

МОДЕЛЬ СЕМАНТИЧЕСКОЙ ИНТЕГРАЦИИ ИНФОРМАЦИОННЫХ СЕРВИСОВ

П. В. Титенков

Кафедра интеллектуальных информационных технологий

Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

E-mail: pavel.titenkov@gmail.com

В работе рассматривается методология интеграции информационных сервисов.

ВВЕДЕНИЕ

В современном мире благодаря быстрому развитию информационных технологий накоплено огромное количество информации, представленной в электронном виде. Такая форма представления информации предоставляет широкие возможности для ее систематизации и структуризации, однако современные технологии и подходы хранения и представления информации не позволяют этого сделать.

В большинстве случаев разработчик сервиса концентрируется на решении конкретной задачи пользователя и не уделяет должного внимания подготовке онтологии используемых понятий, формализации связей между ними, анализу предметной области и, в последствие, создает сервис, который никаким образом не общается с другими сервисами пользователя и сам для этого общения никак не подготовлен.

Такой подход разрозненного существования сервисов приводит к накоплению огромного количества информационного мусора и изначально ограничивает возможности персонификации сервиса под нужды пользователя. В данном случае интеграция сервисов пользователя и создание его единого личного информационного пространства открывает перед ним абсолютно новые возможности и кардинально меняет представление о том, каким должен быть интерфейс взаимодействия пользователя с информационными сервисами.

I. ПРОБЛЕМА ИНТЕГРАЦИИ

Под интеграцией информационных сервисов понимается обеспечение единого унифицированного интерфейса для доступа к некоторой совокупности неоднородных и независимых источников данных. Объединение источников данных является очень трудоемкой задачей во многом из-за многоуровневости их гетерогенности.

Можно выделить следующие уровни гетерогенности информационных сервисов:

- архитектурный;
- системный;
- структурный;
- синтаксический;
- семантический.

Основной причиной этой многоуровневости является использование различных форматов и моделей представления данных и, как следствие, различных программных и аппаратных реализаций. Задача интеграции усложняется также и тем, что совокупность источников данных не является статичной во времени: источники автономны (то есть разработаны и эксплуатируются независимо друг от друга, независимо спроектированы под решение конкретных задач, эволюционируют и умирают с течением времени) и распределены (физически находятся на разных серверах, адреса которых с течением времени могут меняться).

В общем случае можно выделить два подхода для интеграции сервисов — материализованный и виртуальный. В случае материализованной интеграции данных создается новый источник интегрированных данных, который используется автономно от породивших его источников. При необходимости его состояние приходится синхронизировать с их актуальным состоянием. В случае виртуальной интеграции, система использует виртуальный источник, который в любой момент времени имеет доступ к актуальным данным источников, и синхронизации его состояния не требуется. Данные для ответа на запросы пользователя, в этом случае берутся из информационных источников непосредственно в процессе выполнения запроса. Права владельцев исходных интегрируемых источников сохраняются. Они продолжают автономно поддерживать их в своих интересах, предоставляя вместе с тем права доступа к их ресурсам пользователям системы интеграции данных в соответствии с установленным регламентом. Авторизованные пользователи системы интеграции получают непосредственный доступ только к виртуальному источнику, воплощаемому данной системой.

Исходя из совокупности проблем и подходов решения интеграции информационных сервисов можно сформулировать следующие задачи:

- разработка архитектуры системы;
- создание интегрирующей модели данных, которая в последствие станет фундаментом интерфейса пользователя;

- разработка моделей и методов отображения данных гетерогенных сервисов в интегрируемую систему;
- разработка механизмов семантической интеграции источников данных.

Для обеспечения смысловой интероперабельности данных первостепенной становится задача интеграции ресурсов на семантическом уровне.

II. СЕМАНТИЧЕСКАЯ ИНТЕГРАЦИЯ

Семантическая интеграция основывается на знании и учете природы данных информационного источника. Решение задачи семантической интеграции данных невозможно без построения единой базы знаний охваченных предметных областей и разработки единой онтологии понятий (под онтологией понимается формальное описание некоторой области знаний, которое включает в себя релевантные классы объектов, их иерархию, связи и аксиомы, принятые в этой области).

