

О структуре коллекции знаний по экологии региона

В.А. Лебедев, В.Г. Старкова, С.В. Брагин, Н.Ф. Табаков

Состояния и изменения экосистем региона отображаются в большом количестве текстов, таблиц, карт, схем, моделей и т.п. документов. Для ориентирования в них требуется выполнение определенной структуризации знаний. Предложена модель структуризации знаний с целью построения коллекций баз данных и знаний (БДЗ) по экологии региона и обеспечения доступности их в существующих терминах и названиях. Предложенная модель обеспечивает управление поиском информации в зависимости от некоторых ее признаков, контроль полноты и непротиворечивости.

В основу модели положены следующие соглашения. Экосистема региона имеет мозаичную структуру, т.е. в ней выделяется множество частных экосистем, приуроченных к определенным территориям или акваториям. Компоненты экосистем, абиотические и биологические (объекты, вещества, энергии, организмы и т.д.) находятся в устойчивых связях и отношениях, которые определяют структуру знаний. Это отношения типа классификаций (род-вид), агрегаций (целое-часть) и равнозначности (в частности, синонимии). Территории и акватории имеют определенное местоположение на поверхности земли и соответствующее отображение на географической карте; это обеспечивает их идентификацию по названиям, в частности, по координатам. Отношения между компонентами экосистем на территориях (акваториях) представляются как отношения между терминами и названиями. При этом термины и названия можно трактовать как вершины, а отношения между ними как дуги некоторого графа. Точнее, модель структуризации знаний представляет из себя ациклический плоский граф с помеченными вершинами. Причем выделяется несколько типов помет, а именно, указатель ранга вершин, указатель типа вершин и указатель типа знаний, связанных с вершинами. Построение графа структуры знаний по экологии не тривиальная задача, необходимо обеспечить его адекватность, полноту и непротиворечивость, соответствующие реалиям предметной области и целям создания коллекции знаний. Для этого необходимо выбрать подходящее представление графа. Таким представлением является перечисление дуг, обозначаемых парами смежных вершин. Построение будет состоять из нескольких эта-

пов. Сначала необходимо построить перечисление всех узлов - структуру смежности. Далее произвести упорядочивание множества узлов по темам и по лексикографии. После этого следует определить ранг каждого узла и затем к полученной структуре применить алгоритм прохождения всех путей (поиск в глубину). При этом должны быть распечатаны списки корневых и концевых вершин, что позволит определить относительную полноту структуры, а также распечатать список дуг, направленных от вершин более высокого ранга к вершинам более низкого ранга. Последний список является показателем возможной противоречивости структуры; если список пустой, то структура непротиворечива.

Для компьютерной реализации структурного графа необходимо выбрать подходящее его конструктивное представление. Поскольку структурный граф содержит многоаспектно помеченные вершины, то в качестве такого представления принято множественное отношение, или таблица, вид которой показан ниже.

Название вершины истока	Название вершины притока	Тип вершины притока	Тип БДЗ	Имя и путь

Табл.1.

Представление структурного графа в виде таблицы позволяет применить для его построения, контроля и использования средства какой-либо реляционной системы управления базами данных (СУБД). Это удобно, так как описание структуры и состояний компонентов экосистем также часто удобно представлять в виде множественных отношений, что обеспечивает возможность разработки единой технологии взаимодействия совокупности реляционных баз данных, как реальных, так и виртуальных (Виртуальная БД представляется в системе программой вычисления соответствующей таблицы. Такие программы могут быть и реализациями математических моделей).

К преимуществам предложенной модели структуризации знаний по экологии относится следующее.

Возможность построения семантической структуры предметной области.

Построенная структура может быть проверена на полноту и непротиворечивость при помощи специального алгоритма.

Данная модель в отличие от традиционных представлений графов обеспечивает прямой доступ к любой вер-

Первая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
19 - 21 октября 1999 г., Санкт-Петербург

шине, а не только к корневым. Она предоставляет возможность строить запросы правильные по построению.

Наличие в модели различных помет вершин обеспечивает управление выводом как при построении запроса, так и при формировании ответа с учетом определенных признаков информации.

