

УДК 519.6  
ББК В22

**Николай Васильевич Звягинцев<sup>1</sup>**,  
аспирант,  
Тверской государственный университет  
(170100, Россия, г. Тверь, ул. Желябова, 33)  
e-mail: n.zvyagintsev@gmail.com  
**Роман Николаевич Гордеев**,  
кандидат физико-математических наук,  
доцент кафедры информационных технологий,  
Тверской государственный университет  
(170100, Россия, г. Тверь, ул. Желябова, 33)  
e-mail: roman.gordeev@mail.ru

### Архитектура распределённой программной среды для теоретического исследования химических систем

Предложена архитектура распределённой программной среды для теоретического исследования химических систем. В соответствии со спецификой моделирования структуры химических соединений предлагается трёхзвенная архитектура ПО. Сформулированы требования к данному ПО – кроссплатформенность, модульность и наличие Web-интерфейсов пользователя, в качестве инструмента разработки рассматривается технология Java EE. Предполагается, что данная программная среда найдёт применение в области хемоинформатики и биоинформатики.

*Ключевые слова:* распределенные вычисления, программная среда, химические системы, интеллектуальные системы, интеллектуальный анализ данных.

**Nikolay Vasilyevich Zvyagintsev<sup>2</sup>**,  
Postgraduate Student,  
Tver State University  
(33 Zhelyabova st., Tver, Russia, 170100)  
**Roman Nikolaevich Gordeyev**,  
Candidate of Physics and Mathematics, Associate professor,  
Tver State University  
(33 Zhelyabova st., Tver, Russia, 170100)  
e-mail: roman.gordeev@mail.ru

### Architecture of Distributed Software Environment for Theoretical Studies of Chemical Systems

Architecture of distributed software environment for theoretical studies of chemical systems has been proposed. In accordance with the specific of modeling the structure of chemical compounds the proposed architecture is based on three-tier layer software systems. The requirements for this software are: cross-platform, modularity and availability of web-user interfaces. As a tool for developing the technology of Java EE has been chosen. It is assumed that the software environment will be used in the field of bioinformatics and chemoinformatics.

*Keywords:* distributed computing, software environment, chemical systems, intelligent systems, data mining.

**1. Введение.** В настоящее время теоретические методы активно используются в химии и биологии для проведения исследований структуры и свойств химических соединений (в том числе для исследования макромолекул: белков, ДНК и т. д.) [1–2]. За прошедшие десятилетия разработано много программ для проведения исследований методами квантовой химии, молекулярной

---

<sup>1</sup>Звягинцев Н. В. – основной автор, является организатором исследования, формулирует выводы и обобщает итоги реализации коллективного проекта.

<sup>2</sup>Zvyagintsev N.V. is the main author, research organizer, formulates conclusions and summarizes the results of the collective project.

механики и молекулярной динамики [2–7]. Основная особенность большинства подобных программ — консольный ввод/вывод, что существенно усложняет работу пользователя с данным ПО. Кроме того, формат входных/выходных данных большинства расчетных программ различный, поэтому определенную сложность представляет проведение комбинированных расчетов, когда исследование различных частей исследуемой системы требуется выполнить с использованием различных программных средств. С учётом того, что уже разработано большое количество расчетных программ, встает вопрос о интеграции их в единую инструментальную среду, позволяющую упростить работу пользователей с расчетными задачами без внесения изменений в уже созданные программы. Разрабатываемая инструментальная среда должна интегрировать программное обеспечение под наиболее распространенные операционные системы (ОС Windows, GNU Linux). Основной особенностью подобных расчётных задач является их ресурсозатратность. В зависимости от размера исследуемой химической системы, доступного вычислительного ресурса и выбранного приближения расчёт может занять от нескольких секунд до нескольких дней или недель. Поэтому для наибольшей эффективности программная среда должна иметь распределённую архитектуру, иметь возможность легко добавлять новые вычислительные ресурсы и распределять нагрузку между имеющимися вычислительными узлами.

**2. Проблематика.** Основная цель исследования состоит в том, чтобы разработать архитектуру ПО, управляющего расчётными задачами группы исследователей. ПО позволит распределять вычислительные ресурсы, синхронизировать работу, осуществлять обмен данными между задачами для комбинированных расчётных задач (например, квантовая химия + молекулярная динамика), хранить результаты в единой базе данных.

