



## GRID-DL – СЕМАНТИЧЕСКИЙ ИНФОРМАЦИОННЫЙ СЕРВИС ГРИД

**Александр Поспешный, Сергей Стиренко**

Национальный технический университет Украины “КПИ”  
пр. Победы, 37, 03056, Киев, Украина  
pospishniy@kpi.in.ua, stirenko@ugrid.org

**Резюме:** Информационный сервис является краеугольным камнем любой современной Грид-системы. С их помощью другие подсистемы Грид могут получать информацию о имеющихся ресурсах системы, их состоянии и характеристиках. Для борьбы с некоторыми ограничениями традиционных информационных сервисов мы решили применить семантические технологии развивающиеся в рамках концепции Семантической Паутины. В нашей работе мы представляем разработанный нами прототип семантического информационного сервиса Грид.

**Ключевые слова:** Грид, информационный сервис, база знаний, семантика, онтологии, OWL.

## GRID-DL – SEMANTIC GRID INFORMATION SERVICE

**Oleksandr Pospishniy, Sergiy Stirenko**

National Technical University of Ukraine “KPI”  
37 Prospect Peremogy, Kiev 03056, Ukraine  
pospishniy@kpi.in.ua, stirenko@ugrid.org

**Abstract:** The effectiveness of complex Grid systems strongly depends on the availability, accuracy and relevance of information on all connected resources, their characteristics and state. An access to this information plays a very important role in any Grid system, providing necessary information for other Grid components and users. We believe that application of semantic technologies can improve Grid utilization and enhance user interaction with the system. We present our vision of the semantic Grid resource information service.

**Keywords:** Grid, information service, semantics, ontology, OWL.

### ВСТУПЛЕНИЕ: ЦЕЛЬ И ПЕРСПЕКТИВЫ ДАННОЙ РАБОТЫ

Можно с уверенностью заявить, что Грид-системы стали новым мощным и, в некоторой степени, даже необходимым инструментом для современной науки и инженерии. С их помощью стало возможным агрегировать географически распределённые ресурсы и предоставлять к ним коллективный доступ для группы специалистов, которые совместно работают над достижением общих целей. Грид позволяет отделить потребителей вычислительных ресурсов от “производителей” этих ресурсов и реорганизовать связи между ними таким образом, что бы достичь максимальной производительности системы в целом.

Если посмотреть на структуру любой Грид

системы можно легко заметить её исключительную гетерогенность, так как все структурные составляющие Грида уникальны в отношении их программного и аппаратного обеспечения. Поэтому управление и эффективное использование Грид систем невозможно без наличия полной, достоверной и актуальной информации о всех подключенных ресурсах, их характеристиках, состоянии и политике использования. В Грид системах эту функцию выполняют информационные сервисы. Без быстрой, надёжной, масштабируемой и удобной динамической информационной системы в принципе невозможно реализовать качественный Грид-проект.

Любая информационная система играет роль нервной системы Грид и обеспечивает следующие функции:

- описание ресурсов (определение и формализация статических и динамических параметров ресурсов Грид)

- агрегация ресурсов (объединение и структурирование распределённой информации о ресурсах Грид системы)

- поиск ресурсов (обеспечение инструментов по доступу к накопленной информации)

Наиболее известные информационные сервисы это MDS[1] из пакета Globus, BDI[2] из промежуточного программного обеспечения gLite, а также ARC Information System[3], RGMA[4] и другие.

Большинство из вышеупомянутых информационных сервисов основаны на технологиях стандарта серии X.500 для служб распределённого каталога сети. Каталоги полностью структурированы и предоставляют централизованную информацию обо всех именованных объектах сети в форме иерархически скомпонованных объектов.

Изначально стандарт X.500 планировался для использования именованных узлов, адресов и почтовых ящиков, предусмотренных стандартом X.400 и был оптимизирован для хранения статических и редко изменяемых данных для очень быстрого отклика на запросы поиска и чтения. При этом информация хранится в текстовой форме и каждый атрибут элемента, как правило, может иметь несколько значений

Для доступа к информации используется протокол прикладного уровня LDAP (англ. Lightweight Directory Access Protocol – “облегчённый протокол доступа к каталогам”). Для процедур импорта и экспорта данных, всеми серверами LDAP поддерживается единый формат обмена данными – LDIF.

