grid Technologies in Science and Education, Dubna, Russia, 2010* (JINR, Dubna, 2010), p. 309.
4. N. A. Walton, D. K. Witherwick, T. Oinn, and K. M. Benson, *ASP Conf. Ser.* **394**, 309 (2008).
5. C. A. Goble and D. C. De Roure, in *Proc. 2nd Workshop on Workflows in Support of Large-Scale Science, Monterey, USA, 2007* (ACM, New York, 2007), p. 1.
6. D. Briukhov, A. Vovchenko, and L. Kalinichenko, in *Proc. 15th Russian Conf. on Digital Libraries RCDL 2013, Yaroslavl, Russia, 2013*, Ed. by

- V. Smirnov and S. Stupnikov, *CEUR Workshop Proc.* **1108**, 61 (2013).
7. N. Skvortsov, D. Briukhov, L. Kalinichenko, et al., in *Proc. 15th Russian Conf. on Digital Libraries RCDL 2013, Yaroslavl, Russia, 2013*, Ed. by V. Smirnov and S. Stupnikov, *CEUR Workshop Proc.* **1108**, 70 (2013).
8. C. P. Ahn, R. Alexandroff, C. Allende Prieto, et al., *Astrophys. J. Suppl.* **203**, 21 (2012).
9. E. A. Avvakumova, O. Yu. Malkov, and A. Yu. Kniazhev, *Astronomische Nachrichten* **334**, 860 (2013).
10. D. A. Kovaleva, P. V. Kaygorodov, O. Yu. Malkov, et al., *Astronomy and Computing* **11**, 119 (2015).
11. N. A. Skvortsov, O. Yu. Malkov, P. V. Kaygorodov, et al., in *Proc. 15th Russian Conf. on Internet and Modern Society IMS 2012, St. Petersburg, Russia, 2012*, p. 107.
12. A. E. Vovchenko, V. N. Zakharov, L. A. Kalinichenko, et al., in *Proc. 12th Russian Conf. on Digital Libraries RCDL 2010, Kazan, Russia, 2010*, p. 375.
13. A. U. Landolt, *Astron. J.* **88**, 439 (1983).
14. A. E. Vovchenko, A. A. Volnova, D. V. Denisenko, et al., in *Abstracts Russian Astron. Conf. VAK 2010, Nizhny Arkhyz, Russia, 2010*, No. 257.
15. D. G. Monet, S. E. Levine, B. Canzian, et al., *Astron. J.* **125**, 984 (2003).
16. M. F. Skrutskie, R. M. Cutri, R. Stiening, et al., *Astron. J.* **131**, 1163 (2006).
17. J. E. Morrison, S. Röeser, B. McLean, et al., *Astron. J.* **121**, 1752 (2001).
18. N. Zacharias, S. E. Urban, M. I. Zacharias, et al., *Astron. J.* **127**, 3043 (2004).
19. G. Pojmanski, B. Pilecki, and D. Szczygiel, *Acta Astronomica* **55**, 275 (2005).
20. N. N. Samus, O. V. Durlевич, E. V. Kazarovets, et al., *General Catalogue of Variable Stars, VizieR Online Data Catalog*, 102025.
21. P. R. Woźniak, W. T. Vestrand, C. W. Akerlof, et al., *Astron. J.* **127**, 2436 (2004).
22. A. A. Volnova, A. S. Pozanenko, J. Gorosabel, et al., *Monthly Notices Royal Astron. Soc.* **442**, 2586 (2014).
23. M. Fioc and B. Rocca-Volmerange, *Astron. and Astrophys.* **326**, 950 (1997).
24. S. G. Sichevskij, A. V. Mironov, and O. Yu. Malkov, *Astrophysical Bulletin* **69**, 160 (2014).
25. M. Bolzonella, J.-M. Miralles, and R. Pello, *Astron. and Astrophys.* **363**, 476 (2000).
26. K. A. Arnaud, *ASP Conf. Ser.* **101**, 17 (1996).
27. O. Ilbert, S. Arnouts, H. J. McCracken, et al., *Astron. and Astrophys.* **457**, 841 (2006).
28. G. B. Brammer, P. G. van Dokkum, and P. Coppi, *Astrophys. J.* **686**, 1503 (2008).
29. O. V. Verkhodanov, A. I. Kopylov, Yu. N. Parijskij, et al., *Bull. Spec. Astrophys. Obs.* **48**, 41 (1999).
30. O. Yu. Malkov, E. Oblak, E. A. Avvakumova, et al., *Astron. and Astrophys.* **465**, 549 (2007).

## Conceptual Approach to Astronomical Problems

**N. A. Skvortsov, E. A. Avvakumova, D. O. Bryukhov, A. E. Vovchenko, A. A. Volnova,  
O. B. Dluzhnevskaya, P. V. Kaygorodov, L. A. Kalinichenko, A. Yu. Knyazev, D. A. Kovaleva,  
O. Yu. Malkov, A. S. Pozanenko, and S. A. Stupnikov**

New technical capabilities have brought about the sweeping growth of the amount of data acquired by the astronomers from observations with different instruments in various parts of the electromagnetic spectrum. We consider conceptual approach to be a promising tool to efficiently deal with these data. It uses problem domain knowledge to formulate the tasks and develop problem-solving algorithms and data analysis methods in terms of domain concepts without reference to particular data sources, and thereby allows solving certain problems in general form. We demonstrate the benefits of conceptual approach by using it to solve problems related to search for secondary photometric standard candidates, determination of galaxy redshifts, creation of a binary and multiple star repository based on inhomogeneous databases, and classification of eclipsing binaries. We formulate and solve these problems over specifications of astronomical knowledge units such as photometric systems, astronomical objects, multiple stars, etc., and define them in terms of the corresponding problem domains independently of the existing data resources.

*Keywords: methods: data analysis—astronomical data bases: miscellaneous—catalogues—virtual observatory tools*