

КЕРУВАННЯ ПОТОКАМИ НАУКОВИХ РОБІТ В ГРІД-СИСТЕМАХ

Зроблено короткий огляд проблематики використання систем керування потоками наукових робіт для підтримки наукових досліджень. Приводиться аналіз схожості і відмінностей опису Scientific WorkFlow і Business WorkFlow в інформаційних системах. Аналіз проводиться з урахуванням тих змін, що відбулись за останні роки у зв'язку зі збільшенням можливостей, масштабу та глибини при проведенні наукових досліджень. Акцентовано увагу на системі Кеплер (Kepler), специфічній системі керування науковими процесами, що нині активно розвивається разом із численними науковими проектами керування даними. В роботі відмічені деякі ключові особливості системи Кеплер і її основи – системи Птолемея II (Ptolemeu II).

Вступ

Розвиток і вдосконалення технологій, методів і засобів інформаційної підтримки наукових досліджень є одним з найбільш перспективних напрямів інтенсифікації наукового процесу і підвищення його ефективності. Тенденції розвитку інформаційних систем у цій області направлені до створення інтегрованих проблемно-орієнтованих середовищ, призначених для підтримки всього циклу досліджень у вибраній прикладній області: постановка задачі → вибір методів і алгоритмів її вирішення → здобуття структурованих нових знань і представлення їх для загального використання.

Сучасний науковий процес, що розглядається із структурних позицій, характеризується наступними чинниками:

- тривалість процесу, що досягає декількох років при вирішенні масштабних задач;
- великі наукові колективи, що беруть участь у дослідженнях і розробках, зокрема, міжнародні і такі, що знаходяться на віддаленій відстані;
- величезна кількість інформації у літературі по вибраному завданню;
- використання складних комп'ютеризованих приладів і інших технічних засобів для проведення наукових експериментів;
- широке застосування сучасних засобів обчислювальної техніки, програмного забезпечення, Інтернету та Грід-інфра-

структур для отримання інформації і її обробки.

Тому на передній план в інформатизації наукових досліджень виходить проблема автоматизації керування самим науковим процесом (Scientific WorkFlow), яка має стати ядром інтегрованої е-Інфраструктури наукових досліджень, орієнтованих на вирішення наукових проблем [1].

Парадигма WorkFlow не нова, еволюція інформаційних систем у бізнесі вже досить тривалий час проводиться в цьому напрямі. Відомі й широко застосовуються системи рівня управління підприємством (ERP), побудовані на основі парадигми Business WorkFlow, при якій в інформаційній системі паралельно з протіканням у реальності (а іноді – випереджаючи) виконуються відповідні бізнес-процеси, а сама система є засобом їх контролю і управління.

Рівень використання Scientific WorkFlow в наукових інформаційних системах істотно відстає. Водночас аналіз рішень, втілений в бізнесових системах, дає потужний імпульс для розвитку інформаційних систем підтримки наукових досліджень.

Для подальшого прогресу ця нова, керована інформацією та даними наука, отримує великі інвестиції для розвитку телекомунікацій та технічних засобів інфраструктури відомої як «Грід-інфраструктура». Однак учені кінець кінцем зацікавлені в інструментах, які приво-

дять потужність розподілених баз даних і інших обчислювальних та розподілених в мережі ресурсів до робочого столу, і дозволяють їм зручно сполучати і керувати їх власними науковими технологічними процесами. Маються на увазі процеси роботи в мережі, які зазвичай використовують «канали аналізу даних», або порівнюють спостережувані і прогнозовані дані. Також може використовуватись широкий діапазон компонентів, наприклад, для запитів до баз даних, для перетворення даних і вилучення інформації з даних, для виконання моделювання на високопродуктивних комп'ютерах, і т. п.

В ідеалі, учений має мати змогу:

- підключити майже будь-який науковий ресурс даних і обчислювальне обслуговування в науковому технологічному процесі;
- детально вивчити і виразно представити собі дані «на льоту» (безпосередньо в процесі обчислень);
- змінювати параметри у будь-який час за потребою, та після цього повторно перезапускати виконання тільки необхідних кроків;
- отримувати достатньо даних на завершальних етапах процесу;
- отримувати метадані про самі кроки перебігу наукового технологічного процесу;
- отримувати дані не тільки результатів експерименту, а й допоміжні, що пояснюють результати і роблять їх відновлюваними для інших учених.