Информационные сервисы научились эффективно использовать эти технологии для работы с динамической природой Грид систем, но при этом остаются заложниками их ограничений.

Во-первых, все пользователи Грид обязаны использовать единую общую схему[5], с помощью которой происходит именование ресурсов. Как следствие, такие схемы зачастую изрядно перегружены, специализированы, трудны в использовании и поддержке.

Во-вторых, формат атрибутов ресурсов в службе каталогов приводит к частым синтаксическим ошибкам, иногда неоднозначен и способствует низкой эффективности поиска ресурсов.

Ещё одним крупным недостатком традиционных информационных сервисов является принцип симметрического сравнения при поиске ресурсов, как это сделано в Condor[6] и PBS. Такой подход позволяет найти только те ресурсы, описание которых строго соответствует запросу, и почти не допускает никакой гибкости и интел-

лектуального анализа при поиске. Это приводит к потере общей эффективности Грид системы, и повышает количество задач, которые не смогли успешно завершиться[7] или были погружены на не оптимальные узлы.

Вдобавок к уже перечисленным недостаткам можно упомянуть плохую расширяемость традиционных информационных сервисов, сложность их постепенной эволюционной модернизации и не дружелюбность к пользователю. Последнее замечание особо важно в свете движения за популяризацию Грид технологий среди не специалистов в области IT.

Для борьбы с названными недостатками мы решили применить семантические технологии, развивающиеся в рамках концепции Семантической Паутины. С их помощью мы планируем создать “интеллектуальный” информационный сервис Грид. В основе такого сервиса будет лежать база знаний о ресурсах Грид, построенная с помощью системных и пользовательских онтологий. Пользователь сможет применять высокоуровневое и дружественное описание решаемой им задачи в понятных ему терминах, а необходимые для этого программные и аппаратные ресурсы будут определены и найдены при помощи системы логических выводов.

Другим преимуществом такого подхода будет надёжный механизм проверки требований Грид-задачи на корректность, так как можно будет мгновенно проверить запрос на предмет логических несоответствий, что поможет минимизировать ошибочные назначения и простой ресурсов в связи с пользовательскими ошибками.

Мы считаем, что применение семантических технологий позволит существенно упростить использование Грид систем не специалистами в области IT, что будет способствовать популяризации этой технологии.

Посредством предложенного подхода возможно также создание “онтологических мостов” между различными реализациями Грид, где существуют различные концепции и понятия, которые имеют, по сути, схожий смысл. Таким образом, информационные сервисы одной Грид-системы могут корректно интерпретировать данные другой Грид-системы, облегчая процесс их интеграции.

В этой работе мы представляем прототип семантического информационного сервиса Grid-DL на котором мы апробируем нашу гипотезу о возможности и целесообразности применения семантических технологий в Грид.

## 1. АРХИТЕКТУРА

Grid-DL представляет собой автономное Веб-приложение содержащее набор веб-служб и веб-

интерфейс. Проект реализовано на платформе Java версии 1.7. Для компиляции и сборки приложения используется утилита для автоматизации процесса сборки программного обеспечения Apache Ant. Результатом процесса сборки является упакованное веб-приложение в виде war-файла, готовое для внедрения и запуска в контейнере сервлетов Apache Tomcat или J2EE-совместимом сервере приложений.

Для сборки и корректной работы Grid-DL необходимы:

- Oracle Java Development Kit 7
- Apache Ant 1.8+
- Apache Tomcat 7.0+ или совместимый контейнер сервлетов
- Mercurial SCM 1.8+

Указанные версии являются рекомендованными. Возможно применение более старых версий указанных продуктов, однако работоспособность приложения на подобных конфигурациях не проверялась.

На Рис. 1 изображена обобщенная архитектура сервера и его вспомогательных компонентов. Можно выделить 4 основных структурных элемента:

1. опорный информационный сервис Грид,
2. семантический информационный сервис,
3. репозиторий доменных онтологий,
4. пользователи и клиенты сервиса.

