

Анализ и сравнение систем интеграции неоднородных информационных ресурсов*

© Вовченко А.Е., Калиниченко Л.А.
Институт проблем информатики РАН
itsnein@gmail.com, leonidk@synth.ipi.ac.ru

Аннотация

Существует два подхода к решению задач в информационных системах над множеством интегрированных информационных ресурсов: движимый ресурсами и ориентированный на проблему. В статье предпочтение отдается второму. Одной из проблем при таком подходе остается та, что в различных инфраструктурах (веб-сервисы, грид-архитектуры, семантический веб, и др.) до сих пор открытым является вопрос интегрированного представления множественных источников информации для исследователя, решающего задачу. В статье рассматриваются основные принципы интеграции неоднородных информационных ресурсов, на основании чего выбирается набор критериев, по которым сравнивается ряд известных прототипов систем интеграции.

1 Введение

В различных областях науки наблюдается экспоненциальный рост объема получаемых экспериментальных (наблюдательных) данных. Например, в астрономии текущий и ожидаемый темп роста данных от наземных и космических телескопов удваивается в течение периода от шести месяцев до одного года. Сложность использования таких данных увеличивается еще и вследствие их естественной разнородности. Число организаций, получающих данные наблюдений в отдельных областях науки в мире, велико. Разнообразие (информационная несогласованность) получаемой информации вызывается, в частности, не только большим числом организаций, производящих наблюдения, и их независимостью, но и разнообразием объектов наблюдения, непрерывным и быстрым совершенствованием техники наблюдений, вызывающим адекватные изменения структуры и содержания накапливаемой информации. Это приводит к необходимости использования неоднородной, распределенной информации, накопленной в течение значительного периода наблюдений технологически различными инструментами.

Чрезвычайно быстро развивается также программный инструментарий, включающий многообразные сервисы для поддержки различных видов обработки информации при решении научных задач и проведении исследований. Такие сервисы производятся различными научными организациями, их описания неоднородны и неполны. Увеличивающийся разрыв между исследователями и источниками данных и сервисов приводит к необходимости поиска новых путей организации решения задач над множественными распределенными коллекциями данных и программ, которые концентрируются в специализированных центрах данных и вычислительных ресурсов. Разработан (разрабатывается) ряд инфраструктур, которые технически позволяют реализовать решение задач над множественными информационными источниками. Среди них Веб-сервисы, Грид-архитектуры, Семантический Веб, технологии распределенных баз данных, интероперабельные технологии промежуточного слоя, и др. Они составляют техническую среду для организации решения задач.

Традиционно при решении задач специалисты используют привычные для них источники информации, и формулируют задачи, учитывая лишь такие источники. Подобные способы формулирования и решения задач называются далее *движимыми конкретными ресурсами* информации, отобранными до или, в лучшем случае, в процессе формулирования задачи. Очевидна неполнота информации, которую удается охватить при таком подходе. Множество источников данных и сервисов, существующих в Интернете, их разнообразие, вызывают потребность в радикальном изменении такого традиционного подхода. Существо этого изменения заключается в том, что задачи должны формулироваться независимо от существующих источников информации, и лишь после такой формулировки, должна осуществляться идентификация релевантных задаче источников, приведение их к виду, требуемому в задаче, их интеграция, идентификация сервисов, которые позволяют реализовать отдельные части абстрактного процесса решения задачи. Только после этого должно осуществляться конструирование конкретного процесса решения на основе отобранных источников данных и сервисов.

Этот подход, называемый *ориентированным на проблему*, должен значительно больше опираться на использование техники представления знаний, нежели традиционный. Так, формулирование задачи должно быть основано на определении ее проблемной области, включающем ее терминологию и систему понятий, абстрактное описание соответствующей материальной системы, определение адекватных моделей и теорий, абстрактное описание требуемых в задаче характеристик объектов реального мира, определение методов, алгоритмов и процессов решения задачи.

Одной из проблем при таком подходе остается та, что во всех названных выше инфраструктурах до сих пор открытым является вопрос интегрированного представления множественных источников информации для исследователя, решающего задачу.

На сегодняшний день существует множество средств интеграции неоднородных информационных ресурсов, созданных для тех или иных нужд, достаточно лишь взглянуть на их далеко неполный список [1]. Предварительная классификация подходов к интеграции представлена в статье [2]. В данной статье рассматриваются подходы к интеграции ресурсов, основанные на взглядах [3].