Тому система управління науковими потоками робіт стає науковим проблемно-орієнтованим оточенням, налагодженим до все більш і більш поширюваної і орієнтованої на обслуговування інфраструктури Грід.

Проте, перед тим, як це грандіозне бачення зможе стати дійсністю, доводиться звертати увагу на багато істотних факторів. Наприклад, поточне програмне забезпечення Грід все ще дуже складне, щоб його використовувати для пересічного вченого, і швидка зміна версій та стандартів, що постійно розвиваються, вимагають, щоб ці деталі були прихованими від користувача системи управління науковими

потоками робіт. Здається, мережеві послуги забезпечують просту основу для вільного зв'язування та розподілення системи, але основні мережеві службові стандарти як наприклад WSDL тільки забезпечують прості рішення для простих проблем, при більш серйозних проблемах, як, наприклад, оркестрування мережевими службами, передача посередниками (від одного сервісу безпосередньо до іншого, обходячи передачу до ядра керування потоками), і транзакційна семантика потоків, заснованих на сервісах, лишають тему розвитку та майбутнього стандартів мережесервісів актуальною.

У даній роботі проведено попередній аналіз проблематики використання систем керування потоками наукових робіт для підтримки наукових досліджень. Аналіз проводиться з урахуванням тих змін, що відбулись за останні роки у зв'язку зі збільшенням можливостей, масштабу та глибини при проведенні наукових досліджень. Приводиться аналіз схожості та відмінностей опису Scientific Workflow і Business Workflow в інформаційних системах.

Крім цього представлено специфічні особливості системи Кеплер та її основи – системи Птолемея II. Головна перевага системи лежить у парадигмі моделювання і проектування під назвою «актор-орієнтоване моделювання», що, виявляється, дуже важливо, щоб мати справу з результатами складної архітектури проектів наукових процесів.

1. Наукові процеси

Згідно визначенню, приведеному в глосарії WFMC (Workflow Management Coalition), Business Workflow – це автоматизація (повністю або частково) бізнес-процесу, при якій документи, інформація або завдання передаються для виконання необхідних дій від одного учасника до іншого відповідно до набору процедурних правил; набір інструментальних засобів для аналізу, реорганізації і автоматизації інформаційно-насичених завдань і функцій.

Істотними, на наш погляд, у цьому визначенні є наступні моменти:

1) автоматизоване виконання певного процесу у вибраній прикладній області, при якому окремі етапи й операції процесу виконуються автоматично інформаційною системою, а деякі – людиною (користувачем системи);

2) інформаційна система, заснована на парадигмі *Workflow*, є засобом:

- керування процесом, що автоматизується,
- регламентації виконання процесу,
- контролю ходу виконання процесу;

3) взаємодія користувачів при виконанні процесу проводиться, в першу чергу, за допомогою інформаційної системи, в якій крок за кроком відображаються результати етапів виконаного процесу.

Парадигма *Workflow*, на наш погляд, істотним чином змінює місце інформаційної системи – вона стає інформаційним ядром виконання процесів, що керує, у прикладній області. У інформаційній системі створюється так званий «інформаційний конвеєр», утворення якого зіставно з появою конвеєра в промисловості [2, 3].

У інформаційних системах підтримки наукових досліджень визначення *Scientific Workflow* досить схоже, наприклад: засоби автоматизації і контролю процесів у наукових проблемно-орієнтованих середовищах [4], або схеми аналітичних кроків, які включають, наприклад, доступ до бази даних, формування запитів, аналіз даних і інше, а також інтенсивні обчислювальні роботи на кластерних комп'ютерах [5].