Рассмотрим их в подробностях.

## 1.1. ОПОРНЫЙ ИНФОРМАЦИОННЫЙ СЕРВИС

На данном этапе реализации нашей идеи мы решили не создавать свою распределённую архитектуру мониторинга ресурсов, а использовать уже имеющуюся. Таким образом, семантический информационный сервис извлекает необходимые ему для работы данные о Грид ресурсах из корневого информационного сервиса Грид. В качестве такого корневого информационного сервиса мы рассматриваем системы BDII, GIIIS и EGIIIS которые используются в промежуточном программном обеспечении gLite, Globus и ARC соответственно.

Для универсализации мы предусмотрели механизм подключаемых адаптеров, с помощью которых Grid-DL подключается к необходимым типам информационных сервисов. В последствии все данные полученные таким образом будут приведены к обобщённому представлению, которое инвариантно относительно своего источника благодаря применению онтологий.

## 1.2. СЕМАНТИЧЕСКИЙ ИНФОРМАЦИОННЫЙ СЕРВИС

Исходя из самой философии Грид-систем, целесообразно различать два уровня: общий для всей системы в целом, и уровень виртуальной организации, в рамках которой оперируют поль-

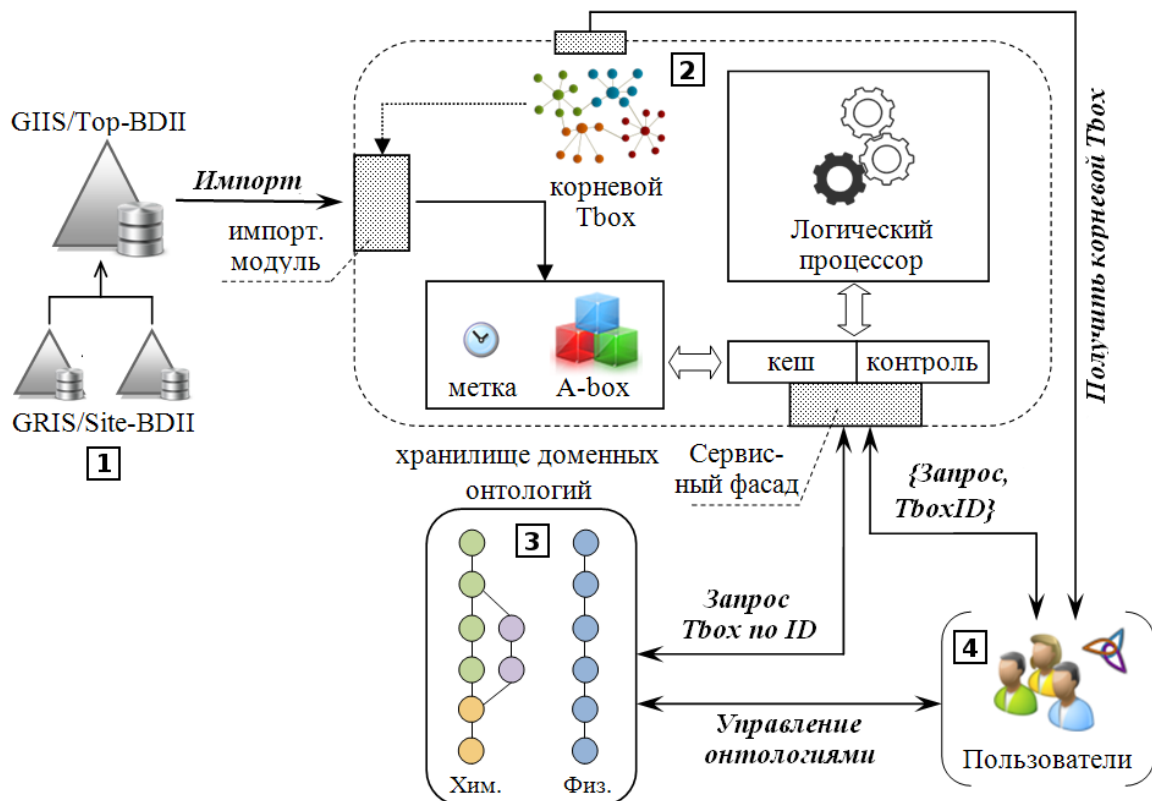


Рис. 1 – Обобщённая архитектура Grid-DL

зователи. Таким образом, семантический информационный сервис (Рис. 2) должен отображать эту особенность путём использования двух онтологий: системной и пользовательской. Общая Грид онтология будет всеохватывающей, относительно статичной и доступной в режиме “только чтение”. А так как виртуальные организации (ВО) формируются для достижения вполне определённых целей, то целесообразно расширять общую онтологию более детальным доменным описанием специфики этих организаций. Участники ВО будут совместно участвовать в формировании такой онтологии, дополняя и уточняя её по мере необходимости. Пользовательская онтология, таким образом, будет относительно динамична, доступна в режиме “чтение-запись” и защищённая системой контроля версий.

