



OSTIS-2014

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822

ИНТЕРНЕТ-КОМПЛЕКС ДЛЯ ПРЕОБРАЗОВАНИЙ КЛАССОВ СЕМАНТИЧЕСКИХ СЕТЕЙ

Тимченко В.А.

*Федеральное государственное бюджетное учреждение науки Институт автоматизации и процессов управления Дальневосточного отделения Российской академии наук,
г. Владивосток, Россия*

vadim@dvo.ru

В работе представлены архитектура и методы реализации комплекса средств для преобразований классов семантических сетей, включающих эндогенные и экзогенные преобразования, а также преобразования в разных технологических пространствах и между ними. Информационные компоненты программного комплекса хранятся в фонде платформы облачных вычислений IASaaS. Каждый информационный компонент формируется и поддерживается в актуальном состоянии с помощью соответствующего специализированного редактора, функционирующего как сервис платформы IASaaS.

Ключевые слова: семантические сети, средства преобразования семантических сетей, представление информации семантическими сетями, структурные проекции.

Введение

В настоящее время обработка информации, включающая в себя задачи по хранению, инженерии и обмену различными видами разнородных данных и знаний между программными сущностями и их компонентами, является одним из важнейших видов деятельности в большинстве теоретических исследований и прикладных областей. Одним из ключевых аспектов обработки является преобразование обрабатываемой программными сущностями информации из одной формы (языка, формализма) представления в другую. Различные графовые структуры и семантические сети успешно зарекомендовали себя в качестве модели представления разных видов данных, знаний, онтологий и т.п. и широко используются во многих предметных областях. Более того, несмотря на совершенно разный конкретный синтаксис форматов представления различных видов информации, в абстрактном синтаксисе они могут быть представлены в общем случае типизированными (мульти-)графами [Agrawal A. et al., 2005] [Ehrig K. et al., 2005].

В силу вышесказанного многие содержательные задачи преобразования семантической информации, возникающие в различных технологических пространствах [Kurtev, 2002] [Bezivin, 2005], а также на их стыке, можно сформулировать в терминах преобразования семантических сетей

следующим образом: по исходной семантической сети и спецификации преобразования сформировать новую семантическую сеть, удовлетворяющую этой спецификации.

К настоящему времени разработано множество прикладных и инструментальных средств поддержки, позволяющих эффективно специфицировать и выполнять преобразования информации из одной формы представления в другую при решении различных классов задач в разных технологических пространствах.

Однако большинство разработанных систем рассчитано, как правило, на выполнение только одно вида преобразований – эндогенных или экзогенных или на работу в рамках одного технологического пространства: в них используются специфические для технологического пространства способы и формализмы представления информации, метainформации, а также описания преобразований (либо используются специализированные адаптеры для импорта/экспорта информации в представление, необходимое для использования технологии и методов другого технологического пространства) [Czarnecki, 2006]. Так, например, средства, позволяющие одновременно выполнять преобразования вида “текст – семантическая сеть” и “семантическая сеть – текст”, можно найти только в технологическом пространстве grammarware (они предназначены для преобразования программ – XML, Stratego/XT). В области преобразования

графов (graphware) графовыми структурами представляют разного вида информацию, однако разработанные программные средства (например, PROGRES, AGG, GrGen.NET) позволяют выполнять только эндогенные преобразования. Помимо этого, большинство средств могут работать только в однопользовательском режиме, либо в локальной сети компании разработчиков, но не могут функционировать в сети Интернет, а значит, информация, обрабатываемая этими системами, хранится на локальных компьютерах и не доступна за рамками этой сети. Это сокращает доступность и масштаб практического применения таких средств, а также замедляет накопление опыта их практического использования.

Таким образом, разные задачи преобразования информации решаются различными методами, единый подход к преобразованию разных видов информации, представленных семантическими сетями, отсутствует. При возникновении новой задачи приходится искать адекватный метод решения и программное средство его поддержки. Поэтому если задача преобразования возникает на пересечении разных технологических пространств, и данный класс задач не охвачен ни одной программной системой, то под этот случай необходимо разрабатывать новое специализированное средство преобразования, либо адаптировать одно из существующих.

В данной работе представлены архитектура и методы реализации комплекса средств для преобразований классов семантических сетей, реализующего метод преобразования классов семантических сетей [Тимченко, 2013], основанный на модели классов семантических сетей и их преобразований, которые являются инвариантными по отношению к технологическим пространствам и позволяют определять как структурные преобразования, так и преобразования из текстового представления информации в структурное (в виде семантической сети) и наоборот [Тимченко, 2012].