Различаются два принципиально различных подхода к проблеме интеграции: 1) двигаясь от ресурсов к задачам (глобальная схема образуется как интегрированная схема множества ресурсов независимо от приложения) и 2) двигаясь от приложения к ресурсам (описание предметной области приложения образуется независимо от ресурсов (в терминах понятий, структур данных, функций, процессов), а затем релевантные приложению ресурсы отображаются в это описание).

Первый подход, движимый информационными ресурсами, является не масштабируемым по отношению к числу ресурсов, не дает возможности достижения семантической интеграции ресурсов в контексте конкретного приложения, не ведет к доказательной идентификации релевантных приложению ресурсов, не способствует повышению стабильности спецификации посредника в процессе эволюции ресурсов, релевантных приложению. Эти недостатки являются характерными для подхода, при котором глобальная схема является взглядом (Global as View (GAV)) [3, 4]. GAV может служить в качестве базовой техники подхода, движимого информационными ресурсами.

Другой подход (движимый приложениями) предполагает создание предметного посредника, который поддерживает взаимодействие между приложением и ресурсами на основе определения прикладной области (определения посредника). Второй подход имеет очевидные преимущества по отношению к подходу, движимому информационными ресурсами. Процесс

регистрации неоднородных информационных ресурсов в предметном посреднике в подходе, движимом приложениями, основан на технике GLAV, комбинирующей два подхода – LAV (Local As View) – когда локальная схема ресурса является взглядом над схемой посредника, и GAV. Согласно LAV [4], схемы регистрируемых ресурсов рассматриваются как материализованные взгляды над виртуальными классами посредника. В этом случае GAV взгляды служат для разрешения различных конфликтов между спецификациями ресурсов и посредника и обеспечивают правила трансформации результатов запроса от ресурса в представление в посреднике. Подобная техника регистрации обеспечивает стабильность спецификации приложения при изменении конкретных информационных ресурсов и их фактического присутствия (удаление ресурса, добавление новых ресурсов, и пр.), а также масштабируемость посредников по отношению к числу ресурсов, регистрируемых в них. В GLAV [5], голова определения взгляда LAV может содержать произвольный запрос над схемой ресурса для устранения конфликтов на основе техники GAV. Кроме перечисленных, существует подход именуемый BAV (Both As View) [6], суть которого заключается в построении реверсивных взглядов. В статье [6] продемонстрировано как из BAV взглядов можно получить взгляды LAV и GAV.

Целью настоящей работы является анализ соответствия ряда известных исследовательских прототипов средств интеграции неоднородных ресурсов некоторому набору критериев полноты средств интеграции. В следующем разделе дано описание выбранного набора критериев для сравнения. Затем представлен ряд известных исследовательских прототипов средств интеграции и осуществлено формирование сводной таблицы анализа этих прототипов.

2 Выбор критериев для сравнения средств интеграции

Выбор критериев является важным при реализации какого-либо сравнения. Набор критериев, с одной стороны, должен быть полным, отражающим по возможности все основные аспекты средства интеграции, а с другой стороны, он должен быть операбельным без необходимости включения в него всего множества различных параметров средств интеграции. Весьма ограниченный набор критериев сравнения предложен в [7], где для оценки подходов интеграции предлагается критерий ASME (Abstraction, Selection, Modeling, Explicit Semantics). Abstraction заключается в абстрагировании пользователя, использующего посредник, от ресурсов, иначе говоря все ресурсы и их детали при интеграции скрываются от пользователя. Selection означает возможность выбора пользователем тех ресурсов, которые необходимы для решения его задачи. Modeling означает возможность создания

пользователем своего собственного представления предметной области для решения конкретной задачи. Наконец, Explicit Semantics означает поддержку возможности связывания схемы с семантическим описанием. Кроме того, данный критерий означает возможность задания запросов к метаданным.

Несмотря на ограниченность критерия ASME, в настоящей работе представляется возможным в качестве базы использовать эти четыре критерия, существенно дополнив их. Критерии “А” и “М” являются характеристикой интеграционного подхода (LAV, GAV, ...), поэтому они заменены указанием используемого подхода. Критерии “S” и “E” остаются без изменения.

Для обоснования критериев далее будут рассмотрены основные принципы построения средства интеграции.

