

ПРИМЕНЕНИЕ ПОДХОДА “SEMANTIC WEB” ПРИ ПРОЕКТИРОВАНИИ ПОРТАЛА “АТМОСФЕРНЫЕ АЭРОЗОЛИ СИБИРИ”*

Ю. И. Молородов

Институт вычислительных технологий СО РАН, Новосибирск, Россия

e-mail: yumo@ict.nsc.ru

Н. В. Голиков, И. А. Сударикова

Новосибирский государственный университет, Россия

e-mail: tensor,sudarikova@ngs.ru

In this paper the “Semantic Web” approach in designing of the portal “Atmospheric aerosols of Siberia” is described. Prospects of the implementation of the ontology of problems for the Atlas and running of the Web-services are considered within this approach.

Введение

Сеть Интернет является средой, позволяющей реализовать основные условия для сотрудничества ученых разных стран и разделов наук. Стало возможным создание распределенных массивов данных и программного обеспечения, которое в последнее время приобретает новую форму — Web-сервис. Появление в Интернете многочисленных информационных порталов по естественным наукам способствует интеграции разных разделов науки, но до сих пор не произошла массовая смена парадигмы по отношению к типу применяемого интерфейса. Подавляющее большинство исследователей продолжают работать в интерфейсе компьютер — человек. Основной недостаток этого интерфейса связан с нерешенностью корневой проблемы коммуникации человека с компьютером — однозначного соотнесения семантики естественного языка и программных средств. Развитие концепции “Semantic Web” неизбежно повлияло на технологию разработки Web-ориентированных информационных систем (WIS) [1]. Под WIS понимаются информационные системы, которые для взаимодействия с пользователем преимущественно используют клиент-серверную архитектуру на основе HTTP-протокола. Смысл концепции “Semantic Web” — сделать так, чтобы документы в Интернете имели структуру аннотаций и связей, подходящую для автоматической программной обработки. Проблемы, породившие данную концепцию, довольно очевидны и представляют собой наиболее актуальные вопросы современного Интернета [2]. Сюда относятся глобальная задача поиска (связанная с ростом общего объема

*Работа выполнена при частичной финансовой поддержке Президентской программы (грант № НШ 2314.2003.19) и Интеграционной программы СО РАН (грант № 169).

© Институт вычислительных технологий Сибирского отделения Российской академии наук, 2005.

информации в Web), сопряжение гетерогенных систем, работающих в сети, разработка интеллектуальных Web-сервисов, Web-агентов. Поэтому разработчиками предпринимались шаги для усовершенствования Web-стандартов описания данных, в результате появились языки описания метаданных (RDFS, DAML+OIL [3], OWL [4, 5]), предназначенные для записи онтологий, т. е. описаний множеств объектов и отношений между ними. В отличие, например, от известного формализма описания знаний KIF (Knowledge Interchange Format), который имеет синтаксис, похожий на LISP, языки DAML+OIL и OWL обладают более слабыми описательными возможностями, однако выигрывают в вычислительных характеристиках.

1. Онтологии

Онтология имеет центральное значение в любой информационной системе, поскольку каждая информационная система имеет под собой систему объектов из реального мира, которую она описывает. Если модель системы (а точнее, ее онтология) имеет формальное представление в виде, удобном для автоматической обработки, то ее можно использовать как программный компонент информационной системы и применять на стадиях разработки и функционирования информационной системы. Формальные онтологии применимы для:

- взаимодействия программных агентов, которые могут “понимать” друг друга, обмениваясь сообщениями в терминах общедоступной онтологии;
- сопряжения гетерогенных источников данных, обеспечивая запросы ко всем ним на одном языке в терминах общей онтологии;
- поиска информации, улучшая точность ответа с помощью “расшифровки” понятий, используемых в запросе, в соответствии с онтологиями ресурсов;
- построения программных интерфейсов с целью их динамического генерирования.

Онтологию также можно применять в качестве компонента баз знаний, схемы объектов в объектно-ориентированных системах, концептуальной схемы базы данных, структурированного глоссария взаимодействующих сообществ, словаря для связи между агентами, определения классов для программных систем. Онтологии позволяют соответствующим программным средствам (интеллектуальным агентам) автоматически (без участия человека) определять смысл терминов, использованных при описании ресурсов, и сопоставлять его со смыслом поставленной задачи. В этом случае агенты создаются таким образом, чтобы они могли использовать онтологии, а онтологии проектируются так, чтобы можно было разделять знания с этими агентами и среди них [6]. Агент применяет онтологию, если его наблюдаемые действия совместимы с определениями в этой онтологии. При этом использование онтологии является гарантией логичности, но не полноты по отношению к запросам и утверждениям, использующим словарь, определенный в этой онтологии.