1. Функциональность комплекса средств для преобразований классов семантических сетей

Этапы решения задачи преобразования информации, представленной семантическими сетями, предполагают выполнение следующих основных видов деятельности:

- описание класса семантических сетей информации (S) и формирование представления информации в виде семантической сети (S_{inst});
- описание спецификации преобразования класса семантических сетей – структурной проекции (M);
- преобразование информации.

Функциональность по описанию S , M , а также формированию S_{inst} реализуется с использованием соответствующих средств редактирования.

Функциональность по преобразованию информации реализуется с помощью Преобразователя классов семантических сетей, предоставляющего следующие возможности:

- возможность выполнять преобразование как структурного, так и текстового представления одной и той же информации, получать на выходе и сохранять как структурное, так и текстовое представление. В этом случае формирование S_{inst} выполняется Преобразователем классов семантических сетей по текстовому представлению этой информации (а не с помощью средства редатирования);
- возможность работы как в “сетевом” (через Интернет/Интранет), так и в “локальном” режиме. Последний позволяет пользователям работать с программным средством при отсутствии доступа к серверу, на котором развернута платформа облачных вычислений IACPaaS [Грибова и др., 2011];
- возможность формировать и сохранять графические представления семантических сетей (S и S_{inst});
- возможность протоколирования процесса преобразования, результатом которого является протокол, содержащий результаты, количественные и временные характеристики, полученные в ходе преобразования.

2. Архитектура комплекса средств для преобразований классов семантических сетей

В данном разделе описывается архитектура программного комплекса для преобразования классов семантических сетей на основе описания структурных проекций.

Пользователей программного комплекса можно разделить на две группы:

- **разработчики спецификации преобразования;**

- **прикладные пользователи.**

В задачи **разработчиков спецификации преобразования** входит:

- формирование Базы описаний классов семантических сетей (доступно только в “сетевом” режиме работы);
- формирование Базы структурных проекций (доступно только в “сетевом” режиме работы);
- формирование файлов с описанием классов семантических сетей;
- формирование файлов с описанием структурных проекций.

Для решения этих задач **разработчик спецификации преобразования** должен знать перечень метасимволов, используемых при описании синтаксических ограничений (S_{sr}), язык для записи выражений в описании структурных проекций, а также уметь работать с редактором семантических сетей и редактором структурных

проекций. Редактор семантических сетей используется для формирования Базы описаний классов семантических сетей в соответствии с моделью для описания классов семантических сетей, а редактор структурных проекций – для формирования Базы структурных проекций в соответствии с моделью для описания структурных проекций.

При формировании файлов – знать формат, в котором эти описания представляются в файлах. Для редактирования этих файлов может использоваться любой стандартный (например, Notepad, WordPad) или специализированный текстовый редактор.

В задачи **прикладных пользователей** входит:

1. формирование Базы семантических сетей – формирование представлений информации в виде семантических сетей и занесение их в базу (доступно только в “сетевом” режиме работы);

2. формирование файлов с текстовым представлением информации;

3. выполнение процесса преобразования семантических сетей представления информации.

Для решения **1-ой задачи прикладные пользователи** должны уметь работать с редактором семантических сетей, который в данном случае используется для редактирования Базы семантических сетей. С его помощью пользователь заносит в базу структурное представление информации, формируемое в соответствии с описанием структуры этой информации, которое затем при работе с Преобразователем классов семантических сетей можно выбрать для преобразования.

Для решения **2-ой задачи** достаточно воспользоваться любым стандартным (например,

Notepad, WordPad) или специализированным текстовым редактором для формирования текста и сохранения его в файл.

Для решения **3-ей задачи** используется Преобразователь классов семантических сетей.

Разработчик и сопровождающий программного средства формирует, а также расширяет (при необходимости) модель описания классов семантических сетей и модель описания структурных проекций с помощью соответствующих редакторов. Занимается сопровождением программного средства, если расширение данных моделей влечет за собой необходимость внесения изменений в его исходный код.

На рис. 1 представлена архитектура средства преобразования классов семантических сетей на основе описания структурных проекций как совокупность программных и информационных компонентов. На данной схеме отражены потоки данных между программными компонентами, а Преобразователь классов семантических сетей представлен в виде совокупности своих подсистем.