Можна сказати, що *Scientific Workflow* є ланкою для керування в інтегрованому проблемно-орієнтованому середовищі автоматизованого наукового дослідження, основним же її призначенням є:

- забезпечення регламенту виконання дослідження;
- взаємодія різних учасників дослідження в процесі вирішення завдання;
- взаємодія науковців з обладнанням, на якому виконуються дослідження і експерименти;
- підбір компонентів програмних засобів, що дозволяють автоматизувати окремі кроки дослідження і забезпечення взаємодії з ними;

- фіксація поточних результатів дослідження на кожному його кроці і результатів дослідження в цілому.

Очевидно, що в реальності *Scientific Workflow* є спектром інформаційних моделей процесів наукової діяльності, що розрізняються за галузями науки, задачах, що вирішуються і класах досліджень, що проводяться.

Спостерігається зростаючий інтерес до інформатизації наукових процесів, що можна побачити з переліку подій, наприклад, the *Scientific Data Management Workshop* [6], the *e-Science Workflow Services Workshop* [7], the *e-Science Grid Environments Workshop* [8], the *Virtual Observatory Service Composition Workshop* [9], the *e-Science LINK-Up Workshop on Workflow Interoperability and Semantic Extensions* [10], і наостанок *Global Grid Forum* [11], тільки, щоб не переховувати більше. Наукові процеси також грають важливу роль у ряді великих дослідницьких проектів, що продовжуються, мають справу з науковим керуванням даними, у тому числі NSF/ITR (*GriPhyN*, *GEON*, *LEAD*, *SCEC*, *SEEK*, ...), NIH (*BIRN*), DOE (*Sci-DAC/SDM*, *GTL*), і подібні зусилля, засновані UK *e-Science initiative* (*myGrid*, *DiscoveryNet*, та інші). Наприклад, проект *SEEK* розвиває *Analysis and Modeling System (AMS)*, яка дозволяє екологам проектувати і виконувати наукові технологічні процеси [12]. *AMS* компонент технологічного процесу використовує *Semantic Mediation System (SMS)*, щоб полегшити проект технологічного процесу і перегляд даних через семантичне друкування [13]. Тому *SEEK* – хороший приклад керованого суспільством проекту в потребі такої системи, яка дозволяє користувачам «без швів» звертатися до джерел даних і послуг, і залучати їх в технологічних процесах, що повторно використовуються. Дійсно *SEEK* – один з головних проектів, що сприяв ініціативі Кеплера і систем керування науковими процесами, який обговорюються далі.

Наукові процеси часто експонують специфічні «штрихи», наприклад, вони можуть використовувати великі масиви даних, інтенсивні обчислення, проводити велику кількість аналізу даних, виконувати багато візуалізації і т. п. Процеси, що зга-

дуються далі, наприклад, експонують різні особливості: орієнтація на сервіси та аналіз даних, які призначені для взаємодії з користувачем та реінжиніринг, високопродуктивні обчислення. В залежності від призначення, для окремих груп користувачів може бути потреба приховати або навпаки зробити наголос на специфічних аспектах і технічних можливостях наукових процесів. Наприклад, “інженер Грід”, можливо, був би зацікавлений у таких низькорівневих аспектах процесу, як, наприклад, передача даних і керування завданням з віддаленого терміналу. Мати компоненти (або актори) технологічного процесу, які діють на цьому рівні, буде вигідно “інженеру Грід”.

З іншого боку, система керування науковим процесом повинна сховати ці ж аспекти від вчених-аналітиків (скажімо, еколога, що вивчає багатство різновидів та їх продуктивність). Система Кеплер націлена на підтримку великої кількості різних видів процесів, від низькорівневих “каналів” процесу, що цікавлять “інженерів Грід”, до аналітичних процесів відкриття знання для вчених, і рівня концептуального дизайну процесів, які, можливо, стали б здійсненими тільки в результаті подальших кроків налаштування [14].

2. Scientific WorkFlow – потоки даних чи потоки робіт

Характеристики і вимоги наукових процесів частково перекликаються з бізнес-процесами. Дійсно, термін «наукові процеси», здається, вказує на дуже близьке відношення з останнім, і тільки детальніше порівняння виявляє перелік істотних відзнак. Історично, бізнес-процеси мають коріння, що ведуть до систем автоматизації офісу 70-их і 80-их років, та придбали інерцію в 90-их під різними іменами, у тому числі «моделювання бізнес-процесів» і «розробка бізнес-процесів»; дивіться, наприклад, [15–17].