Представление о разработанной структуре можно получить по следующим фрагментам. Корневыми вершинами объявлены понятия: экосистемы, биоразнообразие, флора, фауна, антропогенные воздействия. Очевидно, что объемы этих понятий пересекаются, наиболее полное понятие - экосистемы. Принята следующая структура описания экосистем. Они подразделяются на наземные, почвенные и водные. Каждый из этих типов привязан к территории или акватории. При этом территория (акватория) имеет различные разбиения на участки. На территориях выделяются административные, хозяйственные и природные разбиения, соответственно: районы, волости и населенные пункты, лесхозы, сельские хозяйства и охраняемые территории, водосборы, ландшафты и ареалы популяций. Акватории подразделяются на водоемы и водотоки с выделением их частей по конфигурации: заливы, губы, литорали, батигали, пелагиали и т.д. Все участки и части имеют собственные имена.

Собственно экосистемы, приуроченные к определенным участкам и частям, описываются тремя блоками: абиотическая среда, биоценоз, антропогенные воздействия. Абиотическая среда включает космос, магнитосферу Земли, атмосферу и климат, литосферу и гидросферу. Биоценоз описывается указанием видового и генотипического состава населения (растительного, животного и микробиологического), состава внутривидовых и межвидовых отношений и отношений с абиотической средой, состава социальных групп: синузий, консорциев, частных ценозов и описания процессов. Последние могут быть представлены траекториями, моделями или текстовыми описаниями. Антропогенные воздействия подразделяются на глобальные, региональные и локальные, а по видам воздействий на загрязнения и прочие воздействия. Если загрязнения могут быть глобальными, региональными и локальными, то прочие воздействия в основном локальные. К ним относятся, в частности, рубки леса, лесовосстановление, защита леса, мелиорация земель, охота, рыболовство, интродукция растений и животных, регулирование численности и др. воздействия.

Таким образом при описании экосистем задействованы все их компоненты, в том числе флора, фауна, биоразнообразие и антропогенные воздействия. Зачем нужно было опускать эти компоненты до корневого уровня? Каждое из этих понятий по объему и содержанию больше того содержания, которое используется при описании частных экосистем. Это резервное содержание может быть использовано для выработки оценок состояний и изменений экосистем. Вместе с тем значительная часть содержания уже представлена при описании экосистем. Это неравенство содержания и пересечение упомянутых понятий определяет модель структуры.

Структура описания экосистем, как показано выше, представляет из себя дерево. Аналогично этому можно показать, что структуры описаний биоразнообразия, флоры, фауны и антропогенных воздействий также древовидны. Из-за пересечения этих понятий их деревья частично накладываются друг на друга, порождая сложную структуру, которая отображается ациклическим плоским графом, описанным в начале статьи.

Очевидно корневыми и более высокие по уровню по-

нятия абстрактны и требуют определенного толкования. Это первый и наиболее абстрактный уровень знаний в коллекции. Более конкретные знания связаны с висячими вершинами. Фрагменты структуры знаний показаны в таблице Табл.2.

Разработанная структура коллекции знаний по экологии в настоящее время включает около 1000 терминов и названий классов, объектов, свойств, отношений и процессов.

Кроме того в состав структуры включены словари, охватывающие основные классы названий: растений, животных, микроорганизмов, водных объектов, территорий, поселений, лесных и сельских хозяйств, охраняемых территорий, химических элементов и соединений, горных пород, минералов, грунтов и почв, типов экосистем, биоценозов, местообитаний, форм и типов рельефа, названий характеристик, ряд шкал оценок и классификаций различных компонентов, а также синонимические гнезда многих терминов и названий. Всего более 20 тыс. единиц.

При помощи этих словарей осуществляется поиск конкретных объектов и записей в коллекции.

В разработанную структуру естественным образом вписываются картографические материалы. Они подразделяются на две категории: обзорные и узко тематические. Обзорные карты могут быть многослойными и предназначены для фиксации и последующего обозрения справочной информации об объектах, отображенных на карте. Например, сведения о реках, озерах, лесах, болотах, населенных пунктах и т.д. Узко тематические карты обычно содержат небольшое количество слоев и предназначены, например, для отображения ареалов популяций, оценок продуктивности экосистем и др. тем. Обычно они связаны с тематическими базами данных.