Получив доступ к информационному сервису Грид-системы через соответствующий адаптер, импортирующий модуль (a) формирует ABox онтологию (b) содержащую множество утверждений о ресурсах. Все утверждения этой внутренней онтологии формируются с использованием терминов доступных из корневой TBox онтологии (c). Корневая онтология максимально полно и всеобъемлюще описывает все возможные типы ресурсов доступные в Грид-системе и их характеристики. Детальную информацию о содержании корневой онтологии можно почерп-

нуть из нашей ранней работы [8]. Пользователь может получить доступ к корневой онтологии через соответствующий веб-сервис (d). TBox онтология доступна только для чтения и идентифицируется по номеру версии.

Для процессинга онтологий используется OWL-процессор Pellet [9] (e) подключенный через интерфейс OWLAPI. В системных настройках Grid-DL возможно указать любой другой OWLAPI-совместимый процессор онтологий который будет использоваться для работы с онтологиями в системе.

Всё взаимодействие с семантическим информационным сервисом происходит посредством веб-служб (f). Для реализации этого компонента мы воспользовались библиотекой JAX-WS, которая является частью платформы J2EE. Использование веб-сервисов, таким образом, позволяет обеспечить нам интероперабельность с любой платформой и языком программирования.

Все запросы, поступающие на Grid-DL, проходят предварительную валидацию (g) на предмет логического несоответствия и семантических ошибок. Дополнительно, с целью повысить производительность, запросы (а также их результаты) кешируются (h).

### 1.3. РЕПОЗИТАРИЙ ДОМЕННЫХ ОНТОЛОГИЙ

Репозиторий доменных онтологий представляет собой платформу для коллективной разра-

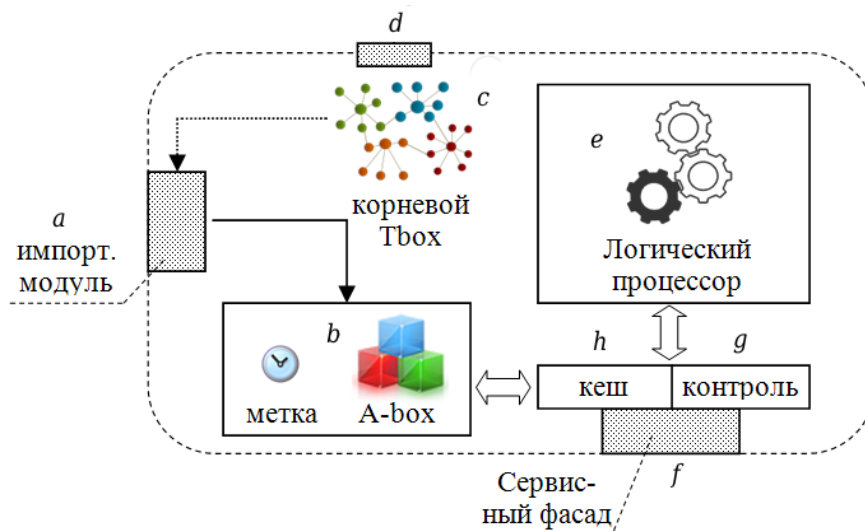


Рис. 2 – Структура семантического информационного сервиса

ботки, уточнения и обмена доменными онтологиями которые будут использованы при работе с семантическим информационным сервисом. Данный компонент представляет собой самостоятельный сервер с установленной системой управления версиями. Таким образом, пользователи могут участвовать в совместной разработке доменных онтологий или пользоваться уже гото-

