

УДК 519.685.5:524.72-13(083)

БАЗА НАБЛЮДАТЕЛЬНЫХ ДАННЫХ ДЛЯ ИЗУЧЕНИЯ БЛИЖНЕЙ ВСЕЛЕННОЙ

© 2012 Е. И. Кайсина, Д. И. Макаров, И. Д. Караченцев, С. С. Кайсин

Специальная астрофизическая обсерватория, Нижний Архыз, 369167 Россия

Поступила в редакцию 16 августа 2011 г.; принята в печать 5 октября 2011 г.

Представлено описание созданной базы данных галактик Местного объема, расположенных в пределах 10 Мпк вокруг Млечного Пути (LVG). Она содержит более 800 объектов. Исходя из анализа функциональных возможностей, в качестве системы управления нашей базой LVG использована СУБД PostgreSQL. С использованием методов семантического моделирования разработана физическая ER-модель базы данных. Приведено описание разработанной схемы структуры таблиц базы и организации Web-доступа к ней <http://www.sao.ru/lv/lvgdb>.

Ключевые слова: *астрономические базы данных: каталоги—астрономические базы данных: обзоры—галактики: фундаментальные параметры*

1. ВВЕДЕНИЕ

За последние 15 лет различными наблюдательными группами предпринято множество усилий по обнаружению близких галактик и систематизации их характеристик. Первый шаг к созданию выборки наиболее близких галактик был сделан в работе [1]. Этот список содержал 179 галактик, удовлетворяющих условию $V_{LG} < 500$ км/с, где V_{LG} — радиальная скорость галактики относительно центра Местной Группы. Спустя 15 лет был опубликован обновленный список галактик Местного объема, который содержал 226 объектов [2]. С появлением новых наблюдательных оптических и ИИ-данных эта выборка в 1999 году была увеличена до 318 объектов [3]. Наконец, в 2004 вышел из печати Каталог ближайших галактик, который содержит 451 объект [4].

На сегодняшний день существует большое количество астрономических баз данных. Наиболее популярными и развитыми базами данных о галактиках являются следующие: NASA Extragalactic Database (NED) — самая крупная в мире база данных, содержащая 163 млн. внегалактических объектов, перекрывающая диапазон от гамма-лучей до радиочастот и включающая данные сотен крупных обзоров неба и тысяч научных публикаций (<http://nedwww.ipac.caltech.edu/>); HyperLeda — база данных внегалактических объектов, которая содержит однородные сведения о 3 млн. объектов (<http://leda.univ-lyon1.fr/>); астрономическая база данных SIMBAD содержит исходные данные, кросс-идентификацию, библиографию и результаты измерений для более 5 млн.

астрономических объектов за пределами Солнечной системы (<http://simbad.u-strasbg.fr/simbad/>).

В последние годы стало очевидным, что наблюдательная космология остро нуждается в создании образцовой выборки галактик Местного объема, которая была бы достаточно представительной и не содержала существенного влияния избирательности. Изучение близких галактик играет особую роль в наблюдательной космологии. Только близкие галактики могут быть подробно исследованы, что невыполнимо на больших расстояниях; только в близких галактиках мы видим индивидуальные звезды, что позволяет изучать процессы звездообразования и использовать высокоточные методы определения расстояний; именно в близкой окрестности нам доступны для наблюдений карликовые галактики, которые составляют основную долю среди звездных систем. Поэтому очевидна необходимость создания базы данных о галактиках Местного объема, которая содержала бы многочисленную информацию о галактиках, расположенных в пределах 10 Мпк.

Таким образом, создание базы данных о галактиках Местного объема является актуальной задачей. Такая база, как совокупность структурированных и взаимосвязанных данных и методов, обеспечит систематизацию, информационную полноту и организационное пополнение данных. Организация и обеспечение быстрого и прозрачного Web-доступа к ним предоставит возможность оперативного отображения данных и может быть востребована широким кругом пользователей.

Существуют принципиальные различия между описанием базы данных и самой базой данных.

Описанием базы данных является ее схема. Схема создается в процессе проектирования базы. При этом предполагается, что она может модифицироваться, но достаточно редко. Однако содержащаяся в базе данных информация может меняться часто, например, при вставке новых данных или изменении существующих. Совокупность информации, хранящейся в базе данных в любой определенный момент времени, называется состоянием базы данных. Следовательно, одной и той же схеме базы данных может соответствовать множество ее различных состояний.