В архитектуре можно выделить следующие информационные компоненты:

- **модель описания классов семантических сетей** – определяет структуру и формализм для формирования описаний классов семантических сетей информации (S);
- **модель описания структурных проекций** – определяет структуру и формализм для описания структурных проекций (M);
- **база описаний классов семантических сетей** – содержит множество описаний классов семантических сетей информации (S);

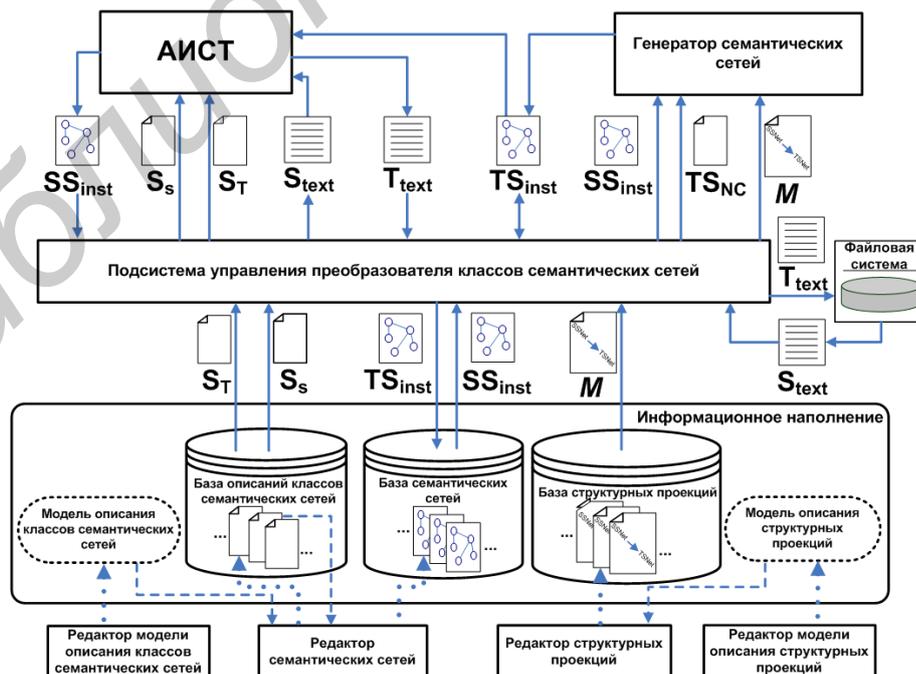


Рисунок 1 – Архитектура программного комплекса

- **база семантических сетей** – содержит множество представлений информации в виде семантических сетей (S_{inst});
- **база структурных проекций** – содержит множество описаний структурных проекций (M);
- под **файловой системой** обобщенно понимаются любые постоянные запоминающие устройства для хранения файлов, доступные на локальном компьютере пользователя, либо на удаленном компьютере, доступном через сетевой интерфейс.

Программными компонентами комплекса средств для преобразований классов семантических сетей являются:

- **редактор модели описания классов семантических сетей** – средство формирования и редактирования модели описания классов семантических сетей;
- **редактор модели описания структурных проекций** – средство формирования и редактирования модели описания структурных проекций;
- **редактор семантических сетей** – средство формирования и редактирования базы описаний классов семантических сетей и базы семантических сетей;
- **редактор структурных проекций** – средство формирования и редактирования базы структурных проекций;
- **преобразователь классов семантических сетей**, состоящий из трех модулей (подсистем):

1. **подсистема анализа и синтеза текстов (АИСТ)** – выполняет синтаксический анализ тестового представления информации ($Text$) и формирует S_{inst} , а также обратную процедуру – синтез $Text$ по S_{inst} ;

2. **генератор семантических сетей** – выполняет собственно преобразование S_{inst} : генерацию представления целевой информации в виде семантической сети (TS_{inst}) на основе представления исходной информации в виде семантической сети (SS_{inst}) по описанию класса сетей понятий целевой информации (TS_{nc}) и описанию структурной проекции (M) описания класса сетей понятий исходной информации (SS_{nc}) на TS_{nc} ;

3. **подсистема управления преобразователя классов семантических сетей** – организует внешний интерфейс с пользователем и внутренние интерфейсы с остальными подсистемами преобразователя, осуществляет загрузку входной информации с внешних источников данных и из информационных баз, а также сохранение выходной информации на внешних носителях.