нующий из нашей ранней работы [8]. Пользователь может получить доступ к корневой онтологии через соответствующий веб-сервис (d). TBox онтология доступна только для чтения и идентифицируется по номеру версии.

ABox онтология, содержащая информацию о ресурсах Грид, аннотируется временной меткой. Таким образом, результат запроса к информаци-

выми. Семантический информационный сервис обращается к репозитарию доменных онтологий при обработке запроса, при этом используя набор онтологий указанных пользователем.

В качестве репозитария мы использовали распределённую систему контроля версий Mercurial[10], в которой поставлен акцент на производительность, простоту и передовые методики слияния веток. Для работы с репозитарием достаточно установить любой совместимый Mercurial-клиент, однако мы планируем в будущем интегрировать его в редактор онтологий Protégé [11].

Взаимодействие клиентов и семантического информационного сервиса с репозитарием происходит посредством протокола HTTP.

Для работы с Mercurial в Grid-DL предусмотрен адаптер для этой системы. Допускается подключение других адаптеров, для использования других систем контроля версии, в частности Git или Subversion.

#### 1.4. ПОЛЬЗОВАТЕЛИ И КЛИЕНТЫ GRID-DL

Так как всё взаимодействие с семантическим информационным сервисом происходит посредством веб-сервисов, клиентом Grid-DL может выступать любое приложение с поддержкой стандарта веб-служб (URI, XML, SOAP, WSDL). Описание предоставляемых веб-сервисов доступно по адресу “адрес сервера:порт/Grid-DL/ServiceFacade?wsdl”.

Для взаимодействия с Grid-DL доступны следующие веб-методы:

- getTBox (получение корневой онтологии)
- getTBoxVersion (получение версии текущей корневой онтологии)
- submitQuery (размещение запроса к семантическому информационному сервису. Сервис возвращает уникальный идентификатор запроса, по которому можно будет проверить его состояние и результат)
- getQueryTask (получение информации о состоянии запроса и его результата. В качестве параметра принимает уникальный идентификатор запроса в системе.)

Для администрирования Grid-DL и просмотра состояния всех поступивших запросов к информационному сервису был разработан веб-интерфейс доступный в веб-браузере по адресу “адрес сервера:порт/Grid-DL/qtasks”.

## 2. ВЗАИМОДЕЙСТВИЕ С СЕМАНТИЧЕСКИМ ИНФОРМАЦИОННЫМ СЕРВИСОМ

В самом простом варианте, не принимая в расчёт процесс разработки и использование доменных онтологий, обращение к семантическому информационному сервису происходит так, как показано на Рис. 3. Пользователь должен предварительно ознакомиться с содержанием корневой TBox онтологии и составить запрос к сервису используя манчестерский OWL синтаксис[12].

Запрос должен представлять собой OWL-класс, экземпляры которого, после процесса классификации, будут представлять искомую информацию. При подаче запроса, семантический информационный сервис возвращает его уникальный идентификатор в системе, по которому можно получить результат его выполнения.