В основу схемы и состояния Базы Данных галактик Местного объема (LVG) нами был положен Каталог ближайших галактик — Catalog of Neighboring Galaxies [4]. Выборка объектов была обновлена и расширена до более чем 800 объектов. В Базу Данных LVG также были включены последние наблюдательные результаты, значительная часть которых получена на 6-метровом телескопе (БТА) сотрудниками лаборатории внегалактической астрофизики и космологии САО. В связи с этим был расширен набор наблюдательных параметров объектов по сравнению с Каталогом ближайших галактик. Также База Данных LVG пополнилась сводкой изображений всех галактик размером $6' \times 6'$ из обзоров SDSS и DSS в различных фильтрах. Все это нашло свое отражение как в схеме, так и в состоянии Базы Данных. В LVG предусмотрено хранение следующей информации: изображения объектов; измерения (размеры, степень сжатия, фотометрические параметры, скорости, типы, потоки); физические параметры (расстояния, светимости, пространственные характеристики, поглощение, потоки и звездные величины в различных фильтрах). Разработка базы данных проводилась с учетом возможности расширения хранимой информации.

В данной работе мы представляем разработанную физическую ER-модель и соответствующую схему структуры таблиц базы данных LVG, описание организации Web-доступа к ней.

2. ВЫБОР СИСТЕМЫ УПРАВЛЕНИЯ БАЗЫ ДАННЫХ

Выбор системы управления базы данных (СУБД) представляет собой сложную многопараметрическую задачу и является одним из важных этапов при разработке приложений баз данных. Выбранный программный продукт должен удовлетворять как текущим, так и будущим потребностям. Перечень требований к СУБД, используемых при анализе той или иной информационной системы, может изменяться в зависимости от поставленных целей [5]. Тем не менее можно сформулировать несколько критериев:

- мощные и надежные механизмы транзакций и репликации;
- возможность программирования базы данных и наличие расширяемой системы встроенных языков программирования;
- поддержка со стороны многих языков программирования: C/C++, Java, PHP, Perl;
- механизмы создания объектно-реляционных связей (наследование);
- легко расширяемая система типов данных;
- наличие свободной лицензии и стабильная поддержка программного продукта;
- возможность программирования Web-интерфейса;
- производительность;
- поддержка БД большого размера.

Наиболее функциональной и перспективной в этом плане нам представляется БД PostgreSQL. Она относится к категории объектно-реляционных систем управления базами данных (ОРСУБД). На сегодняшний день PostgreSQL считается наиболее развитой СУБД, распространяемой на условиях открытых исходных кодов [6, 7]. Ниже перечислены основные функциональные возможности, предоставляемые PostgreSQL ([8], <http://www.postgresql.org/>, <http://postgresmen.ru/>):

- контроль параллельного доступа, поддержка многопользовательского доступа;
- использование транзакций, оптимизация запросов;
- поддержка наследования и массивов;
- простота расширения — поддержка пользовательских операторов, функций, методов доступа и типов данных;
- строгое соответствие стандартам и полноценная поддержка SQL;
- правильность данных в базе обеспечивается проверкой целостности ссылок;
- гибкость интерфейса программирования приложений (Application Programming Interface, API), позволяющая создавать интерфейсы к PostgreSQL;



Рис. 1. Схема общей структуры таблиц данных базы.

- поддержка встроенных процедурных языков, таких как PL/pgSQL, Perl, Python, TCL;
- использование технологии Multi-Version Concurrency Control (MVCC) для предотвращения лишних блокировок (locking);
- использование архитектуры “клиент-сервер”
- с распределением процессов между пользователями;
- опережающая регистрация изменений (Write Ahead Logging, WAL) и репликация повышает надежность данных;
- индексы, система управления буферами памяти и кэширования, масштаби-

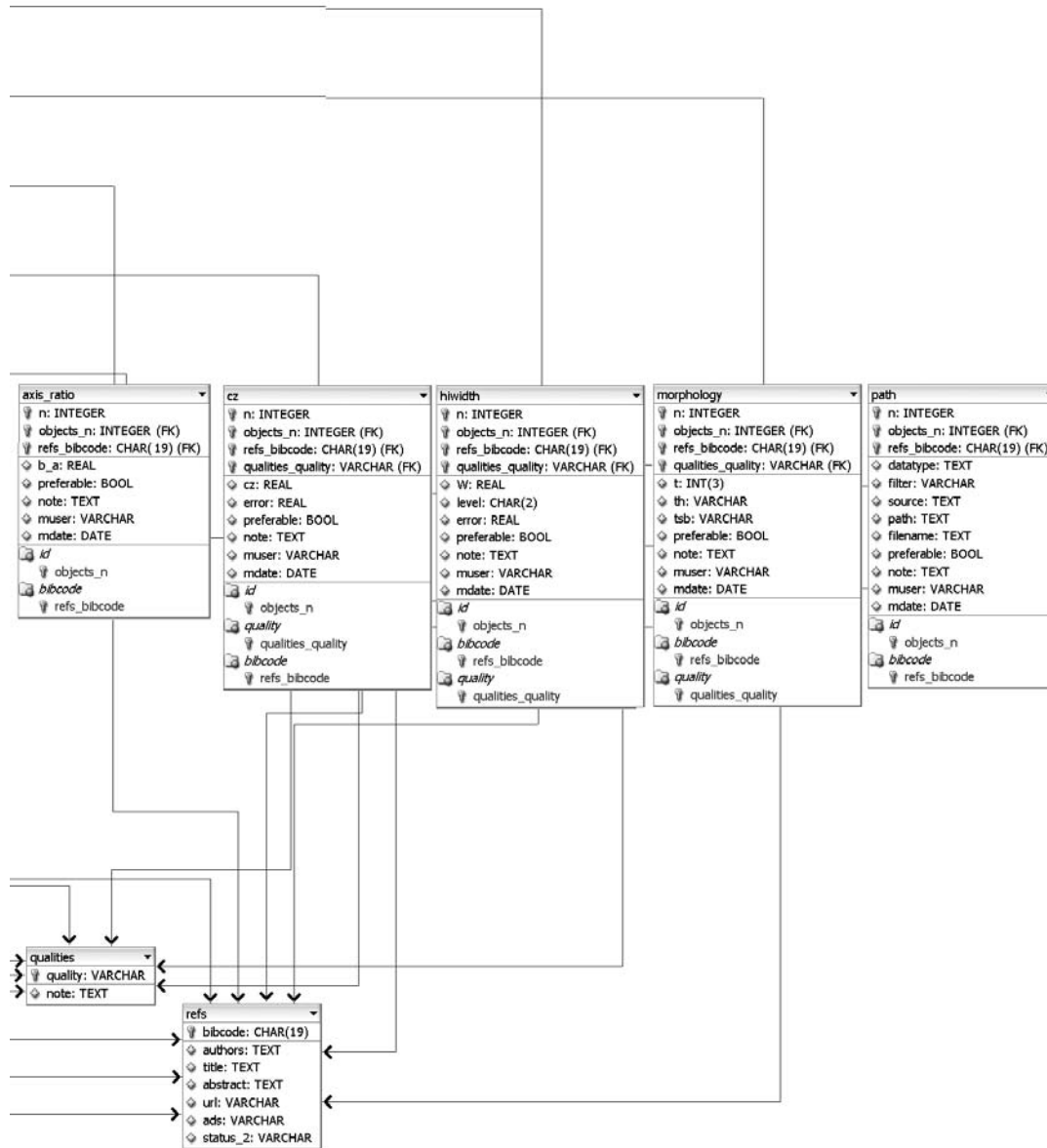


Рис. 1 Продолжение.

руемость обеспечивают производительность PostgreSQL;

- четырехуровневая система безопасности данных.

Как можно заметить, PostgreSQL обладает необходимой функциональностью, удовлетворяет требуемым критериям выбора и предоставляет

возможность создать легко расширяемую и надежную во всех смыслах базу данных в рамках существующих международных стандартов SQL. Возможности этой СУБД позволяют работать с различными типами данных и создавать специализированные типы данных в конкретной области знаний, что является немаловажным при работе с астрономическими данными. PostgreSQL изначально ориентирована на наличие интерфейсов.

В настоящее время в PostgreSQL реализован программный интерфейс для языков Object Pascal, Python, Perl, PHP, ODBC, Java/JDBC, Ruby, TCL, C/C+ и Pike, что позволяет с уверенностью говорить о надежных механизмах создания интерфейсов.

Важную роль здесь также играет операционное окружение, в котором реализуется поддержка базы данных как информационной системы, а также аппаратная база, на которой развертывается программное обеспечение.