Различные сведения в коллекции кроме текстов и таблиц могут содержать разнообразные изображения, привязанные к соответствующим местам или названиям в текстах или таблицах. Изображения хранятся в экономном графическом формате.

Важную роль в системе экологических знаний играют математические модели. Они могут быть представлены в виде текстов (неидентифицированные модели) или программ, реализующих идентифицированные для региона модели. Представление программ в коллекции возможно либо в виде виртуальных баз данных, либо в виде процедурных модулей. В последнем случае модуль должен быть оформлен как объект с прямым указанием области определения. Совокупность этих модулей регистрируется в специальной базе данных. Доступность модуля для исполнения определяется по списку названий понятий, заданному в модуле, сравнением с именем выбранного фрагмента знаний. Результаты моделирования высвечиваются для просмотра и могут быть импортированы в коллекцию.

Разработанная структура положена в основу программной оболочки для построения и просмотра коллекций знаний по экологии региона. Просмотр знаний в коллекции выборочный. Предусмотрены следующие способы поиска фрагментов знаний:

- прохождением по структурному графу (поиск в глубину);
- поиск терминов и названий по первым буквам;
- поиск фрагментов по их названиям при помощи словаря ключевых слов;
- поиск фрагментов по спискам их дескрипторов (в частности, по названиям в колоннитулах таблиц) при помощи словаря ключевых слов;

поиск фрагментов указанием объектов, отображенных на обзорных или тематических картах.

Для построения коллекции знаний предусмотрен ряд функций, учитывающих особенности экологической информации и разнородность источников ее получения. Предусмотрены следующие функции:

импорт удаленных баз данных или их представлений, распределенных в локальной сети;

импорт текстовых файлов;

импорт изображений с привязкой их к текстам или БД;

импорт обзорных и тематических карт с привязкой их к текстам или БД;

импорт программных модулей;

пополнение и изменение структурного графа в соответствии с импортируемыми фрагментами;

пополнение и изменение собственных БД и текстов;

пополнение и изменение словарей.

Создание БД, текстовых и графических файлов и программ не входит в задачу данной оболочки и должно выполняться при помощи подходящих инструментальных средств: СУБД, текстовых и графических редакторов, языков программирования.

Система реализована на базе Visual FoxPro и Map Object в среде локальной сети, управляемой ОС Novel в режиме "клиент-сервер". Клиенты сети могут работать в среде Windows и обладать собственной локальной системой баз данных, работающей под управлением одной из стандартных СУБД (Access, FoxPro, Paradox, Excel, Lotus и др.).

Собственно программная часть комплекса в виде отдельного exe-файла и библиотечных программ инсталлирована на сервере локальной сети и доступна для любого пользователя.

Взаимодействие пользователей с системой осуществляется через комплекс многооконных меню. При этом взаимодействие происходит в основном на основе выбора значений из соответствующих словарей терминов и названий. Набор на клавиатуре используется только указания первых букв соответствующих названий, а также для дополнения словарей при импорте удаленных баз данных и моделей.

Дистрибутив комплекса занимает 8 Мб пространства жесткого диска. Exe - модуль программы, не считая программ моделей и виртуальных БД, занимает 2,6 Мб памяти жесткого диска. Для работы программе требуется 6,5 Мб оперативной памяти и оптимально 2 Мб буферной памяти для информации. Информационное наполнение системы (структурный граф и словарь названий без БДЗ) занимает более 2 Мб памяти жесткого диска. Количество включенных в систему БДЗ и подключаемых клиентов ограничивается только возможностями компьютеров.