3. Методы реализации комплекса средств для преобразований классов семантических сетей

В данном разделе описываются методы реализации комплекса средств для преобразования классов семантических сетей на основе описания

структурных проекций: описана реализация программного комплекса как клиент-серверного приложения, методы реализации средств редактирования и подсистем преобразователя классов семантических сетей.

3.1. Клиент-серверная реализация

Для обеспечения возможности работы с программным комплексом через Интернет/Инtranет он реализован как клиент-серверное приложение. На рис 2 представлена его клиент-серверная архитектура.

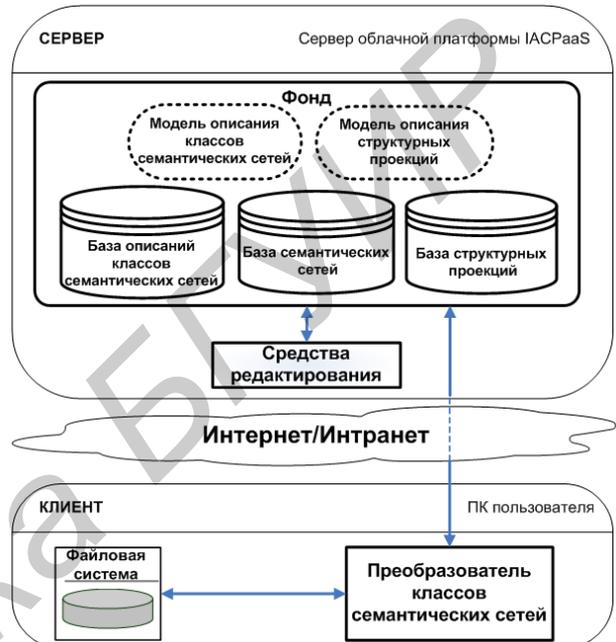


Рисунок 2 – Клиент-серверная архитектура программного комплекса

Все информационные компоненты программного комплекса расположены на **сервере**, на котором платформа облачных вычислений *IACPaaS*, в *фонде* платформы. Фонд представляет собой совокупность единиц хранения – программных и информационных ресурсов различных типов.

Весь фонд разделен на предметные области, а те, в свою очередь, на разделы. Для каждого информационного компонента комплекса средств для преобразований классов семантических сетей на основе описания структурных проекций через административную систему платформы *IACPaaS* зарегистрирован и создан соответствующий ему *информационный ресурс* в фонде.

Все средства редактирования представляют собой облачные сервисы платформы *IACPaaS*, разработанные в соответствии с методологией и технологией проектирования интеллектуальных многоагентных сервисов с использованием облачной платформы *IACPaaS*.

На **клиентском компьютере** пользователя находится Преобразователь классов семантических сетей, который для доступа к информационным компонентам обращается к функциям программного

интерфейса (API) для (удаленного) доступа сервисов к фонду платформы IACPaaS.

3.2. Методы реализации средств редактирования

Функциональность редактора модели описания классов семантических сетей и редактора модели описания структурных проекций реализуется интеллектуальным сервисом редактирования программных и информационных ресурсов различного уровня общности (редактором IWE), который позволяет создавать и модифицировать информацию в соответствии с моделью структурированного повторно используемого представления единиц хранения фонда платформы IACPaaS [Орлов, 2006].

В фонде IACPaaS основной формой представления любого программного и информационного ресурса является семантическая сеть понятий. Таким образом, каждый информационный компонент комплекса средств для преобразования классов семантических сетей на основе описания структурных проекций хранится в информационном наполнении в виде сети или совокупности сетей понятий.

Редактор IWE используется, также, как инструментальное средство для создания специализированных (прикладных) средств редактирования информационных ресурсов фонда IACPaaS. Чтобы получить специализированный (прикладной) редактор “объектной информации”, необходимо специфицировать метаинформацию и занести ее в фонд IACPaaS. Требуемый специализированный редактор получается подключением такой метаинформации к редактору IWE в качестве управляющей информации. Таким образом, управляемый в процессе редактирования метаинформацией, редактор IWE становится по существу специализированным (прикладным) редактором “объектной информации” в терминах, специфицированных в метаинформации (рис. 3).