2.1 Интеграционный подход (LAV, GAV и другие)

Как было отмечено во введении, существуют два подхода: движимого ресурсами (GAV) и движимого приложением (LAV). LAV представляет интерес в контексте решения задач над неоднородными информационными ресурсами. Используя подход, ориентированный на проблему, специалист формулирует задачу, описывая базовые сущности и понятия предметной области, в рамках которой предполагается решать задачу. Среди прочего, специфицируются структуры данных, специфицируются понятия предметной области, выражаемые соответствующими онтологическими спецификациями, описываются функции, процессы и пр. Спецификация предметной области, представляет собой спецификацию предметного посредника, для решения класса задач. Сущности и понятия предметной области, определенные таким образом, не зависят от существующих информационных ресурсов. Также в пользу данного подхода могут служить доводы, высказанные в [7], при описании критериев Abstraction, Modeling.

Поэтому в качестве критерия был выделен критерий используемого подхода (Approach, “А”).

2.2 Создание расширяемой канонической информационной модели

При интеграции неоднородных ресурсов в посреднике нужно уметь семантически отождествлять объекты, представленные в различных информационных моделях, и семантически правильно отображать схемы интегрируемых ресурсов в схему посредника. Поскольку в общем случае ресурсы неоднородны (представлены в различных моделях), при интеграции неоднородных ресурсов для однородного представления их семантики требуется приведение различных информационных моделей к унифицированному виду в рамках некоторой унифицирующей информационной модели, которая называется *канонической*.

Для унификации разнородных спецификаций прежде всего требуется умение сопоставлять

спецификации различных ресурсов друг с другом так, чтобы можно было отвечать на вопрос, можно ли при реализации посредника использовать спецификацию существующего ресурса вместо фрагмента спецификации посредника. Для этого достаточно доказать, что рассматриваемые спецификации находятся в отношении уточнения. Говорят, что спецификация А уточняет спецификацию D, если А можно использовать вместо D так, что пользователь D не будет замечать этой замены. Средства доказательства факта уточнения спецификации некоторого компонента спецификацией другого компонента (реализуемые на основе теоретико-модельных нотаций и соответствующего инструментария [10–12]) составляют фундамент применяемых методов конструирования унифицирующих (канонических) моделей представления информации в посредниках. Каноническая информационная модель служит в качестве общего языка, эсперанто, для адекватного выражения семантики разнородных моделей представления информации, используемых в разнообразных информационных ресурсах, релевантных посреднику. Методы отображения информационных моделей и синтеза расширяемых канонических информационных моделей для слоя предметных посредников подробно рассмотрены в [13]. Там же предложена архитектура Унификатора информационных моделей.

Задачей Унификатора моделей является унификация множества информационных моделей (называемых *исходными*), использующихся в различных информационных ресурсах некоторой информационной системы. Унификация исходной модели R есть приведение ее к канонической информационной модели C, т. е. создание такого расширения E ядра канонической модели (которое может быть и пустым) и такого отображения M исходной модели в расширенную каноническую, что исходная модель *уточняет* расширенную каноническую модель. Уточнение моделей означает, что для любой допустимой спецификации r модели R ее образ M(r) при отображении M уточняется спецификацией r. Синтез канонической модели заключается в построении указанных расширений ядра канонической модели для всех исходных моделей и объединении таких расширений. В качестве ядра канонической информационной модели применяется язык СИНТЕЗ [14], который ориентирован на спецификацию предметных посредников и синтез канонических моделей.

Во многих проектах подобные вопросы остаются без должного внимания, прототип разрабатывается только для нескольких моделей данных (чаще всего реляционной, XML). Поэтому возможность Унификации (Unification “U”) произвольных информационных моделей представляется важным критерием.

2.3 Семантический поиск соответствий спецификаций схемы посредника и схем информационных ресурсов

Регистрация релевантных посреднику ресурсов рассматривается как задача композиционного проектирования систем [8, 9]. Регистрация ресурсов есть процесс целенаправленной трансформации спецификаций, включающий декомпозицию спецификаций посредника на непротиворечивые фрагменты, поиск среди спецификаций релевантных ресурсов подходящих типов данных — кандидатов для уточнения ими спецификаций типов посредника, построение выражений, определяющих классы ресурсов в виде композиции классов посредника.