<p>*ЭКОСИСТЕМЫ НАЗЕМНЫЕ ЭКОСИСТЕМЫ АДМИНИСТРАТИВНЫЕ ТЕРРИТОРИИ РЕСПУБЛИКА КАРЕЛИЯ АТМОСФЕРА СОСТАВ АТМОСФЕРЫ ПРИРОДНЫЙ СОСТАВ АТМОСФЕРЫ ПРИРОДНЫЙ СОСТАВ АТМОСФЕРЫ СОСТАВ АТМОСФЕРЫ ЗАГРЯЗНЯЮЩИЕ АТМОСФЕРУ ВЕЩЕСТВА ЗАГРЯЗНЯЮЩИЕ АТМОСФЕРУ ВЕЩЕСТВА ЗАГРЯЗНЯЮЩИЕ АТМОСФЕРУ ВЕЩЕСТВА ЗАГРЯЗНЯЮЩИЕ АТМОСФЕРУ ВЕЩЕСТВА МОНИТОРИНГ АТМОСФЕРНЫХ ЗАГРЯЗНЕНИЙ БИОСФЕРНЫЙ МОНИТОРИНГ БИОСФЕРНЫЕ СТАЦИОНАРЫ БИОСФЕРНЫЕ СТАЦИОНАРЫ БИОСФЕРНЫЕ СТАЦИОНАРЫ БИОСФЕРНЫЕ СТАЦИОНАРЫ МОНИТОРИНГ АТМОСФЕРНЫХ ЗАГРЯЗНЕНИЙ ИМПАКТНЫЙ МОНИТОРИНГ ПРЕДПРИЯТИЯ ПРЕДПРИЯТИЯ ПРЕДПРИЯТИЯ ПРЕДПРИЯТИЯ ПРЕДПРИЯТИЯ ПРЕДПРИЯТИЯ ПРЕДПРИЯТИЯ ПРЕДПРИЯТИЯ ЗАГРЯЗНЕНИЕ ТЕРРИТОРИЙ ЗАГРЯЗНЕНИЕ ТЕРРИТОРИЙ ЗАГРЯЗНЕНИЕ ТЕРРИТОРИЙ ЗАГРЯЗНЕНИЕ ТЕРРИТОРИЙ ЗАГРЯЗНЕНИЕ ТЕРРИТОРИЙ ПРЕДПРИЯТИЯ ЗАГРЯЗНЕНИЕ ВОДОЕМОВ ЗАГРЯЗНЕНИЕ ВОДОЕМОВ ЗАГРЯЗНЕНИЕ ВОДОЕМОВ ЗАГРЯЗНЕНИЕ ВОДОЕМОВ РЕСПУБЛИКА КАРЕЛИЯ РАСТЕНИЯ РАСТЕНИЯ НАЗВАНИЯ РАСТЕНИЙ НАЗВАНИЯ РАСТЕНИЙ РАСТЕНИЯ ФИТОИНДИКАТОРЫ ФИТОИНДИКАТОРЫ ФИТОИНДИКАТОРЫ ФИТОИНДИКАТОРЫ ВОДНЫЕ ФИТОИНДИКАТОРЫ ВОДНЫЕ ФИТОИНДИКАТОРЫ ВОДНЫЕ ФИТОИНДИКАТОРЫ ХИМИЧЕСКИЙ СОСТАВ ЛЕСНЫХ РАСТЕНИЙ ХИМИЧЕСКИЙ СОСТАВ ЛЕСНЫХ РАСТЕНИЙ ФИТОИНДИКАТОРЫ НАЗЕМНЫЕ ХИМИЧЕСКИЙ СОСТАВ КОРМОВЫХ РАСТЕНИЙ ХИМИЧЕСКИЙ СОСТАВ КОРМОВЫХ РАСТЕНИЙ ФИТОИНДИКАТОРЫ НАЗЕМНЫЕ ФИТОИНДИКАТОРЫ НАЗЕМНЫЕ РАСТЕНИЯ РАСТЕНИЯ РАСТЕНИЯ РАСТИТЕЛЬНОСТЬ РАСТИТЕЛЬНОСТЬ НАЗЕМНАЯ БИОЦЕНОЗЫ НАЗЕМНЫЕ БИОЦЕНОЗЫ ЛХ РАСТИТЕЛЬНОСТЬ ЛХ РАСТИТЕЛЬНОСТЬ ЛХ РАСТИТЕЛЬНОСТЬ ЛХ РАСТИТЕЛЬНОСТЬ ЛХ РАСТИТЕЛЬНОСТЬ ЛХ РАСТИТЕЛЬНОСТЬ ЛХ РАСТИТЕЛЬНОСТЬ ЛХ БИОЦЕНОЗЫ ЛХ ПОПУЛЯЦИИ ЖИВОТНЫХ ЛХ ПОПУЛЯЦИИ ЖИВОТНЫХ ЛХ ПОПУЛЯЦИИ ЖИВОТНЫХ ЛХ ПОПУЛЯЦИИ ЖИВОТНЫХ ЛХ ПОПУЛЯЦИИ ЖИВОТНЫХ ЛХ ПОПУЛЯЦИИ ЖИВОТНЫХ ЛХ ПОПУЛЯЦИИ ЖИВОТНЫХ ЛХ ПОПУЛЯЦИИ ЖИВОТНЫХ ЛХ БИОЦЕНОЗЫ ЛХ ОТНОШЕНИЯ В БИОЦЕНОЗАХ ЛХ МЕЖВИДОВЫЕ ОТНОШЕНИЯ ЛХ ПАРАЗИТИЗМ НАЗЕМНЫЙ ПАРАЗИТИЗМ НАЗЕМНЫЙ МЕЖВИДОВЫЕ ОТНОШЕНИЯ ЛХ МЕЖВИДОВЫЕ ОТНОШЕНИЯ ЛХ</p>	<p>НАЗЕМНЫЕ ЭКОСИСТЕМЫ АДМИНИСТРАТИВНЫЕ ТЕРРИТОРИИ РЕСПУБЛИКА КАРЕЛИЯ АТМОСФЕРА СОСТАВ АТМОСФЕРЫ ПРИРОДНЫЙ СОСТАВ АТМОСФЕРЫ МОНИТОРИНГ КИСЛОРОДА* МОНИТОРИНГ УГЛЕКИСЛОГО ГАЗА* ЗАГРЯЗНЯЮЩИЕ АТМОСФЕРУ ВЕЩЕСТВА ТРАНСГРАНИЧНЫЕ ПЕРЕНОСЫ* СУММАРНЫЕ ВЫБРОСЫ В АТМОСФЕРУ* ПРЕДЕЛЬНО ДОПУСТИМЫЕ КОНЦЕНТРАЦИИ ЗАГРЯЗНЕНИЙ В АТМОСФЕРЕ* МОНИТОРИНГ АТМОСФЕРНЫХ ЗАГРЯЗНЕНИЙ БИОСФЕРНЫЙ МОНИТОРИНГ БИОСФЕРНЫЕ СТАЦИОНАРЫ МОНИТОРИНГ ЗАГРЯЗНЕНИЙ АТМОСФЕРЫ* КИСЛОРОДНО-УГЛЕРОДНЫЙ БАЛАНС* МЕСТОПОЛОЖЕНИЕ СТАЦИОНАРОВ* ЗАГРЯЗНЯЮЩИЕ ВЕЩЕСТВА В ПОВЕРХНОСТНЫХ ВОДАХ* ИМПАКТНЫЙ МОНИТОРИНГ ПРЕДПРИЯТИЯ СВОДНАЯ ХАРАКТЕРИСТИКА* НАЗВАНИЯ ПРЕДПРИЯТИЙ* ЭКОЛОГИЧЕСКИЕ УГРОЗЫ* ОТРАСЛИ ХОЗЯЙСТВА* СУММАРНЫЕ АТМОСФЕРНЫЕ ВЫБРОСЫ* СТОЧНЫЕ ВОДЫ* ТВЕРДЫЕ ОТХОДЫ ЗАГРЯЗНЕНИЕ ТЕРРИТОРИЙ КАРТЫ ЗАГРЯЗНЕНИЙ СОДЕРЖАНИЕ ЗАГРЯЗНЯЮЩИХ ВЕЩЕСТВ В ПОЧВАХ* СОДЕРЖАНИЕ ЗАГРЯЗНЯЮЩИХ ВЕЩЕСТВ И ИЗМЕНЕНИЯ В ФИТОИНДИКАТОРАХ* СОДЕРЖАНИЕ ЗАГРЯЗНЯЮЩИХ ВЕЩЕСТВ И ИЗМЕНЕНИЯ В ЗООИНДИКАТОРАХ* СВАЛКИ И ЗАХОРОНЕНИЯ ЗАГРЯЗНЕНИЕ ВОДОЕМОВ СОДЕРЖАНИЕ ЗАГРЯЗНЯЮЩИХ ВЕЩЕСТВ В