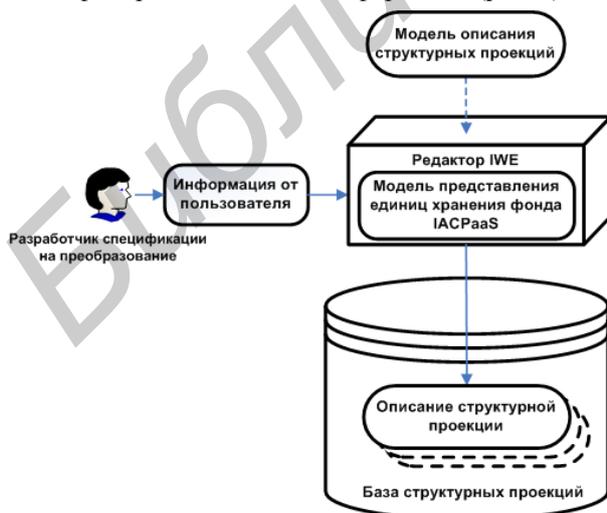


Рисунок 3 – Редактирование структурных проекций с помощью редактора IWE

Так, специализированные редакторы для

формирования описаний классов семантических сетей (S) и структурных проекций (M), ориентированные на разработчика спецификации преобразования, получаются в результате подключения к редактору IWE модели описания классов семантических сетей и модели описания структурных проекций соответственно. Специализированный редактор для формирования представлений информации в виде семантических сетей (S_{inst}), ориентированный на прикладного пользователя, получается в результате подключения к редактору IWE описания класса сетей понятий (S_{nc}) в качестве метаинформации. После этого редактор используется для формирования множества S_{inst} , причём для каждого элемента этого множества S_{nc} является метаинформацией.

3.3. Методы реализации преобразователя классов семантических сетей

Подсистема анализа и синтеза текстового представления информации (АИСТ) преобразователя классов семантических сетей состоит из двух компонентов – модуля анализа и модуля синтеза, реализующих базовые методы преобразования “текст – семантическая сеть”, “семантическая сеть – текст” соответственно [Тимченко, 2013].

Генератор семантических сетей реализует базовый метод преобразования “семантическая сеть – семантическая сеть” [Тимченко, 2013]. Генерация может быть завершена успешно, если каждое правило в M определяет описание устройства подсети TS_{inst} , которую возможно породить на основе подсети из описания класса сетей понятий TS_{nc} . Если же в M обнаружено правило, которое определяет описание устройства подсети TS_{inst} , которая не может быть порождена на основе подсети из описания класса сетей понятий TS_{nc} , то процесс генерации завершается с ошибкой.

Подсистема управления преобразователя классов семантических сетей состоит из трёх компонентов – *Интерфейса пользователя*, *Организатора* и *Отображателя*.

Интерфейс пользователя инкапсулирует реализацию классического WIMP-интерфейса, характерного для оконных операционных систем.

Организатор инкапсулирует основную функциональность подсистемы управления. Он организует внутренние интерфейсы с остальными компонентами и подсистемами программного средства и управляется пользователем через интерфейс. Организатор осуществляет кэширование информации в оперативной памяти, ведёт протокол процесса преобразования, а также загружает входные данные и сохраняет выходные данные в файловой системе компьютера пользователя, запоминает и восстанавливает настройки программного средства между сеансами работы.

Отображатель инкапсулирует функциональность по загрузке входных данных из фонда платформы IACPaaS во внутренние структуры данных

преобразователя в оперативной памяти, а также сохранения выходных данных в фонде. Взаимодействие Отображателя с фондом IACPaaS осуществляется через API доступа к фонду. Чтобы добавить поддержку отображения входных и выходных данных в любое другое нужное представление, понадобится расширить только этот компонент подсистемы управления, остальные подсистемы эти изменения не затронут.

В качестве языка программирования всех компонентов Преобразователя классов семантических сетей использовался язык Java. Общий объем исходного кода преобразователя (без учета подключаемых библиотек) ~ 27000 строк.

Заключение

Разработаны методы реализации программного комплекса для преобразования информации, представленной семантическими сетями, на основе описания структурных проекций. Для обеспечения процесса раздельного, независимого (от программных компонентов) сопровождения информации соответствующими специалистами программный комплекс спроектирован с соблюдением принципа разделения программных компонентов и обрабатываемой ими информации, которая должна храниться отдельно и редактироваться независимо от средств ее обработки. Все информационные компоненты выделены в отдельные информационные ресурсы фонда облачной платформы IACPaaS. Для каждого компонента разработан свой специализированный редактор, позволяющий удаленно формировать и поддерживать в актуальном состоянии соответствующий информационный ресурс. Преобразователь классов семантических сетей обеспечивает возможность работы как в “сетевом” (через Интернет/Интранет), так и в “локальном” режиме, а также позволяет формировать и сохранять графические представления семантических сетей.