Принципиальным моментом в этой схеме является реализация доказательства уточнения фрагментов спецификаций посредника спецификациями ресурсов в процессе построения отображений таких спецификаций. Идентификация релевантных ресурсов (предшествующая регистрации) основана на использовании трех моделей — *модели метаданных*, характеризующих свойства ресурсов, собранных в некотором реестре, *онтологической модели*, позволяющей формально определять понятия предметной области посредника или ресурса, и *канонической модели*, позволяющей формально определять структуру и поведение объектов предметной области посредника и информационных ресурсов. Рассуждения в канонической и в онтологической модели основаны на семантике канонической модели и средствах доказательства уточнения. При этом в онтологической модели необходимо достичь согласования понятийной семантики спецификаций посредника и регистрируемых в нем ресурсов. Рассуждения в модели метаданных являются эвристическими на основе нефункциональных требований к нужным в классе задач ресурсам. Необходимые модели метаданных и алгоритмы поиска составляют часть метода. В целях проектирования спецификации посредника и ресурсов задаются в однородном их представлении в канонической модели, хотя для этого может потребоваться преобразование в такую модель из некоторого другого языка спецификаций, например из UML.

Выше описан подход, используемый в проекте SYNTHESIS, в других проектах семантика может использоваться иначе, но несомненна важность наличия возможности использования семантических описаний, а также возможности поиска по метаданным спецификаций релевантных ресурсов, или для других задач. В [7] такой критерий был предложен, и назван Explicit Semantics (“E”).

2.4 Ориентация средства интеграции на решение задач

Важным аспектом является ориентация средства интеграции на решения научных задач. В связи с

этим целесообразно введение двух дополнительных критерия. Как было сказано выше при описании предметной области, специалист может задавать спецификации понятий, онтологий, структур данных, а также спецификации функций (методов) и процессов. Последнее особенно важно в контексте решения задач, т.к. обработка данных с помощью различных средств является неотъемлемой частью процесса решения задачи. Таким образом, возникает критерий достаточной функциональности (Functionality, “F”).

Вторым важным аспектом при решении задач, является расширяемость архитектуры. В настоящий период существуют множество инфраструктур решения задач: веб-сервисы, семантический-веб, виртуальные обсерватории в астрономии, грид-инфраструктуры. Последние представляют особую важность, т.к. различные грид-инфраструктуры обладают огромными ресурсами. Поэтому важным аспектом является возможность создания гибридных инфраструктур (Hybrid, “H”), иначе говоря, возможность встраивания средств интеграции в существующие инфраструктуры решения задач.

Последним аспектом при решении задач, является наличие удобных в использовании инструментов для работы со средством интеграции. В контексте решения задач, основной предполагаемой аудиторией пользователей средством интеграции являются специалисты в некоторой прикладной области. В астрономии – астрономы, в биологии – биологи, и т.д. Надеяться на то что, специалисты в своей области окажутся еще и хорошими ИТ специалистами не представляется возможным, поэтому наличие инструментов (Tools, “T”) с хорошим пользовательским интерфейсом является важным критерием.

Таким образом, для сравнения были выделены 7 критериев: Approach, Unification, Selection, Explicit Semantic, Functionality, Hybrid, Tools – AUSEFHT.

3 Выбор средств интеграции для сравнения

Выбор проектов для сравнения, учитывая их число и разнообразие, является непростой задачей. Прежде всего, в данной работе интересны те проекты, которые можно было бы достаточно содержательно сопоставлять с проектом SYNTHESIS [15-19] разрабатываемым в ИПИ РАН. С другой стороны, интерес представляют и отличные от проекта SYNTHESIS работы, отличающиеся оригинальными решениями. В работе были рассмотрены следующие проекты: LAV – Agora [20, 21], Infomaster [22, 23], SIRUP [24-26], PICSEL [27-29], Information Manifold [30]; GAV – MedMaker [31], MOMIS [32, 33], TSIMMIS [34, 35]; BAV – AutoMed [36-39]; GLAV – SYNTHESIS. Основное внимание уделено LAV сводимым проектам (LAV, BAV, GLAV), но также рассмотрены и некоторые известные GAV проекты.