На основе анализа специфики решения расчётных задач сформулированы следующие требования к архитектуре разрабатываемой программной среды:

1. Модульность, изменение одной компоненты не требует внесения изменений в уже созданные компоненты (например, в компонент работы с базой данных). Исключение составляет переработка интерфейсов компонентов.
2. Возможность интеграции в инструментальную среду любой расчётной программы путём создания компоненты – адаптера, преобразующего интерфейс интегрируемой программы в соответствии с требованиями самой инструментальной среды.
3. Кроссплатформенность программного обеспечения.
4. Реализация базы данных для хранения результатов расчётов.
5. Поддержка языка сценариев для управления комбинированными исследованиями.
6. Простота добавления новых вычислительных узлов вне зависимости от типа операционной системы.
7. Управление вычислительными ресурсами и ходом выполнения расчётной программы.
8. Разграничение доступа к вычислительным ресурсам и результатам вычислений.
9. Отсутствие лицензионных ограничений на все компоненты разрабатываемой программной среды.

Для управления комплексными расчетами предложено использовать язык сценариев Java Script (используя библиотеку Rhino, входящую в состав Java SE), также сценарии на основе XML (*eXtensible Markup Language*).

Для работы пользователя с единой инструментальной средой предусмотрен web-интерфейс, при помощи которого можно управлять расчётом и осуществлять контроль за выполнением расчётной задачи.

Важной задачей является создание единого формата обмена данными между вычислительными программами. Предложен единый формат обмена данными, основанный на XML. Преобразование данных непосредственно в формат расчётной программы осуществляется самой средой. В качестве альтернативы единому формату обмена данными программная среда поддерживает традиционные форматы интегрируемых вычислительных программ.

Общая схема предложенной архитектуры программной среды представлена на рис. 1.

Клиент в данной архитектуре является тонким. В предлагаемой инструментальной среде клиентом может являться как Web-браузер, так и специальная программа, управляющая выполнением расчёта после соединения с сервером.

Под сервером предполагается программная платформа, состоящая из Java EE сервера приложений и сервера база данных, обеспечивающего хранение всех результатов работы инструментальной среды. На сервере приложений развёрнуты модули работы с базой данных, модули для обеспечения

web-интерфейса пользователя, а также модули, управляющие работой инструментальной среды. Схема работы сервера показана на рис. 2.

Поставщики, изображённые на рис. 1, – это компьютер или группа компьютеров, готовая предоставить вычислительный ресурс. Реализуется в виде клиентского приложения, настроенного на соответствующий сервер. При запуске поставщика на сервер передается информация о доступном вычислительном ресурсе и доступном на данном узле программном обеспечении.

Задача – запущенное из поставщика приложение для проведения расчёта (например, Games [3], Tinker [5]), управляемое поставщиком и передающее ему данные о ходе расчёта.

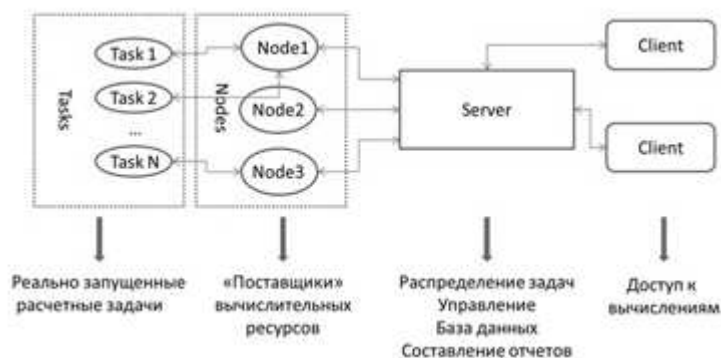


Рис. 1. Общая архитектура программной среды для теоретического исследования химических систем

На рис. 2 показана схема работы сервера. На сервере приложений расположены компоненты, которые реализуют основной функционал среды в соответствии с разработанными интерфейсами. В виде овала изображен главный компонент (ГК), который управляет работой. Это постоянная сущность, которая управляет бизнес-процессом исследования. Стрелками на рис. 2 показаны связи ГК с остальными компонентами. ГК осуществляет следующие функции:

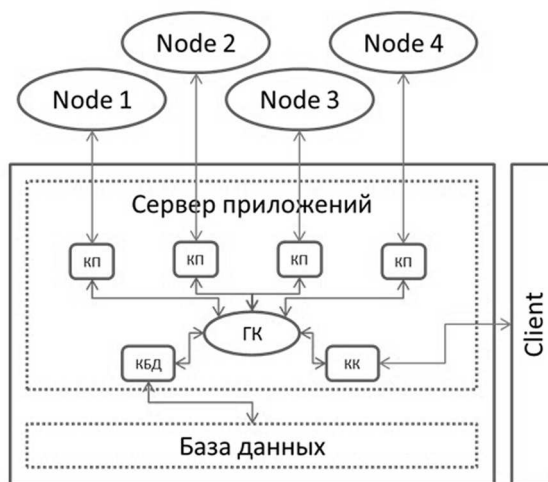


Рис. 2. Схема работы сервера

1. Регистрация новых поставщиков.
2. Отправка и получение данных при проведении расчётов.
3. Сохранение получаемой информации с поставщиков в БД через модуль работы с БД.
4. Формирование отчётов о ходе и результатах вычислений, обработка запросов клиентов.

5. Реализация бизнес-процесса при проведении серий расчётов или комбинированных расчётов.

При запуске клиентской программы на поставщике автоматически на сервере приложений появляются соответствующие компоненты поставщика (КП), через которые ГК управляет работой поставщика.

Работа с базой данных осуществляется через компонент управления БД (КБД). Стрелкой на рис. 2 обозначена связь КБД с самой базой данных. Этот компонент существует на сервере приложений постоянно и связан с остальными компонентами через ГК. В случае, если сервер связан с несколькими БД, на сервере будет существовать несколько КБД.

Компонент, обрабатывающий запросы от клиента (КК), получает доступ к МБД и поставщикам тоже через ГК. Этот компонент появляется при соединении клиента с сервером и после отключения пользователя исчезает.

Поставщик является клиентским Java-приложением, которое управляет вычислительным ресурсом на том компьютере, где он запущен. Схема работы поставщика представлена на рис. 3.

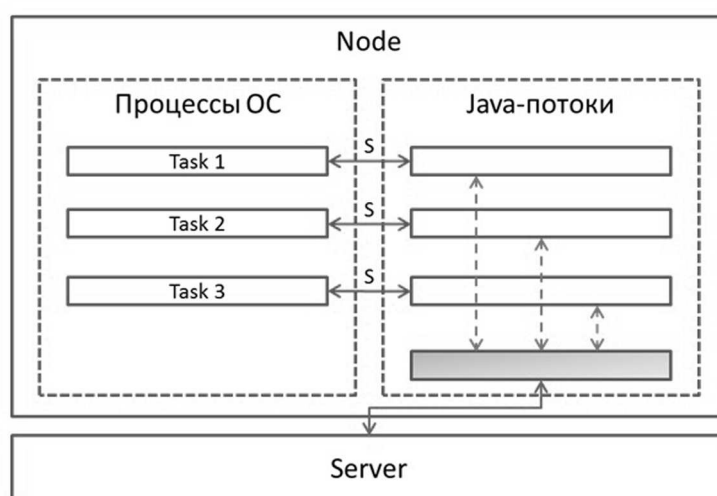


Рис. 3. Схема работы поставщика

При запуске поставщика на сервере автоматически создается КП, который регистрирует его в ГК. Поставщик реализован в виде нескольких Java-потоков: главного (который связан с сервером, на рисунке обозначен заштрихованным прямоугольником) и нескольких второстепенных, контролирующих выполнение задач. Связи между потоками обозначены пунктирной линией.

В случае, когда, если ГК принимает решение запустить на компьютере поставщика расчетную задачу, поставщик создает отдельный Java-поток, который запускает расчетную программу как процесс операционной системы (если запускается расчетная нативная программа), осуществляет контроль за её выполнением и обмен данными. Стрелками с литерой S на рис. 3 обозначен обмен данными между процессом и Java-потоком.

В качестве вспомогательного инструмента, отвечающего за представление информации о химическом строении соединений и работы с форматами данных расчетных программ, предполагается использовать фреймворк CDK [8]. Данный фреймворк представляет собой библиотеку Java-классов, позволяющих работать с 3D-структурой химических соединений, проводить анализ информации о структуре и свойствах химических соединений, а также работать с форматами данных, принятыми в хемоинформатике.