На данный момент получена онтология задачи для атласа “Атмосферные аэрозоли Сибири”. В системе можно выделить два уровня метаданных:

- специфические (динамические) метаданные — данные о месте наблюдения, наблюдаемых элементах, проведенных измерениях и т. д. Эти метаданные могут быть получены из базы данных путем сокрытия полей с конфиденциальными данными;
- общие метаданные (статические) — информация о том, где, когда данные собраны, как часто пополняются, в каком формате сохранены и т. д. Эти метаданные являются дополнительными, поскольку в самой базе данных, как правило, не хранятся.

Поддержка специфических метаданных в актуальном состоянии — непростая задача. Большинство известных баз метаданных накапливают информацию централизованно и пополняются крайне нерегулярно. Целесообразно рассредоточить формирование баз метаданных и поручить поддержание их в актуальном состоянии специальным автоматическим агентам или администраторам. Осуществлять этот процесс следует с использованием единой методологии, которая должна обеспечивать следующие функции:

- отражение баз данных в базы метаданных;
- установление эквивалентностей между понятиями и объектами различных баз метаданных;
- обеспечение поиска информации в распределенной базе;
- контроль пополнения баз метаданных.

На этапе построения использовался редактор онтологий Protege (версия 3.1 с использованием OWL Plugin 2.1) [<http://protege.stanford.edu/>]. Сформированный документ онтологии проверен как Full ontology [<http://phoebus.cs.man.ac.uk:9999/OWL/Validator>], (рис. 1).

Природа физических величин задана исследуемой предметной областью, хотя расчетные параметры могут быть не связаны с ней. Ведется работа над реализацией динамического интерфейса поиска информации в системе на основе свойств и понятий, описанных в онтологии.

2. Web-сервисы

“Web-сервис” — это реализуемая программными средствами система для поддержки межмашинного взаимодействия через сеть. Интерфейс сервиса описывается на языке, читаемом машиной, например WSDL. Другие системы взаимодействуют с Web-сервисом способом, указанным в его описании, используя сообщения в стандарте SOAP, передаваемые с использованием HTTP и XML и в сочетании с другими стандартами, относящимися к Web. Физически “Web-сервис” представляет собой фрагмент программного обеспечения, способный передавать и принимать сообщения. Цель Web-приложений для данной системы — увеличение презентационных и вычислительных возможностей портала “Атмосферные аэрозоли Сибири”. Построение системы на базе программного продукта SMART [7] [<http://web.ict.nsc.ru/smart/>] допускает расширение функциональности атласа добавлением Web-сервисов и Java-библиотек.

Сервис “Альbedo” предоставляет возможность интеграции существующей базы данных альbedo, представленной в виде набора файлов, с атласом “Атмосферные аэрозоли Сибири”; предназначен для получения значений коэффициента отражения от земной поверхности световых лучей — альbedo для заданной точки и промежутка времени. Каждый файл альbedo содержит массив значений альbedo для точек земной поверхности, принадлежащих одной квадратной области определенного размера [8]. Получаемые данные впоследствии могут быть визуализированы GIS-программой jMapper [9]. Работа ведется совместно с Институтом космических исследований РАН (г. Москва).

Сервис “Аэрозольный калькулятор” предназначен для определения физико-химических характеристик атмосферного аэрозоля: размерных, температурных параметров, давления и т. д. Получаемые характеристики планируется использовать в дальнейших расчетах других сервисов. Данное направление в работе ведется совместно с Институтом физики атмосферы СО РАН (г. Томск).