3. РАЗРАБОТКА СТРУКТУРЫ БАЗЫ ДАННЫХ

Существует целый ряд важных моментов, принципиальным образом влияющих на процесс создания базы наблюдательных данных, которые необходимо учесть на первых этапах разработки и принимать во внимание при дальнейшей работе [9].

Организация поддержки обеспечивает регулирование поступления данных, корректное наполнение базы, необходимую реорганизацию как структуры, так и данных, и т.п. Для осуществления организационной поддержки и обеспечения доступа к архивным данным была разработана система политик доступа к БД на уровне групп пользователей. Системой политик было обусловлено дополнительно создание следующих групп пользователей с различным уровнем доступа к базе данных: *managers* (полный доступ); *editors* (выбор, редактирование, пополнение данных); *guest* (выбор данных). В процессе работы уровень доступа к данным определяется принадлежностью пользователя к определенной группе.

Создание группы пользователей (*guest*) продиктовано исключительно необходимостью организации доступа к базе данных через Web-интерфейс.

В настоящее время весьма актуально обеспечение информационной безопасности. Помимо введения групп пользователей, решением является использование безопасного подключения пользователей к базе. При проверке паролей пользователей применяется алгоритм шифрования md5.

Проблемы разнородности данных, форм их представления и информационной полноты очевидны, и на них приходится обращать внимание при создании баз данных различного уровня и содержания. На этапе проектирования базы данных все эти вопросы сыграли свою немаловажную роль.

При проектировании структуры базы данных мы использовали метод семантического моделирования, который представляет собой моделирование структуры данных, опираясь на смысл этих данных, и отношений между ними. Таким образом,

осуществляется детализация хранилищ данных. В качестве инструмента семантического моделирования используются различные варианты диаграмм сущность—связь (ER — Entity-Relationship) [10]. ER-диаграмма содержит информацию о сущностях системы и способах их взаимодействия, включает идентификацию объектов, важных для предметной области (сущностей), свойств этих объектов (атрибутов) и их отношений с другими объектами (связей). Основное достоинство метода состоит в том, что модель строится методом последовательных уточнений первоначальных диаграмм. Как уже отмечалось выше, база данных галактик Местного Объемы построена на основе БД PostgreSQL, поэтому, при построении ER-диаграммы мы учитывали особенности этой конкретной СУБД. В связи с этим, мы строим физическую ER-модель, которая учитывает такие особенности СУБД, как допустимые типы и наименования полей и таблиц, ограничения целостности и т.п.

При разработке ER-моделей на первом этапе необходимо определить следующую информацию:

- список сущностей предметной области;
- список атрибутов сущностей;
- описание взаимосвязей между сущностями.

При проектировании базы данных LVG первоначально нами были выделены следующие сущности: *objects* — объекты, входящие в БД LVG; *names* — имена галактик; *magnitude* — звездные величины; *fluxes* — величины потоков в различных фильтрах; *distance* — расстояния; *diameter, axis_ratio* — размеры; *cz* — лучевые скорости; *hiwidth* — ширины линий; *morphology* — морфологические типы; *path* — параметры, описывающие данные файлового архива; *refs* — источники данных. Атрибут сущности — именованная характеристика, являющаяся некоторым свойством сущности, соответственно, список атрибутов определяется для каждой сущности индивидуально. Для создания взаимосвязей между сущностями используется связь типа “один-ко-многим” — один экземпляр первой сущности связан с несколькими экземплярами второй сущности.

Как правило, все варианты диаграмм сущность—связь основываются на том, что рисунок всегда нагляднее текстового описания. Все такие диаграммы используют графическое изображение сущностей предметной области, их свойств (атрибутов) и взаимосвязей между сущностями. На Рис. 1 мы представляем разработанную нами физическую ER-модель базы данных LVG. В PostgreSQL логически связанные данные хранятся в двумерных структурах, называемых таблицами [8]. Каждая сущность

в модели представляет собой таблицу базы данных, каждый атрибут (изображается ромбиком) становится колонкой соответствующей таблицы (на этом этапе учитываются допустимые для данной СУБД типы данных и наименования столбцов) и изображается в виде прямоугольника с наименованием. Вводится некий избыточный набор атрибутов — первичный ключ сущности — значения их в совокупности являются уникальными для каждого экземпляра сущности. Сущность может иметь несколько различных ключей. В некоторых сущностях появляются дополнительные атрибуты — это ключевые атрибуты родительских таблиц, мигрировавших в дочерние таблицы для того, чтобы обеспечить связь между таблицами посредством внешних ключей. Таким образом, связи реализуются путем миграции ключевых атрибутов родительских сущностей и создания внешних ключей. Ключевые атрибуты изображаются на диаграмме со значком ключа, внешние ключи дополнительно помечены как FK (Foreign Key). Взаимосвязи между двумя сущностями изображаются линией со стрелкой, отношение читается вдоль линии, начиная со стрелки.