4 Результаты сравнения

На таблице 1 видно, какой прототип соответствует описанным критериям. Помимо проекта SYNTHESIS, разрабатываемого в ИПИ РАН, выделены пять LAV – ориентированных проектов: AutoMed, Infomaster, Information Manifold, PICSEL, SIRUP. AutoMed основан на подходе, известном как both as view (BAV). BAV обеспечивает отображение между схемами в обоих направлениях, что немало важно для P2P приложений или веб-сервисов. В качестве языка запросов используется Automed Intermediate Query Language (AIQL). Благодаря отображению этих моделей в модель, основанную на гиперграфе, названную HDM [9], AutoMed позволяет осуществлять трансляцию схем между различными моделями. К настоящему времени описано отображение следующих моделей: реляционной, XML, ER, UML, ORM. Несоответствие критерию S не столько существенно, а вот отсутствие поддержки функций, является существенным недостатком. PICSEL – подход, ориентированный на интеграцию сервисов. Все информационные источники и сервисы обработки информации представляются в

виде сервисов. В качестве языка запросов используется расширенный язык дескриптивной логики CARIN (разработанный в проекте Information Manifold). В проекте реализованы возможности автоматической генерации описания сервисов по их спецификациям на языке WSDL. Главным недостатком проекта является отсутствие возможности расширения. Инструменты будут работать лишь с теми сервисами, для которых заранее были написаны адаптеры. Вопрос о возможности интеграции данного прототипа в различные инфраструктуры авторами рассмотрен не был.

Проект SIRUP, соответствует всем критериям предложенным в [7], но имеет два существенных недостатка: отсутствие поддержки функций и возможности расширения при желании интегрировать ресурсы с новыми информационными моделями. Основными же достоинствами проекта является ориентация на онтологии. Язык описания, названный IConcept, позволяет связывать каждое описываемое понятие с онтологическим понятиями. Кроме того разработан АПИ, позволяющий осуществлять запросы к онтологиям на языке AIQL (SQL-like).

Табл. 1. Результат соответствия рассмотренных прототипов выбранным критериям

Название проекта	A	U	S	E	F	H	T
Agora	LAV	No	No	No	No	No	Yes
AutoMed	BAV	Yes	No	Partially	No	Yes	Yes
Infomaster	LAV	No	No	No	No	No	No
SYNTHESIS	GLAV	Yes	Yes	Yes	Yes	Yes	Yes
PICSEL	LAV	No	Yes	Yes	Yes	No	Yes
SIRUP	LAV	No	Yes	Yes	No	Yes	Yes
Information Manifold	LAV	No	No	No	No	No	Yes
MedMaker	GAV	No	Yes	Partially	No	No	Yes

Заключение

В работе рассмотрены основные принципы для построения системы интеграции неоднородных ресурсов в контексте решения научных задач. На основе этих принципов были сформулированы 7 критериев для сравнения различных прототипов средств интеграции.

Хотя единственным проектом, удовлетворяющим всем выбранным критериям, является проект SYNTHESIS, другие аспекты, такие как, например, поддержка интеграции сервисов, генерация адаптеров, декларативная спецификация посредников и т.д., в других проектах проработаны лучше. Поэтому планируется дальнейшее расширение системы критериев.

Литература

[1] <http://www.ifi.unizh.ch/~pziegler/IntegrationProjects.html>.

- [2] Patrick Ziegler, Klaus R. Dittrich. Data Integration — Problems, Approaches, and Perspectives. In John Krogstie, Andreas L. Opdahl, and Sjaak Brinkkemper, editors, Conceptual Modelling in Information Systems Engineering, pages 39–58. Springer, Berlin, 2007.
- [3] Jeffrey D. Ullman. Information Integration Using Logical Views. In Proc. of the 6th Int. Conf. on Database Theory (ICDT'97), 1997.
- [4] Alon Y. Halevy. Answering Queries Using Views: A Survey. VLDB Journal, 10(4), 2001.
- [5] Marc Friedman, Alon Levy, and Todd Millstein. Navigational Plans for Data Integration. In Proceedings of the National Conference on Artificial Intelligence (AAAI), 1999.
- [6] P.J. McBrien, A. Poulouvasilis. Data Integration by Bi-Directional Schema Transformation Rules. In Proceedings of ICDE03, IEEE, Pages 227-238, 2003.