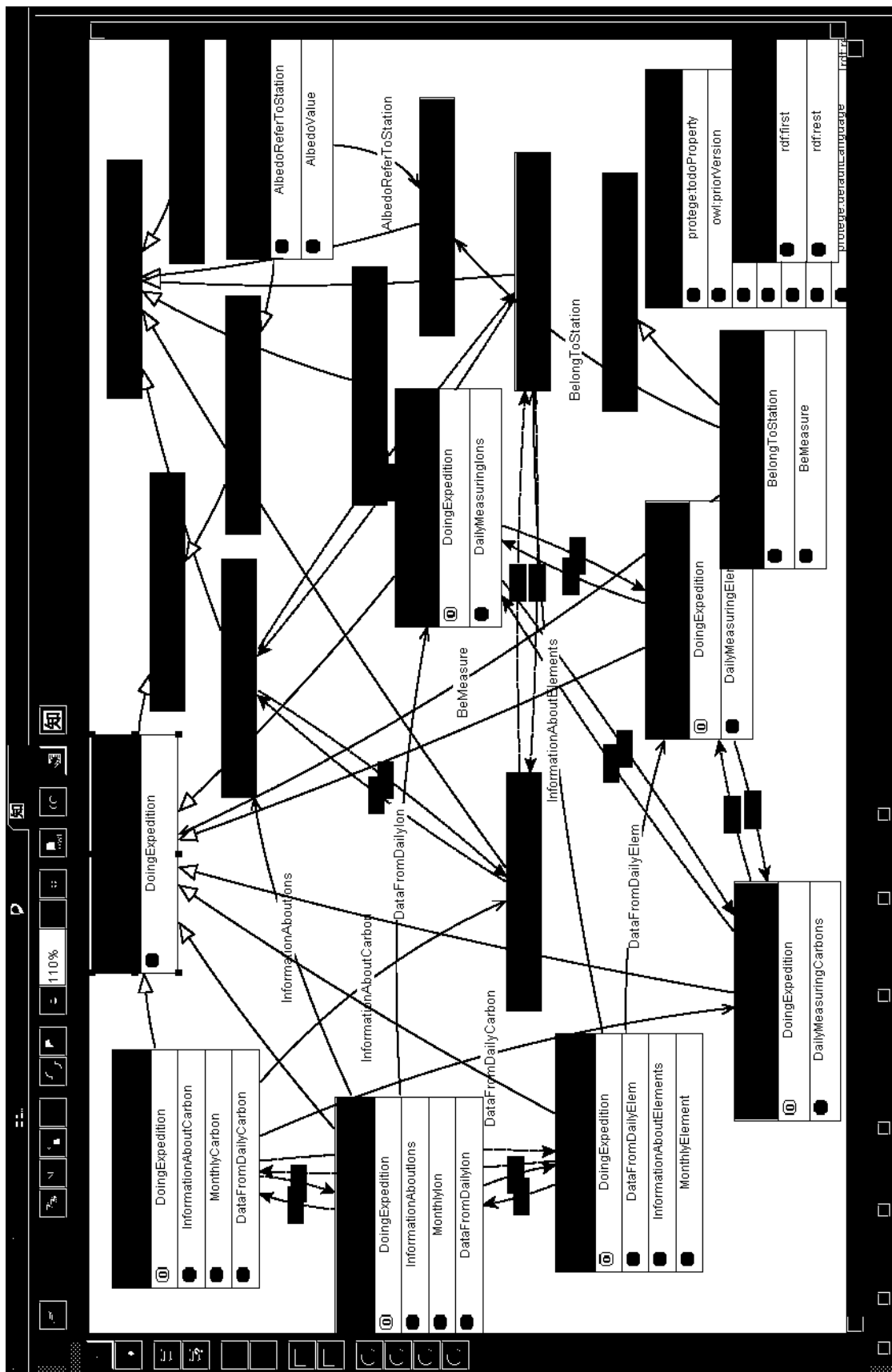


Рис. 1. Графическое представление связей в онтологии атласа “Атмосферные аэрозоли Сибири”.

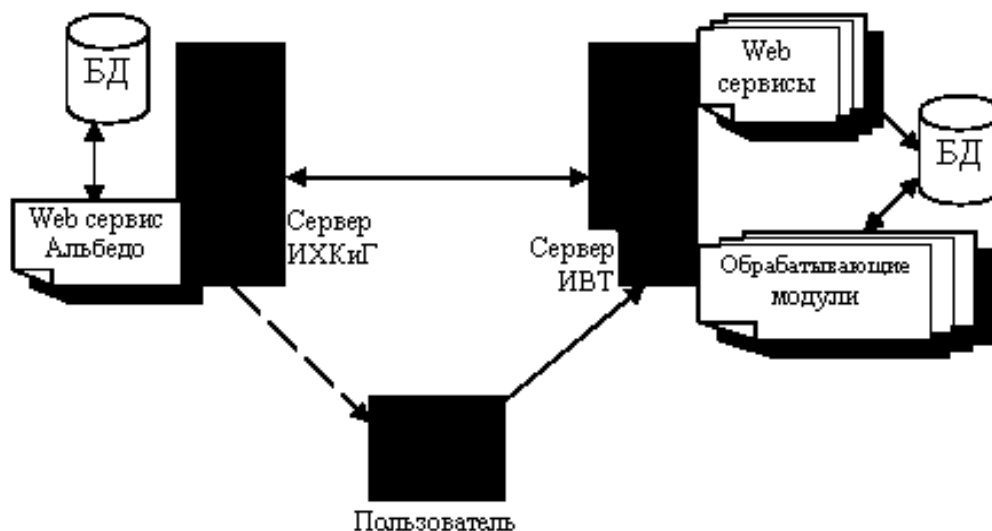


Рис. 2. Распределение данных и Web-сервисов на серверах СО РАН.