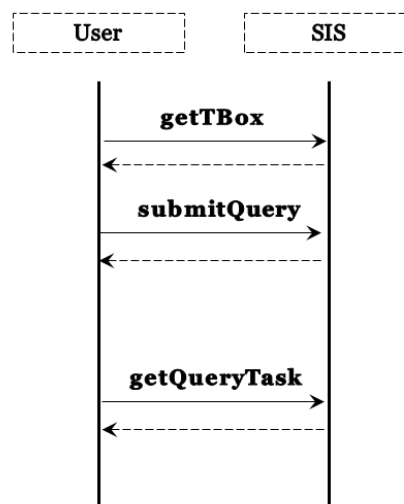


Рис. 3 – Простейшее взаимодействие пользователя с семантической информационной системой

Теперь рассмотрим схему взаимодействия с Grid-DL, принимая во внимание использование репозитария доменных онтологий. Как показано на Рис. 4 пользователю необходимо указать в своём запросе те онтологии, которые необходимо будет использовать при обработке запроса. Путь к онтологии задаётся относительно корня репозитария доменных онтологий. Например: “/general/operatingSystems.owl”

После подачи запроса семантический информационный сервис проведёт синхронизацию с репозитарием доменных онтологий, получит новую версию онтологии, если такова будет и подключит её к OWL-процессору перед выполнением обработки запроса.

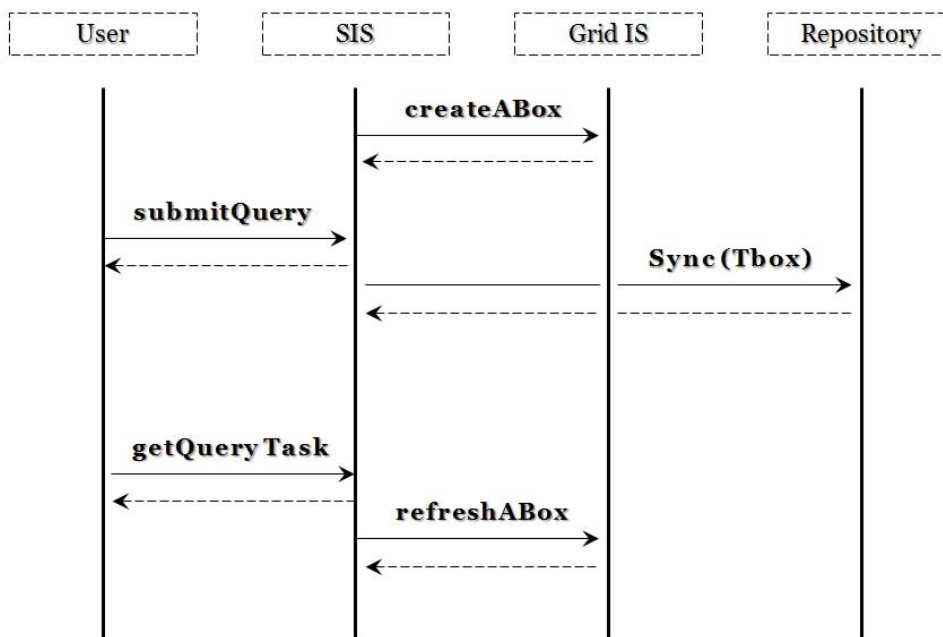


Рис. 4 – Типичное взаимодействие пользователя с семантической информационной системой

Взаимодействие пользователей с репозитарием доменных онтологий происходит по правилам работы с распределённой системой контроля версий SCM Mercurial (Рис. 5). Она обеспечивает целостность файлов, простой механизм доступа к ним, просмотр истории развития онтологий и механизмы по коллаборативной разработке онтологий пользователями. Мы рассматриваем возможность интегрировать Mercurial клиент в редактор онтологий Protégé для более комфортной работы пользователя с системой.