Все сущности на диаграмме образуют условно 4 уровня, что определяет структуру таблиц базы данных LVG. Сущность *objects* относится к **1-му уровню**, представляя основополагающую таблицу, содержащую идентификатор (*id*) объекта — первичный ключ данной сущности и внешний ключ для сущностей второго уровня, *base name* объекта — уникальный ключ, координаты и ряд вычисляемых параметров объекта (атрибутов). Таблица *objects* является родительской таблицей для таблиц второго уровня.

2-й уровень образуют таблицы основных наблюдательных параметров галактики: *names; magnitude; fluxes; distance; diameter, axis_ratio; cz; hiwidth; morphology; path*. На данный момент сформировано 10 таблиц 2-го уровня. Это таблицы параметров, значения которых могут корректироваться. Мы храним каждое значение со ссылкой на источник, одно из которых с помощью атрибута *preferable* считаем основным на данный момент и визуализируем. Существует возможность просмотра всех значений данного параметра. Эти таблицы содержат различный набор полей и строк (список атрибутов), который определяется каждым конкретным параметром. Обязательными атрибутами здесь являются: *id* — внешний ключ для связи с таблицей *objects*; атрибут, определяющий непосредственно основной параметр данной таблицы; атрибуты *preferable; note* (комментарии); *muser* (пользователь — автор данной записи); *mdate* (дата модификации записи).

Принимая во внимание разнородность наблюдательных данных, необходимо учесть максимально возможную их параметризацию и различные специфические особенности. В связи с этим для ряда атрибутов таблиц второго уровня необходимо образование дополнительных таблиц. Соответственно, **3-й уровень** образуют две сущности, *refs* (данные по источникам данных) и *qualities* (данные, определяющие качество параметра), являясь родительскими для таблиц второго уровня, таким образом, образуя также два обязательных атрибута этих таблиц. И, наконец, **4-й уровень** образуют 5 сущностей, соответственно, являющихся также родительскими для таблиц второго уровня. Выделение этого уровня обусловлено исключительно особенностями наблюдательных параметров. К примеру, таблица *magnitude* требует 3-х дополнительных характеристик: *magtypes* — тип параметра, *units* — единицы измерения, *passband* — фильтры; таблица *distance* — также 3-х таблиц: *method* — методы измерения расстояний, *measurement* — типы измерения, *passband*.

Такая структура таблиц, как и сами таблицы, могут быть модифицированы и расширены, что является чрезвычайно важным, и в чем нам видится заметное преимущество. Кроме этого, в дальнейшем в базу данных возможно дополнительно вносить характеристики, важные с точки зрения организации базы данных как информационной системы.

В базу LVG также входят представления (*view*), необходимые для организации Web-доступа, последовательности и функции.

4. ОРГАНИЗАЦИЯ WEB-ДОСТУПА К БАЗЕ ДАННЫХ LVG

Реализация методов доступа к архивным данным зависит от архитектуры используемой базы данных. Как уже говорилось, к преимуществам БД PostgreSQL относится гибкость API, позволяющая создавать программные интерфейсы. В настоящее время в PostgreSQL реализован программный интерфейс для достаточно широкого спектра программных языков.

Для реализации Web-доступа к базе данных LVG мы использовали программные языки PHP и JavaScript ([11–13], <http://www.php.net/>, <http://www.php.ru/>). Реализованный программный Web-интерфейс находится по адресу <http://www.sao.ru/lv/lvgdb>.

Список галактик, входящих в БД LVG представлен в двух видах:

1. Общий список имен галактик Местного Объемы с возможностью выхода на основную страницу объекта.