Показана практическая применимость разработанного комплекса программных средств при решении задач преобразования информации, возникающих в разных технологических пространствах. Для этого была подготовлена выборка примеров задач преобразования информации, которая покрывала бы весь спектр видов преобразований: эндогенных и экзогенных, преобразований внутри технологических пространств и между ними. При этом применяются все базовые типы преобразований: “семантическая сеть – семантическая сеть”, “текст – семантическая сеть” и “семантическая сеть – текст”. Помимо этого, задействуются все три комбинации базовых типов преобразований: “семантическая сеть – семантическая сеть – текст”, “текст – семантическая сеть – семантическая сеть” и “текст – семантическая сеть – семантическая сеть – текст”. Результаты практического применения показали, что разработанный программный комплекс позволяет успешно решать практические задачи.

Работа выполнена при финансовой поддержке РФФИ, проект 12-07-00179-а, проект 13-07-00024-а и ДВО РАН по Программе ОНИТ, проект 12-I-ОНИТ-04.

Библиографический список

- [Agrawal et al., 2005] Reusable Idioms and Patterns in Graph Transformation Languages / Agrawal, A. [et al.]; // Journal Electronic Notes in Theoretical Computer Science (ENTCS). – Vol. 127. – Issue 1. – 2005. – Pp. 181 – 192.
- [Ehrig et al., 2005]. Model Transformation by Graph Transformation: A Comparative Study / Ehrig, K. [et al.]; // Proceedings of Model Transformations in Practice Workshop, MoDELS Conference. – Jamaica. – 2005. [Electronic resource]. URL: <http://www.inf.mit.bme.hu/FTSRG/Publications/varro/2005/mtip05.pdf> (дата обращения: 26.11.2012).
- [Kurtev, 2002] Kurtev, I. Technological Spaces: An Initial Appraisal / I. Kurtev, J. Bezivin, M. Aksit // Int. Federated Conf. (DOA, ODBASE, CoopIS), Industrial track, Irvine. – 2002.
- [Bezivin, 2005] Bezivin, J. Model-based Technology Integration with the Technical Space Concept / J. Bezivin, I. Kurtev // In Proceedings of the Metainformatics Symposium, Springer-Verlag. – 2005.
- [Czarnecki, 2006] Czarnecki, K. Feature-based survey of model transformation approaches / K. Czarnecki, S. Helsen // IBM Systems Journal, special issue on Model-Driven Software Development. 45(3). – 2006. – Pp. 621 – 645.
- [Тимченко, 2013] Тимченко, В. А. Метод преобразования классов семантических сетей / В. А. Тимченко // Материалы международной научно-технической конференции OSTIS-2013. – Минск БГУИР. – 2013. – С. 81 – 86.
- [Тимченко, 2012] Тимченко, В. А. Модель классов семантических сетей и их преобразований / В. А. Тимченко // Материалы международной научно-технической конференции OSTIS-2012. – Минск БГУИР. – 2012. – С. 63 – 70.
- [Грибова и др., 2011] Грибова В.В. [и др.]; // Материалы международной научно-технической конференции OSTIS-2011. – Минск БГУИР. – 2011. – С. 5 – 14.
- [Орлов, 2006] Орлов, В. А. Компьютерные банки знаний. Универсальный подход к решению проблемы редактирования информации / В. А. Орлов, А. С. Клещев // Информационные технологии. – 2006. – №5. – С. 25 – 31.

THE INTERNET SOFTWARE TOOL FOR SEMANTIC NETWORK CLASSES TRANSFORMATION

Timchenko V.A.

*Federal State Budget Institution of Science
Institute of Automation and Control Processes Far
Eastern Branch of the Russian Academy of
Sciences, Vladivostok, Russia*

vadim@dvo.ru

The paper presents the architecture and implementation methods of the toolset for semantic network classes transformations. These transformations include endogenous and exogenous transformations as well as transformations in different technological spaces and between technological spaces. Information components of the toolset are stored in the data warehouse of cloud-computing software platform IACPaaS. Each information component is created and supported in an actual state by the appropriate specialized editor, which is running as an internet-service of the software platform IACPaaS.