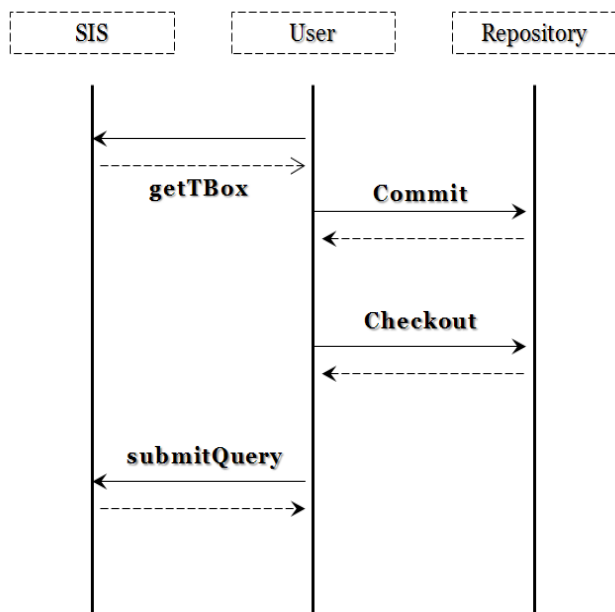


Рис. 5 – Взаимодействие пользователя с репозитарием онтологий

### 3. ПРИМЕР ИСПОЛЬЗОВАНИЯ СЕМАНТИЧЕСКОГО ИНФОРМАЦИОННОГО СЕРВИСА

Представим себе неопытного пользователя Грид, который хочет решить свою задачу молекулярной динамики в рамках доступных для его виртуальной организации ресурсов.

Грид, к которому у него есть доступ, содержит тысячи рабочих узлов, с разными конфигурациями. Он обращается к онтологии своей виртуальной организации и получает следующие определения:

*MolecularDynamicsSim*  $\equiv$  *GROMACS\_App* or *LAMMPS\_App*

*GROMACS\_App*  $\equiv$  *ApplicationSoftware* and *hasRunTimeEnvironment* some string [pattern "GROMACS"]

*LAMMPS\_App*  $\equiv$  *ApplicationSoftware* and *hasRunTimeEnvironment* some string [pattern "LAMMPS"]

*GROMACS\_Host*  $\equiv$  *Host* and *describedBy* some *GROMACS\_App*

*LAMMPS\_Host*  $\equiv$  *Host* and *describedBy* some *LAMMPS\_App*

*GROMACS\_Cluster*  $\equiv$  *MPI\_SubCluster* and *X86\_64\_SubCluster* and (*SubCluster* and *describedBy* some *GROMACS\_Host*)

*LAMMPS\_Cluster*  $\equiv$  *MPI\_SubCluster* and *X86\_64\_SubCluster* and (*SubCluster* and *describedBy* some *LAMMPS\_Host*)

*MolDynSubCluster*  $\equiv$  *GROMACS\_Cluster* or *LAMMPS\_Cluster*

*MolDynCE*  $\equiv$  *ComputingElement and partOf some (Cluster and contains some MolecularDynamicsSubCluster)*

В этой онтологии, среди прочего, определены программные пакеты необходимые для работы и определение кластера способного их выполнить. При этом определения *MPI\_SubCluster* и *X86\_64\_SubCluster* находятся в более общей онтологии, которая используется всеми виртуальными организациями. В частности там будет определение кластера с поддержкой технологии MPI и платформы x86-64:

*OPENMPI*  $\equiv$  *ApplicationSoftware and hasRunTimeEnvironment value "OPENMPI"*

*MPICH*  $\equiv$  *ApplicationSoftware and hasRunTimeEnvironment value "MPICH"*

*MPI\_Library*  $\equiv$  *MPICH or OPENMPI*

*MPI\_Host*  $\equiv$  *Host and describedBy some MPI\_Library*

*MPI\_SubCluster*  $\equiv$  *SubCluster and describedBy some MPI\_Host*

*MPI\_Cluster*  $\equiv$  *Cluster and contains some MPI\_SubCluster*

*X86\_64\_Arch*  $\equiv$  *Architecture and hasPlatformType value "x86\_64"*

*X86\_64\_Host*  $\equiv$  *Host and describedBy some X86\_64\_Arch*

*X86\_64\_SubCluster*  $\equiv$  *SubCluster and describedBy some X86\_64\_Host*

*X86\_64\_Cluster*  $\equiv$  *Cluster and contains some X86\_64\_SubCluster*

На этом этапе пользователь может добавить свои личные определения, как например понятие незагруженного вычислительного элемента и окончательно определить узел для работы:

*Available\_CE*  $\equiv$  *ComputingElement and hasState some (CEState and hasRunningJobs value 0 and hasWaitingJobs value 0 and hasFreeJobSlots some integer[>0])*

*MyVoACL*  $\equiv$  *AccessControlBaseRule and hasPrefix value "VO" and hasSCN value "myVO"*

*CEForMyWork*  $\equiv$  *MolDynCE and Available\_CE and hasAccessControlBaseRule some MyVoACL*

Запрос к семантическому информационному сервису *CEForMyWork* с указанием используемых онтологий вернёт список подходящих пользователю узлов Грид. При этом пользователь изолирован от сложности информационной схемы Грид-системы онтологиями.