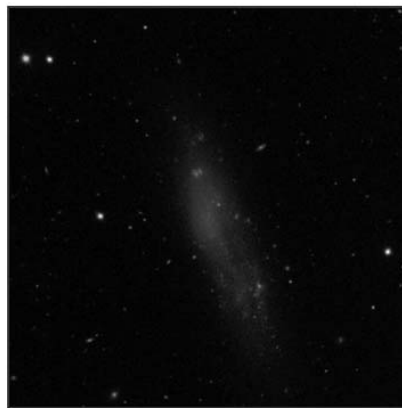
Catalog of LV Galaxies

Information View Search Authorization

The Data for Object "UGC08837"

Name	R.A.	Decl.	V_h	B_t
	EguJ2000.0			
UGC08837	13 54 45.1	+53 54 17	144	13.76

Object Names
DDO185
CGCG 272-013
MCG +09-23-017
PGC049448
UGC08837
HolmIV
Holmberg IV



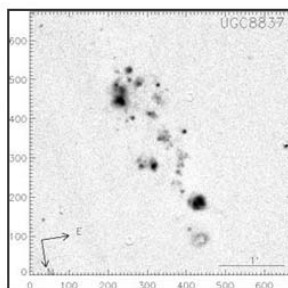
SDSS

▶ Invert image

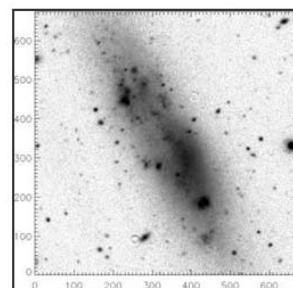
a_{Holm}	b/a	m_{FUV}	B_t	m_{Ha}	m_{21}	A_g	T	Morph	V_h	V_{LG}	W_{50}	D	Method
4.07	0.300		13.76	13.99	14.24	0.06	8		144	276	77	6.83	rgb

A_{25}	i	V_m	A_i	M_B	Σ_B	M_{HI}/L	Σ_{HI}	$\log M_{25}/M_{sun}$	M_{25}/L	Θ	MD
8.26	74	30	0.00	-15.90	25.0	0.91	6.8	9.07	3.32	0.0	M101

	Flux, ($\text{erg cm}^{-2} \text{s}^{-1}$)	$\lg[SFR],$ ($M_{sun} \text{yr}^{-1}$)
$H\alpha$	$(4013 \pm 9.8) \times 10^{-16}$	-1.61
UV		—



H α -continuum
FN657-SED607



H α -continuum
FN657+SED607



Рис. 2 Пример основной страницы объекта.

- Общий список имен галактик Местного Объем с сортировкой по координате R.A. (на эпоху 2000.0) и разбивкой на страницы по 25 объектов на каждой, также с возможностью выхода на основную страницу объекта.

Весь набор параметров объекта, находящихся в БД, сведен в единую WEB-страницу, образующую

визитную карточку объекта — основную страницу объекта (Рис. 2). На нем можно увидеть, что основная страница включает в себя:

- список существующих имен объекта;
- оптическое изображение галактики, для изображений SDSS реализована функциональная возможность преобразования и просмотра инвертированного изображения;

- таблицу, содержащую набор базовых наблюдаемых параметров;
- таблицу, содержащую набор вычисляемых параметров;
- таблицу, содержащую основные параметры, которые характеризуют светимость галактики в различных фильтрах и дают представление о темпах звездообразования в ней, включая изображения объекта в фильтре $H\alpha$.

Для каждой таблицы в LVG реализована возможность просмотра описаний параметров объекта по таблицам в целом или по конкретному параметру. Как уже упоминалось, значения параметров могут корректироваться, поэтому мы храним каждое из них с источником-публикацией, соответственно, существует возможность просмотра всех значений данного параметра в отдельном Web-окне (вкладке).

В рамках развития функциональных возможностей Web-интерфейса реализован поиск объектов в БД по имени, при этом предусмотрен поиск по полному имени, по маске имени, по координатам с заданным радиусом поиска. Результатом поиска является страница, содержащая таблицу найденных имен объектов и ряда основных параметров для каждого найденного объекта с возможностью выхода на его страницу.

Для дальнейшей работы непосредственно с данными БД реализована возможность авторизации в БД на уровне пользователей в соответствии с системой политик доступа к БД. Для этого Web-интерфейс LVG сделан двухслойным: гостевой и авторизованный пользователь. Авторизованному пользователю предоставляется возможность работы с именами объектов: редактирование существующих имен объектов и добавление новых имен к данному объекту с возможностью изменения базового имени объекта, добавление новых объектов в базу данных.