В любом информационном сервисе большинство объектов находятся в непрерывном взаимодействии между собой. Информация о характеристиках этих объектов, иерархия их наследования, как правило, хранится в отдельных базах данных в отличие от правил их поведения и связей, которые не хранятся в явном виде в хранилищах данных, а скрыты в реализациях различных алгоритмов того или иного сервиса. Формализовав всю имеющуюся информацию об информационном ресурсе в единую семантическую модель предметной области и представив ее в виде семантической сети, можно построить интеллектуальную информационную систему, которая сможет не только собирать и выдывать пользователю информацию из различных источников, но также сможет производить операции логического вывода и принимать решения за пользователя, предоставляя ему интерфейс взаимодействия кардинально другого уровня. Не стоит также забывать и об знаниях, которые могут находиться "на стыке" предметных областей определенных информационных сервисов и оставаться скрытыми для пользователя в силу отсутствия общения между ними.

Ключевой задачей семантической интеграции информационных сервисов является задача интеграции их онтологий понятий. В общем, интеграцию онтологий принято определять как процесс нахождения сходства двух онтологий А и В и, как результат, создание новой онтологии С, объединяющей и согласующей семантические представления исходных онтологий. В результате, две системы, основанные на онтологиях А и В, получают возможность взаимодействовать между собой, используя онтологию С.

Различают различные уровни интеграции онтологий понятий:

- соответствие
- совместимость
- унификация

В случае построения интеллектуальной системы с общей для всех интегрируемых сервисов базой знаний наиболее предпочтительным уровнем интеграции онтологий понятий является унификация, которая обеспечит гибкость и простоту расширяемости системы. В случае наличия унифицированной онтологии понятий агенту, осуществляющему сбор информации из определенного сервиса, будет необходимо лишь приводить операцию трансформации ее к единому формату согласно заранее заготовленному шаблону вместо поиска соответствий или совместимых понятий.

III. ЗАКЛЮЧЕНИЕ

Объединение знаний поступающих из привычных для пользователя источников в единую интеллектуальную систему открывает перед ним новые возможности для персонализации информационных сервисов, систематизации и структуризации личного информационного пространства, автоматизации привычных рутинных процессов.

1. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский – СПб: Питер, 2000.
2. Представление и обработка знаний в графодинамических ассоциативных машинах /В.В. Голенков [и др.]; под ред. В.В. Голенкова. – Минск : БГУИР, 2001.
3. Интеграция знаний в информационных системах. / Н.А. Гулякина, В.П. Ивашенко // Доклады БГУИР. – 2004. – №6. – С. 113-119.
4. Ломов П. А., Шишаев М. Г. Интеграция семантически связанных информационных ресурсов на основе онтологий для эффективного информационного обеспечения рационального природопользования / П. А. Ломов, М. Г. Шишаев // Глубокая переработка минеральных ресурсов: Сборник материалов IV школы молодых ученых и специалистов «Сбалансированное природопользование» (6-8 ноября 2007 г.) – Апатиты: Изд-во КНЦ РАН. 2008. – С.243-247
5. Смирнов А.В., Пашкин М.П., Шилов Н.Г., Левашова Т.В. Онтологии в системах искусственного интеллекта: способы построения и организации// Новости искусственного интеллекта.
6. Wache H., and other. Ontology-Based Integration of Information. A Survey of Existing Approaches. Proceedings of the IJCAI-01 Workshop: Ontologies and Information Sharing, 2001. Режим доступа: <http://www.cs.vu.nl/heiner/public/ois-2001.pdf>
7. Bianchini D., De V. Antonellis «Ontology-based Integration for Sharing Knowledge over the Web». Режим доступа: <http://www.doc.ic.ac.uk/pjm/diweb2004/DIWeb2004 Part8.pdf>
8. Menzel, Christopher, "Ontology Theory," in J. Euzenat, A. Gomez-Perez, N. Guarino, and H. Stuckenschmidt (eds.), Ontologies and Semantic Interoperability, CEUR Workshop Proceedings, 64 (2002). Режим доступа: <http://CEUR-WS.org/Vol-64/menzel.pdf>.