#### 4. ЗАКЛЮЧЕНИЕ

Применение семантических технологий открывает множество возможностей и перспектив дальнейшего совершенствования базовых элементов Грид систем. С их помощью предпола-

ется создание семантического информационного сервиса, который будет основан на технологиях интеллектуального управления знаниями.

Представленный в данной работе прототип свободно доступен для применения и доработки [13]. Мы используем его как платформу для проведения исследований в области Грид и семантических технологий и побуждаем всех желающих присоединиться к этому процессу.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Czajkowski K., Fitzgerald S., Foster I., Kesselman C., Grid information services for distributed resource sharing, *Proc. of the 10-th IEEE International Symposium on High Performance Distributed Computing*. – IEEE Press. (2001). – pp. 181-195.
- [2] Документация к проекту Berkeley Database Information Index V5. –<https://twiki.cern.ch/twiki/bin/view/EGEE/BDII/>
- [3] Konya B., The NorduGrid/ARC information system, *The NorduGrid Collaboration*, [http://www.nordugrid.org/documents/arc\\_infos.pdf](http://www.nordugrid.org/documents/arc_infos.pdf).
- [4] Byrom R., Cornwall L., Djaoui A., et al., R-GMA an information integration system for grid monitoring, *Proc. of International Conference on Cooperative Information Systems*, Springer Press. (2003).
- [5] Andreozzi S., Burke S., Donno F. et al., *GLUE Schema Specification (version 1.3)*, <http://glueschema.forge.cnaf.infn.it/Spec/V13>.
- [6] Thain D., Tannenbaum T., Livny M., Distributed computing in practice: the Condor experience, *Concurrency and Computation: Practice and Experience*. 17 (2-4) 2005. – p. 323-356.
- [7] Neocleous K., Dikaiakos MD., Fragopoulou P., et al. Failure Management in Grids: the Case of the EGEE Infrastructure // *Parallel Processing Letters*. – vol. 17. –2007. –pp. 391-410.
- [8] Поспешный А.С., Стиренко С.Г. Онтология ресурсов для семантического информационного сервиса Грид, *Электронное моделирование*. 33 (4) (2011).
- [9] Sirin E., Parsia B., Grau B. et al., Pellet: A practical OWL-DL reasoner, *Web Semantics: science, services and agents on the World Wide Web*, 5 (2) (2007). – pp. 51-53.
- [10] Mercurial SCM – <http://mercurial.selenic.com/>
- [11] The Protégé Ontology Editor and Knowledge Acquisition System – <http://protege.stanford.edu/>
- [12] Horridge M., Drummond N., Goodwin J., The Manchester OWL syntax, *Second International*

*Workshop OWL: Experiences and Directions (OWLED 2006)*. 216 (2006).

[13] Grid-DL project page – <https://github.com/pospishniy/Grid-DL>

---



**Поспешный Александр Сергеевич**, аспирант и ассистент кафедры вычислительной техники факультета информатики и вычислительной техники НТУУ “КПИ”.

Научные интересы: Грид-системы, высокопроизводительные вычисления, семантические технологии, интеллектуальные системы, представление знаний, искусственный интеллект, системное программирование, встраиваемые системы, пользовательские интерфейсы.



**Стиренко Сергей Григорьевич**, кандидат технических наук, доцент кафедры вычислительной техники факультета информатики и вычислительной техники НТУУ “КПИ”.

Научные интересы: организация эффективных вычислительных процессов в гетерогенных высокопроизводительных средах; экстремальные параллельные вычисления для проблемно-ориентированных предметных областей; построение схем метаданных и онтологий для повышения качества интеллектуального поиска Грид ресурсов.