Проведено соответствие имен галактик с базами данных NASA Extragalactic Database (NED) и Lyon Extragalactic Database (LEDA) с основной страницы объекта LVG реализован выход на данные каждого объекта в этих базах.

5. ЗАКЛЮЧЕНИЕ

Проведенный комплекс работ включает в себя:

- создание базы данных галактик Местного объема — LVG;
- подбор и развертывание необходимой аппаратной и операционной базы;

- подбор базы данных, исходя из ее функциональных возможностей, и удовлетворяющей требованиям параметризации астрономических данных;
- анализ и реорганизацию имеющегося на первом этапе объема данных;
- разработку структуры базы данных и подбор программного обеспечения;
- разработку, написание и развертывание интерфейса Web-доступа к базе данных.

В результате впервые сформирована и сопровождается систематизированная и доступная база данных галактик Местного Объемы — LVG.

Расширенная выборка объектов и развернутый анализ свойств галактик Местного Объемы будут даны в следующей публикации.

Представленная база данных галактик Местного Объемы будет в значительной степени способствовать получению новых астрофизических результатов и повышению эффективности работы при подготовке к наблюдениям на БТА.

Использование форм хранения информации в рамках общепринятых международных стандартов позволит реализовать взаимодействие и обмен данными с другими базами данных.

БЛАГОДАРНОСТИ

В работе использованы базы данных HyperLEDA (<http://leda.univ-lyon1.fr>) и NED (<http://nedwww.ipac.caltech.edu>), данные обзора SDSS (<http://cas.sdss.org/dr7/en/>) и DSS/DSS-2 (<http://archive.eso.org/dss/dss>). Работа поддержана Министерством образования и науки Российской Федерации (государственные контракты 14.740.11.0901 и 16.518.11.7073), грантом РФФИ 11-02-00639 и российско-украинским грантом 11-02-90449.

СПИСОК ЛИТЕРАТУРЫ

1. R. C. Kraan-Korteweg and G. A. Tammann, *Astron. Nachr.* **300**, 181 (1979).
2. I. D. Karachentsev, *Astron. Astrophys. Trans.* **6**, 1 (1994).
3. I. D. Karachentsev, D. I. Makarov, and W. K. Huchtmeier, *Astronom. and Astrophys. Suppl. Ser.* **139**, 97 (1999).
4. I. D. Karachentsev, V. E. Karachentseva, W. K. Huchtmeier, and D. I. Makarov, *Astronom. J.* **4**, 2031 (2004).
5. Е. М. Балдин, Цикл статей о PostgreSQL, Linux Format, 85–91 (2006–2007).

6. О. Бартунов, <http://postgresrussia.org/articles/view/57>.
7. А. Шетухин, PostgreSQL vs MySQL, Системный администратор 7, 38 (2007).
8. John C. Worsley and Joshua D. Drake, *Practical PostgreSQL* (O'Reilly, 2002) [in Russian].
9. В. К. Кононов и В. Е. Панчук, Препринт САО 139Т (2000).
10. С. J. Date, *An Introduction to Database Systems*, 8th Edition (Addison-Wesley, 2004) [in Russian].
11. Luis Argerich et al., *Professional PHP4* (Apress, 2003) [in Russian].
12. E. Geshwinde and H.-J. Schonig, *PHP and PostgreSQL. Advanced Web programming* (SAMS, 2002) [in Russian].
13. А. Мазуркевич и Д. Еловой, *PHP: настольная книга программиста* (Новое знание, М., 2004).

OBSERVATIONAL DATABASE FOR STUDIES OF NEARBY UNIVERSE

E. I. Kaisina, D. I. Makarov, I. D. Karachentsev, S. S. Kaisin

We present the description of a database of galaxies of the Local Volume (LVG), located within 10 Mpc around the Milky Way. It contains more than 800 objects. Based on an analysis of functional capabilities, we used the PostgreSQL DBMS as a management system for our LVG database. Applying semantic modelling methods, we developed a physical ER-model of the database. We describe the developed architecture of the database table structure, and the implemented web-access, available at <http://www.sao.ru/lv/lvgdb>.

Keywords: *astronomical databases: catalogs—astronomical databases: surveys—galaxies: fundamental parameters*