Важной задачей является визуализация результатов работы расчетных программ: 3D-структуры химических соединений, а также табличных данных, графиков и диаграмм. Для визуализации 3D-структуры предложено воспользоваться инструментом Jmol [10]. Данный инструмент можно встраивать как в графический интерфейс пользователя, так и встраивать в веб-страницу в качестве апплета. Для построения таблиц и визуализации графиков и диаграмм предложено использовать BIRT (Business Intelligence and Reporting Tools) [11].

В качестве сервера приложений предложен JBoss. В качестве системы управления базами данных предполагается использовать PostgreSQL.

**3. Заключение.** Предложена архитектура программной среды, позволяющая автоматизировать работу научных групп, занимающихся теоретическими исследованиями в области хемоинформатики и биоинформатики. Главное преимущество данной архитектуры — модульность, что значительно расширяет возможности и сферу применения данного ПО. Предлагаемая архитектура позволит легко интегрировать в данную среду любое расчётное ПО путём разработки компоненты-адаптера, который обеспечит интерфейс расчётной программы с самой программной средой. Предлагаемое ПО упростит работу пользователя с вычислительными программами и обеспечит использование единого формата данных. Также вычислительная среда обеспечит распределение задач между поставщиками вычислительных ресурсов.

### Список литературы

1. Jensen F. Introduction to Computational Chemistry / F. Jensen. 2006. Wiley.
2. Цирельсон В. Квантовая химия. Молекулы, молекулярные системы и твёрдые тела. М.: Бином, 2010. 496 с.
3. Официальный Web-сайт программы Gaussian. URL: <http://www.gaussian.com/> (дата обращения: 14.02.2014).
4. Официальный Web-сайт программы Gamess. URL: <http://www.msg.ameslab.gov/gamess/> (дата обращения: 10.02.2014).
5. Официальный Web-сайт программы Tinker. URL: <http://dasher.wustl.edu/tinker/> (дата обращения: 10.02.2014).
6. Официальный Web-сайт программы Gromacs. URL: <http://www.gromacs.org/> (дата обращения: 25.01.2014).
7. M. Valiev NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations / M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. A. de Jong // Comput. Phys. Commun. 2010. Vol. 181. № 1477.
8. The Chemistry Development Kit. URL: <http://sourceforge.net/projects/cdk/> (дата обращения: 25.01.2014).
9. Официальный Web-сайт программы Jmol. URL: <http://www.jmol.org> (дата обращения: 14.02.2014).
10. Статья из Википедии. URL: [http://en.wikipedia.org/wiki/BIRT\\_Project](http://en.wikipedia.org/wiki/BIRT_Project) (дата обращения: 10.02.2014).

### References

1. Jensen F. Introduction to Computational Chemistry / F. Jensen 2006. Wiley.
2. Tsirel'son V. Kvantovaya Khimiya. Molekuly, molekulyarnye sistemy i tverdye tela / Tsirel'son V. M.: Binom, 2010. 469 s.
3. Oficial'nyi sait Web-sait programmy Gaussian. URL: <http://www.gaussian.com/> (data obrashcheniya: 14.02.2014).
4. Oficial'nyi sait Web-sait programmy URL: <http://www.msg.ameslab.gov/gamess/> (data obrashcheniya: 10.02.2014).
5. Oficial'nyi sait Web-sait programmy Tinker. URL: <http://dasher.wustl.edu/tinker/> (data obrashcheniya: 10.02.2014).
6. Oficial'nyi sait Web-sait programmy Gromacs. URL: <http://www.gromacs.org/> (data obrashcheniya: 25.01.2014).
7. M. Valiev NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations / M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma,

H. J. J. van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, W. A. de Jong // Comput. Phys. Commun. 2010. Vol. 181. № 1477.

8. The Chemistry Development Kit. URL: <http://sourceforge.net/projects/cdk/> (data obrashcheniya: 25.01.2014).

9. Ofitsial'nyi sait Web-sait programmy Jmol. URL: <http://www.jmol.org> (data obrashcheniya: 14.02.2014).

10. Stat'ya iz Vikepedii. URL: [http://en.wikipedia.org/wiki/BIRT\\_Project](http://en.wikipedia.org/wiki/BIRT_Project) (data obrashcheniya: 10.02.2014).

*Статья поступила в редакцию 10.04.2014*