

Система доступа к коллекции знаний по экологии региона

В.А. Лебедев, В.Г. Старкова, С.В. Брагин
Карельский научный центр РАН
г. Петрозаводск

Идея системы доступа к коллекции знаний, описываемой в докладе, состоит в том, чтобы построить своего рода тезаурус (структуру знаний) предметной области и использовать его для поиска интересующих фрагментов знаний. При этом требовалось обеспечить доступ несколькими путями: прохождением по структуре от корней к листьям; по алфавиту (набором первых букв названий вершин); с использованием словаря ключевых слов при обращении к названиям вершин или к спискам дескрипторов, характеризующих фрагменты знаний. Кроме того следовало обеспечить доступ к пространственно распределенным данным через географическую карту и/или отображение таких данных на карту.

В публикациях [1,2] рассмотрены вопросы структуризации знаний, построения и представления структуры (тезауруса), обсуждены некоторые аспекты построения распределенной, разнородной коллекции знаний. Данная публикация посвящена описанию системы доступа к коллекции и интерфейса, частично изложенных в [3].

Моделью структуры знаний является ациклический граф достаточно общего вида с помеченными вершинами, причем вершины - это термины и названия предметной области, а дуги - отношения классификации, агрегации или равнозначности. Пометы вершин указывают тип вершины и тип знаний, связанных с этой вершиной. Выделяются следующие типы вершин: корни, листья и срединные; последние подразделяются на единичные и множественные, которые, являясь названиями классов объектов, связаны с соответствующими списками названий (например, Озера - список названий озер).

В системе доступа предусмотрены следующие списки названий: растений, животных и микроорганизмов (русские и латинские названия), водных объектов (акваторий), территорий, населенных пунктов, лесных и сельских хозяйств, охраняемых территорий, химических элементов и соединений, горных пород, минералов, почв и

грунтов, типов экосистем, биоценозов, местообитаний, форм и типов рельефа, различных органов, систем и частей организмов, ряд шкал оценок и классификаций.

Знания подразделяются на общие и конкретные, причем общие знания (толкования понятий и их взаимосвязей, рефераты, описания методик экспериментов и расчетов и др.) могут сопровождать любую вершину, конкретные связываются только с листьями. Конкретные знания могут представляться в виде текстов, таблиц, географических карт, снимков, программ решения задач. Каждый из фрагментов конкретных знаний может иметь описание в виде списка дескрипторов.

На основе структуры знаний и множества списков дескрипторов формируется словарь ключевых слов, представляющий из себя список неизменяемых частей всех слов, встречающихся в структуре и множестве дескрипторов. Кроме того составляется словарь синонимов посредством выделения из того же множества слов синонимических гнезд.

Компоненты структуры знаний, а именно: список всех вершин с пометами, список дуг, таблица соответствий внешних и внутренних имен фрагментов, таблица списков дескрипторов, образуют единую базу данных, включающую перечисленные компоненты в качестве таблиц. Это позволяет однократно и согласованно вносить изменения и дополнения в базу данных и обеспечивает формирование и контроль ее ограничений целостности. Основным условием при импорте фрагментов знаний является наличие пути от импортируемого фрагмента к одной из вершин-корней. Условие целостности при удалении фрагмента - вычеркивание всех строк таблиц, содержащих внутреннее имя фрагмента и фрагмента пути до первой вершины, имеющей исходящие дуги, кроме вычеркиваемой.

Основные признаки (пометы) вершин-листьев представлены в таблице соответствия их внешних и внутренних имен. Внешнее имя - это название вершины-листа; внутреннее имя показывает путь, имя и формат файла. Далее следует перечисление слоев карты, связанных с базой данных, имя карты и ключевое поле шейп-файла, список номеров задач (функций), определенных на данной БД.

Таблица задач имеет следующую структуру: номер по порядку, русское название задачи, тип результата, имя файла программного модуля, список аргументов (для задач, определенных на конкретной БД) или список полей

©Вторая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
26-28 сентября 2000г., Протвино

для передачи аргументов, поле для записи формулы сравнительно простых арифметических вычислений на конкретной БД. Тип результата показывает, что результат - это либо новая колонка, либо новая строка, либо новая таблица. Соответственно этому название задачи пересылается либо в название колонки, либо в название строки при формировании таблицы - результата.