- [7] Patrick Ziegler. User-Specific Semantic Integration of Heterogeneous Data: What Remains to be Done? Technical Report ifi-2004.01, Department of Informatics, University of Zurich, 2004.
- [8] Briukhov D. O., Kalinichenko L. A. Component-based information systems development tool supporting the SYNTHESIS design method // 2nd East-European Conference “Advances in Databases and Information Systems” Proceedings.—Berlin-Heidelberg:Springer-Verlag, 1998. P. 305–327.
- [9] Briukhov D.O., Kalinichenko L. A., Skvortsov N. A. Information sources registration at a subject mediator as compositional development // East-European Conference “Advances in Databases and Information Systems” Proceedings. Lithuania, Vilnius, Springer, LNCS No. 2151, 2001.
- [10] Abrial J.-R. B-technology: Technical overview. — B-Core (UK) Ltd., 1993.
- [11] Abrial J.-R. The B-book: Assigning programs to meanings.— Cambridge: Cambridge University Press, 1996.
- [12] The B-toolkit online documentation. <http://www.bcore.com/ONLINEDOC/BToolkit.html>.
- [13] Zakharov V.N. , Kalinichenko L.A., Sokolov I.A., Stupnikov S.A. Development of Canonical Information Models for Integrated Information Systems. Informatics and Applications. -- 2007. -- V. 1, Issue 2. -- P. 15--38. (In Russian)
- [14] Kalinichenko L. A., Stupnikov S. A., Martynov D. O. SYNTHESIS: A language for canonical information modeling and mediator definition for problem solving in heterogeneous information resource environments. — M.: IPI RAS, 2007. 171 p.
- [15] Briukhov D.O., Kalinichenko L.A., Martynov D.O. Source Registration and Query Rewriting Applying LAV/GLAV Techniques in a Typed Subject Mediator. Proc. of the Ninth Russian Conference on Digital Libraries RCDL'2007.
- [16] Briukhov D.O., Vovchenko A. E., Zakharov V.N., Zhelenkova O. P., Kalinichenko L.A., Martynov D.O., Skvortsov N.A., Stupnikov S.A. The Middleware Architecture of the Subject Mediators for Problem Solving over a Set of Integrated Heterogeneous Distributed Information Resources in the Hybrid Grid-Infrastructure of Virtual Observatories Informatics and Applications. -- 2008. -- V. 2, Issue 1. -- P. 2—34.
- [17] Kalinichenko L.A., Briukhov D.O., Martynov D.O., Skvortsov N.A., Stupnikov S.A. Mediation Framework for Enterprise Information System Infrastructures. Proc. of the 9th International Conference on Enterprise Information Systems ICEIS 2007. -- Funchal, 2007. -- Volume Databases and Information Systems Integration. -- P. 246--251.
- [18] Kalinichenko L.A., Stupnikov S.A., Martynov D.O. SYNTHESIS: a Language for Canonical Information Modeling and Mediator Definition for Problem Solving in Heterogeneous Information Resource Environments. Moscow: IPI RAN, 2007.
- [19] Zakharov V.N. , Kalinichenko L.A., Sokolov I.A., Stupnikov S.A. Development of Canonical Information Models for Integrated Information Systems. Informatics and Applications. -- 2007. -- V. 1, Issue 2. -- P. 15--38.
- [20] Ioana Manolescu, Daniela Florescu, Donald Kossmann. Answering XML Queries over Heterogeneous Data Sources. Proc. of the Int'l. Conf. on Very Large Databases (VLDB) 2001, Roma, Italy,.
- [21] Ioana Manolescu, Daniela Florescu, Donald Kossmann, Florian Xhumari, Dan Olteanu. Agora: Living with XML and Relational. Proc. of the Int'l Conf. on Very Large Databases (VLDB) 2000, Cairo, Egypt.
- [22] Michael R. Genesereth, Arthur M. Keller, Oliver Duschka. Infomaster: An Information Integration System. in proceedings of 1997 ACM SIGMOD Conference, May 1997.
- [23] Oliver M. Duschka, Michael R. Genesereth. Infomaster - An Information Integration Tool. in proceedings of the International Workshop "Intelligent Information Integration" during the 21st German Annual Conference on Artificial Intelligence, KI-97. Freiburg, Germany, September 1997.
- [24] Patrick Ziegler. Evaluation of SIRUP with the SIRUP Classification of Data Integration Conflicts. Technical Report ifi-2007.07, Department of Informatics, University of Zurich, 2007.
- [25] Patrick Ziegler, Christoph Sturm, Klaus R. Dittrich. Unified Querying of Ontology Languages with the SIRUP Ontology Query API. In 11. GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web (BTW 2005), volume P-65 of Lecture Notes in Informatics, pages 325-344, Karlsruhe, Germany, March 2-4, 2005.
- [26] Patrick Ziegler, Klaus R. Dittrich. User-Specific Semantic Integration of Heterogeneous Data: The SIRUP Approach. In Mokrane Bouzeghoub, Carole Goble, Vipul Kashyap, and Stefano Spaccapietra, editors, First International IFIP Conference on Semantics of a Networked World (ICSNW 2004). volume 3226 of Lecture Notes in Computer