Ответ пользователю может приходиться напрямую с сервера Института химической кинетики и горения, где расположена база данных космической съемки. Если работать с географической ГИС-картой системы, то результат запроса клиента поступит на сервер Института вычислительных технологий СО РАН, а затем, соответственно, пользователю системы (рис. 2).

Заключение

Переходя к практическому применению онтологии, необходимо отметить, что онтологические аксиомы чаще всего используются неявно. Решая те или иные задачи, человек обычно не задумывается об "онтологическом базисе", предпочитая опираться на научные теории, обладающие относительной полнотой и непротиворечивостью, а главное, возможностью получать дедуктивные выводы с помощью того или иного исчисления. Однако при изучении сложных систем-организаций на этом пути чаще всего его ожидают разочарования по причине невозможности построения целого на основе его частей. Было принято решение вернуться к онтологическим представлениям и попытаться на этом "более высоком" уровне решить проблему целостного, а не фрагментарного описания системы. Информационно-вычислительная система переживает период становления, и многие проблемы еще не решены и даже не сформулированы (например, как можно сравнивать или интегрировать онтологии), но к настоящему времени в литературе уже есть некоторая информация по этим проблемам [10]. Атлас уже содержит значительные цифровые ресурсы. Это информация о научных публикациях, базы данных, алгоритмы и программы. Онтологии могут занять важное место среди средств интеграции, обеспечив семантическую интеграцию этих ресурсов. Применение подхода Semantic Web к созданию портала предоставляет возможность не только объединять различные информационные ресурсы в одну концептуально информационную среду, но и оперативно управлять, актуализировать информацию, хранящуюся в разнородных базах данных и источниках, а также обрабатывать данные с целью их дальнейшего использования для климатического моделирования. Система расположена по адресу [<http://web.ict.nsc.ru/aerosol/>].

Список литературы

- [1] ROUSSEAU B., ROUSSEAU F. Some ideas concerning the Semantic Web // Library and Information Service. 2002. Vol. 8. P. 39–49.
- [2] BERNS-LEE T. Web Architecture from 50,000 feet. [<http://www.tamark.ca/students/?p=1582>]
- [3] FENSEL D., HARMELEN F., HORROCKS I. ET AL. OIL: An ontology infrastructure for the semantic Web // IEEE Intelligent System. 17:2. March/April 2001. Vol. 16, N 2.
- [4] PATEL-SCHNEIDER P.F., HAYES P., HORROCKS I. OWL Web Ontology Language Semantics and Abstract Syntax. W3C Working Draft. [<http://www.w3.org/TR/WCAG20/>]
- [5] MCGUINNESS D.L., HARMELEN F. OWL Web Ontology Language Overview. W3C Working Draft. [<http://www.w3c.org/TR/owl-features/>]
- [6] BENJAMINS V., CONTRERAS J., CORCHO O., GOMEZ-PEREZ A. Six Challenges for the Semantic Web. [http://www.iswc2002.semanticweb.org/posters/VR_Benjamins.pdf]
- [7] ШПРАЙБМАН В.Б., ГУСЬКОВ А.Е. Разработка информационных систем на основе RDF-технологии // Тр. ХLI Междунар. науч. студенческой конф. “Студент и научно-технический прогресс”. Новосибирск: НГУ, 2003. Ч. 1. 153 с.
- [8] ДУБРОВ И.С. Сервис Альbedo. Технический проект. Версия 1.0. [<http://www.ng.ru/science/2005-04-27/Albedo.html>]
- [9] GIS приложение jMapper: [<http://www.uiggm.nsc.ru/uiggm/geology/>]
- [10] VAN HEIJST G., SCHREIBER A.T., WIELINGA B.J. Using explicit ontologies in KBS development // Intern. J. of Human and Computer Studies. 1996. Vol. 46, N (2–3). P. 183–292.

Поступила в редакцию 2 июня 2005 г.