Задание аргументов для сложных задач формирования новых таблиц, например, геостатистики, выполняется по дополнительному меню, вызов которых определяется по графе "Тип результата".

Таблица дескрипторов состоит из двух колонок: имени файла фрагмента знаний и списка его дескрипторов. При этом, если файл - таблица базы данных, то порядок написания списка дескрипторов - это порядок расположения колонок в этой таблице, в остальных случаях порядок дескрипторов несуществен.

Пометы вершин используются следующим образом. Программа, выполняющая сканирование графа, контролирует тип очередной вершины и в зависимости от этого разрешает или запрещает дальнейшее движение по графу структуры, вызов пояснений, списка названий или фрагмента конкретных знаний. При вызове списка названий инициируется функция формирования предварительного фильтра отбора записей из БД, который впоследствии используется для формирования окончательного фильтра.

Вызов конкретного фрагмента инициирует ветвь программы, которая по типу знаний (формату файла) назначает соответствующий браузер или программу манипулирования данными. Например, если тип знаний "база данных dbf, dbc или иная", то инициируется ветвь манипулирования базой данных. Предварительно высвечивается меню для выбора действий над базой данных: Просмотр, Дополнение, Изменение, Задачи, Диаграммы. Далее вызывается меню для задания параметров манипулирования (три первых пункта), или список задач, определенных на данной БД, или мастер диаграмм в зависимости от выбранного пункта.

Если база данных содержит пространственные данные и связана с картой, то инициируется ветвь программы вызова и манипулирования картой, представляющая из себя модуль OLE ActiveX. Блоки сложных функций манипулирования картой такие, как поиск объекта по названию, выборка, раскраска оформлены в виде DLL-модулей. Доступ к удаленным базам данных осуществляется в среде файловой системы локальной сети через ODBC по их представлениям, зарегистрированным в структуре знаний. Такие базы данных могут быть связаны с тематическими слоями карты, хранящимися на том же компьютере.

Блок инициации задач читает список определенных на данной БД задач, делает выборку из таблицы задач и формирует меню задач.

Если выбранный фрагмент - текст, то расширение имени его файла или тип адреса (например, WWW-адрес), используется для вызова соответствующего браузера. В системе предусмотрено использование текстов трех типов: txt-файлов (небольшие документы типа мемо-полей), документы в формате Word.doc и документы в формате HTML. Формат HTML предпочтителен для сравнительно больших документов с оглавлениями и иллюстрациями.

На базе указанных структур построена иерархическая

система оконных меню для обеспечения доступа к различным фрагментам знаний. На верхнем уровне предоставляется выбор направлений взаимодействия с коллекцией: Поиск, Импорт, Создание виртуальных баз данных (БД), Печать, Выход. Доступ для использования коллекции осуществляется через п. "Поиск". Высвечивается меню вариантов поиска: 1) по объектам и темам, 2) по дескрипторам. При выборе первого варианта высвечивается меню, показанное на рис. 1 (рисунки в конце статьи). На нем в правом окне виден фрагмент списка всех вершин структуры знаний. Для выбора конкретной вершины предоставляется три возможности: 1) следовать по графу структуры от корней к листьям, 2) набрать в верхней строке первые буквы названия вершин, 3) вызвать словарь ключевых слов и выбрать там слова, ключевые для интересующего названия вершины. Общие знания, относящиеся к данной вершине, вызываются нажатием кнопки "Пояснения". Конкретные фрагменты знаний вызываются через листья, которые отмечены знаком "*" справа от названия. На рис. 1 приведен пример поиска фрагмента по первым буквам "Биоц" "Биоценозы" "Б. наземные" "Растительность лесных хозяйств". Важной особенностью является возможность движения по графу структуры как от корней с листьями, так и от листьев к корням.

При выборе поиска фрагмента по дескрипторам вызывается словарь ключевых слов, формируется фильтр, по которому осуществляется поиск в множестве списков дескрипторов. Результат поиска - список названий релевантных фрагментов, из которого пользователь выбирает нужный (см. рис. 2).