- Science, pages 44-64, Paris, France, June 17-19, 2004. Springer.
- [27] Chantal Reynaud, Gloria Giraldo. An application of the mediator approach to services over the Web. *Concurrent Engineering*, 2003.
- [28] Chantal Reynaud. Building scalable mediator systems. *Topical Day in Semantic Integration of Heterogeneous Data*, IFIP World Computer Congress, 2004.
- [29] Marie-Christine Rousset, Chantal Reynaud. Picles and Xyleme: two illustrative information integration agents. Book chapter in *Intelligent Information Agents Research and Development in Europe*, Springer-Verlag, 2003.
- [30] Alon Levy, Anand Rajaraman, Joann Ordille. Querying Heterogeneous Information Sources Using Source Descriptions. *Proceedings of the Twenty-second International Conference on VLDB 1996*.
- [31] Yannis Papakonstantinou, Hector Garcia-Molina, Jeffrey Ullman. MedMaker: A Mediation System Based on Declarative Specifications. *Proceedings of the 12th International Conference on Data Engineering*, 1995.
- [32] D. Beneventano, S. Bergamaschi. The MOMIS Methodology for Integrating Heterogeneous Data Sources. *IFIP World Computer Congress*. Toulouse France, 22-27 August 2004.
- [33] D. Beneventano, S. Bergamaschi, F. Guerra, M. Vincini. Building a Tourism Information Provider with the MOMIS System. *Information Technology & Tourism Journal*, 2005.
- [34] S. Chawathe, H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman, J. Widom. The TSIMMIS Project: Integration of Heterogeneous Information Sources. In *Proceedings of IPSJ Conference*, pp. 7-18, Tokyo, Japan, October 1994.
- [35] H. Garcia-Molina, Y. Papakonstantinou, D. Quass, A. Rajaraman, Y. Sagiv, J. Ullman, V. Vassalos, J. Widom. The TSIMMIS approach to mediation: Data models and Languages. In *Journal of Intelligent Information Systems*, 1997.
- [36] M. Boyd, P.J. McBrien. Comparing and Transforming Between Data Models via an Intermediate Hypergraph Data Model. *Journal on Data Semantics IV*, Pages 69-109, Springer-Verlag, 2005.
- [37] M. Boyd, S. Kittivoravitkul, C. Lazanitis, P.J. McBrien, N. Rizopoulos. AutoMed: A BAV Data Integration System for Heterogeneous Data Sources. In *Proceedings of CAiSE04*, Springer Verlag LNCS Vol 3084, Pages 82-97, 2004.
- [38] L. Zamboulis, A. Poulouvassilis. Using AutoMed for XML Data Transformation and Integration. In *Proceedings of DIWeb'04, CAiSE Workshop Proceedings Volume 3*, Pages 58-69.
- [39] L. Zamboulis, N. Martin, A. Poulouvassilis. A Uniform Approach to Workflow and Data Integration. In *Proceedings of U.K. e-Science All Hands Conference*, September, 2007.

Analysis and comparison of systems for heterogeneous information resources integration

Vovchenko A.E., Kalinichenko L.A.

Institute of informatics problems of RAS

Two approaches for problem solving in information systems over multiple heterogeneous integrated information resources are distinguished: resource driven and problem driven. In the paper the second approach is emphasized. One of the obstacles for such approach remaining to be unresolved consists in the integrated representation of multiple heterogeneous information resources for a researcher solving the problem in various infrastructures (such as grid architectures, Web services, Semantic Web, etc.). In the paper the basic principles of heterogeneous information resources integration are considered. These principles are used to make a choice of criteria for comparison of several well known integration system prototypes.

* Работа выполнена при частичной финансовой поддержке РФФИ (гранты 06-07-08072-офи-а, 06-07-89188-а, 08-07-00157-а) и программы ОНИТ РАН «Фундаментальные основы информационных технологий и систем» (проект 1-10).