ВОДОПРИЕМНИКАХ* СОДЕРЖАНИЕ ЗАГРЯЗНЯЮЩИХ ВЕЩЕСТВ В ДОННЫХ ОТЛОЖЕНИЯХ* СОДЕРЖАНИЕ ЗАГРЯЗНЯЮЩИХ ВЕЩЕСТВ В ГИДРОБИОНТАХ* ПРЕДЕЛЬНО ДОПУСТИМЫЕ КОНЦЕНТРАЦИИ ЗАГРЯЗНЕНИЙ В ВОДОЕМАХ* РАСТЕНИЯ СИСТЕМАТИКА ВЫСШИХ РАСТЕНИЙ* НАЗВАНИЯ РАСТЕНИЙ НАЗВАНИЯ РАСТЕНИЙ (ЛАТ.)* НАЗВАНИЯ РАСТЕНИЙ (РУС.)* ФИТОИНДИКАТОРЫ ФИТОИНДИКАТОРЫ ХИМ. СОЕДИНЕНИЙ* ТЕСТОВЫЕ ХАРАКТЕРИСТИКИ ФИТОИНДИКАТОРОВ* ФИТОИНДИКАТОРЫ ВОДНЫЕ ХИМИЧЕСКИЙ СОСТАВ ВОДНЫХ РАСТЕНИЙ БИОХИМИЧЕСКИЙ СОСТАВ ВОДНЫХ РАСТЕНИЙ ВИДОВОЙ СОСТАВ ФИТОПЛАНКТОНА ЗАГРЯЗНЕННЫХ ВОДОЕМОВ ФИТОИНДИКАТОРЫ НАЗЕМНЫЕ СОДЕРЖАНИЕ ОСНОВНЫХ ХИМ.ЭЛ.-ТОВ В РАСТЕНИЯХ* СОДЕРЖАНИЕ ТЯЖЕЛЫХ МЕТАЛЛОВ В РАСТЕНИЯХ* ХИМИЧЕСКИЙ СОСТАВ КОРМОВЫХ РАСТЕНИЙ СОСТАВ В АБСОЛЮТНО СУХОМ ВЕЩЕСТВЕ* СОСТАВ ПРИ ЕСТЕСТВЕННОЙ ВЛАЖНОСТИ* ЗАГРЯЗНЕНИЕ ТЕРРИТОРИЙ МОРФОЛОГИЧЕСКИЕ ИЗМЕНЕНИЯ ЛЕСНЫХ РАСТЕНИЙ НАЗВАНИЯ НАЗЕМНЫХ РАСТЕНИЙ* НАЗВАНИЯ ВОДНЫХ РАСТЕНИЙ* РАСТИТЕЛЬНОСТЬ РАСТИТЕЛЬНОСТЬ НАЗЕМНАЯ БИОЦЕНОЗЫ НАЗЕМНЫЕ БИОЦЕНОЗЫ ЛХ РАСТИТЕЛЬНОСТЬ ЛХ ДРЕВОСТОЙ* РАСТИТЕЛЬНОСТЬ ЛЕСНАЯ ПРОЧАЯ* РАСТИТЕЛЬНОСТЬ ВОЛОТНАЯ* МОРФОЛОГИЧЕСКИЕ ИЗМЕНЕНИЯ ЛЕСНЫХ РАСТЕНИЙ РЕДКИЕ И ОХРАНЯЕМЫЕ ВИДЫ РАСТЕНИЙ* ФИТОИНДИКАТОРЫ НАЗЕМНЫЕ ПОПУЛЯЦИИ ЖИВОТНЫХ ЛХ МЛЕКОПИТАЮЩИЕ* ПТИЦЫ* АМФИБИИ И РЕПТИЛИИ* НАСЕКОМЫЕ* ПАУКООБРАЗНЫЕ* ПОЧВЕННЫЕ БЕСПОЗВОНОЧНЫЕ* МИКРООРГАНИЗМЫ ОТНОШЕНИЯ В БИОЦЕНОЗАХ ЛХ МЕЖВИДОВЫЕ ОТНОШЕНИЯ ЛХ ПАРАЗИТИЗМ НАЗЕМНЫЙ ВРЕДИТЕЛИ И БОЛЕЗНИ ЛЕСА* КРОВОСОСУЩИЕ КЛЕЩИ* КОНКУРЕНЦИЯ НАЗЕМНАЯ МУТВАЛИЗМ НАЗЕМНЫЙ</p>
---	--

Табл.2.