Дальнейшие действия с выбранным фрагментом определяются его типом знаний. Так, текстовый фрагмент может начинаться с оглавления, которое позволяет просматривать его выборочно в произвольном порядке. Географическая карта вызывается с дополнительной серией меню действий над картой таких, например, как поиск объекта, разбиение объектов некоторого слоя на классы и их раскраска, увеличение-уменьшение масштаба, оформление для печати и печать (рис. 3).

В случае необходимости действий по конкретизации запроса высвечивается дополнительное меню. Например, на рис. 4 представлено меню функции поиска объекта по названию. Набираются первые буквы названия, что вызывает соответствующий фрагмент списка названий, из которого выбирается требуемое. В результате на экран выдается фрагмент карты с отмеченным искомым объектом.

Действия над базами данных определяются последовательностью меню. Сначала определяются функции: Просмотр, Добавление, Изменение (записей), Задачи, Диаграммы. При просмотре (добавлении, изменении) появляется меню действий: проекция, селекция, сортировка (см. рис. 5), при помощи которого строится фильтр отбора колонок и записей из БД и ключ сортировки результата. При этом для отбора записей (или добавления, изменения) по колонкам, содержащим названия, используются соответствующие словари (списки названий).

Отдельный интерфейс разработан для п. "Задачи". Для каждой БД вызывается меню выбора задач, в котором виден список задач и функций, определенных для данной БД, двух категорий: определенных на всех БД или только

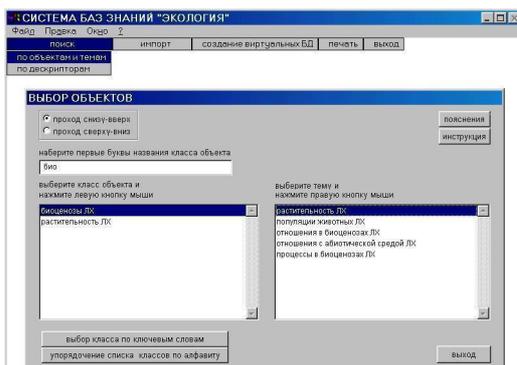


Рис. 1

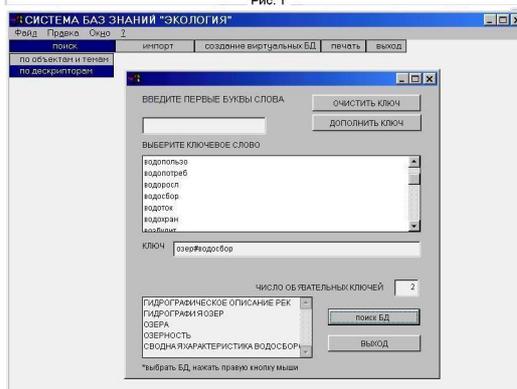


Рис. 2

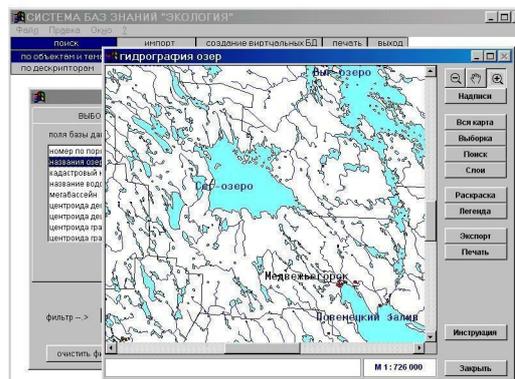


Рис. 3

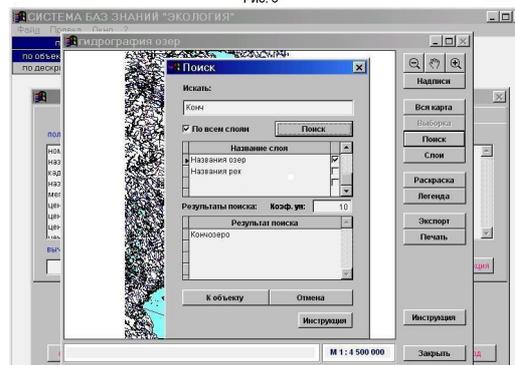


Рис. 4

на части их множества. Пользователь выбирает некоторый набор задач и функций, после чего появляется меню (рис. 6), в левом окне которого виден список названий колонок таблицы, а справа - список выбранных функций. Выбрав аргументы для очередной функции из числа существующих колонок БД и /или вновь вычисляемых колонок или строк (типа сумма, число строк, дисперсия и т.д.), отмечаем ее и в нижнем окне видим формулу для исполнения. При этом система следит, чтобы последовательность функций и задач была корректной (по определенности аргументов).

Программы решения сложных задач (например, моделирование какого-либо процесса) рассматриваются как самостоятельные фрагменты знаний и вызываются для исполнения по названиям вершин-листьев. При этом, как правило, такие задачи имеют собственные меню для задания аргументов, граничных значений и т.п.

Выбор направления "Импорт" вызывает меню конкретизации задачи: Импорт БД, Импорт текста, Импорт задачи, Импорт тематических слоев карт, при выборе которых инициируется соответствующая серия меню. Так, при импорте БД необходимо получить ее формат и адрес, русское название БД и названия колонок таблиц, ключи сортировки, названия вершин структуры знаний для регистрации их в структуре, а также некоторые другие сведения (например, для связи с картой, если БД содержит пространственные данные). При импорте задач необходимо получить сведения о входе и выходе, о виде и расположении меню исходных данных, а также сведения для включения в структуру знаний. Импорт текстов требует указания формата и браузера, а также адресов.

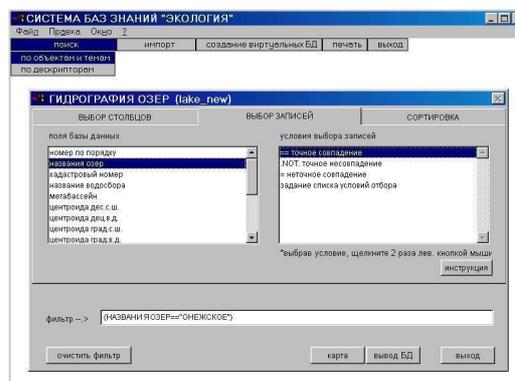


Рис. 5

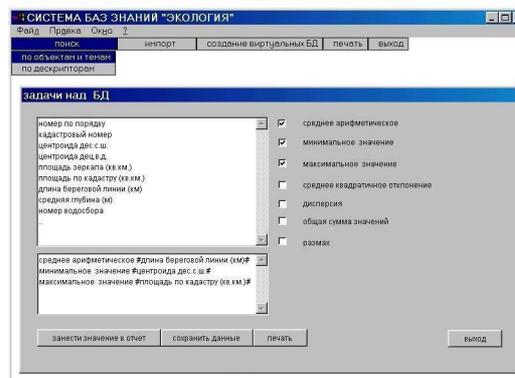


Рис. 6

Направление "Создание виртуальных БД это программирование заданий на обработку БД (например, SQL-запросов) и регистрация задания в структуре знаний.

Печать выборок из текстов, баз данных или фрагментов карты может выполняться без специального оформления непосредственно из меню соответствующих функций. При необходимости оформить выдаваемый документ используется направление "Печать", выбор которого позволяет оформить либо печатаемую таблицу при помощи конструктора таблиц, либо фрагмент карты при помощи графического редактора, либо текстовый документ.

Отметим, что данная система открыта для развития, благодаря тому, что все компоненты структуры знаний оформлены как базы данных и зарегистрированы в ней. Это дает возможность изменить, добавить и удалить любые фрагменты знаний, соответственно изменяя структуру, словари, списки задач и пр. Фактически программная оболочка может быть использована также для создания других коллекций знаний.

Список литературы

- [1] Лебедев В.А., Старкова В.Г., Брагин С.В., Табаков Н.А., *О структуре коллекции знаний по экологии региона*, Труды I Всероссийской конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", С.-Петербург, 1999.
- [2] Лебедев В.А., *Модель структуризации знаний по экологии*, Труды Института прикладных математических исследований, Вып.1, Петрозаводск, 1999.
- [3] Лебедев В.А., Старкова В.Г., Брагин С.В., Табаков Н.А., *Информационная система комплексного отображения экологии региона*, Труды Института прикладных математических исследований, Вып.1, Петрозаводск, 1999.