Сьогодні ми бачимо деякий вплив стандартів бізнес-процесів у сфері Веб-послуг, особливо стандарти для хореографії Веб-послуг. Наприклад, Мова Виконання Бізнес-Процесів для Веб-Послуг (BPEL4WS) [18], поглинання двох більш

ранніх стандартів, IBM WSFL і Microsoft XLANG привернули деяку увагу.

Аналізуючи основні принципи проектування і підходи моделювання бізнес-процесів, фокусування на шаблонах потоків керування подіями стає очевидним, тоді як керування потоками даних залишається в тіні.

Системи керування науковими процесами, з іншого боку, прагнуть мати моделі виконання, які значно більше орієнтовані на потоки даних. Це вірно, наприклад, для академічних систем, у тому числі Кеплер, Taverna [19], і Triana [20], і для комерційних систем, як, наприклад, Inforsense's DiscoveryNet або Scitegic's Pipeline-Pilot.

Різницю між орієнтацією на потоки даних і потоки робіт можна також спостерігати в їх формальному описі. Наприклад, візуалізація бізнес-процесів часто нагадує календарні графіки процесу, діаграми зміни станів, або UML діаграми діяльності, всі з яких роблять наголос на керування подіями по зрівнянню з потоками даних. Формальний аналіз технологічних процесів зазвичай залучає вивчення їх зразків керування потоками [21], і часто проводиться, використовуючи мережі Петрі.

Довгий час в інформаційних системах підтримки наукових досліджень акценти розставлялися таким чином, що бізнес-процеси (Business WorkFlow) орієнтовані в основному на керування складними потоками робіт, а в наукових процесах (Scientific WorkFlow) набагато більшу увагу слід було приділяти потокам даних [22]. Проте, із зростанням гетерогенності середовища, складності і різноманітності програмного забезпечення, що залучається до наукового процесу, пріоритети змінилися.

По-перше, людські ресурси є обов'язковою частиною системи моделювання наукових процесів, і взаємодіють з нею набагато тісніше, ніж в бізнес-процесах. Вчені можуть вносити корективи практично у будь-який момент і в будь-який стан процесу (змінюючи параметри, додаючи додаткові параметри, змінюючи оточення виконання процесу). Таке втручання особливе часто зустрічається на початкових етапах роботи обчислювальних процесів.

По-друге, на сьогодні обчислювальне середовище необхідно розглядати як

гетерогенне. Обчислювальне середовище в загальному випадку містить в собі засоби різної складності, організовані в систему. Спеціалізовані вимірювальні системи, суперкомп'ютери і робочі станції можуть бути об'єднані в Грід-мережі. Керування таким середовищем є важливим, але непростим завданням, оскільки гетерогенність гарантує додаткову складність завданням підтримки і керування обчислювальними процесами.

По-третє, із-за підвищеної складності програмного забезпечення, що використовується, вірогідність виникнення помилок і виняткових ситуацій підвищується, що вимагає додаткової обробки таких ситуацій. Для забезпечення цілісності і коректності даних, що використовується в процесі, у разі виникнення виняткових ситуацій необхідно провести так звану «семантичну відміну» змін, які були виконані в ході тих або інших операцій процесу. Така необхідність зв'язана з тим, що дуже часто звичайна відміна транзакцій в сховищах даних просто неможлива із-за великої тривалості часу виконання деяких стадій обчислювального процесу або великого об'єму змінних даних. Таким чином, необхідно вводити функціонал фіксації станів процесу, відмінний від таких функцій сховищ даних, як транзакція сесії (transaction) і точка збереження (savepoint). Крім того, необхідний функціонал, що забезпечує відміну змін процесу до зафіксованих раніше станів процесу.

На підставі приведених фактів можна стверджувати, що моделювання потоків робіт (керування) у системах інформатизації наукових процесів по-перше, має не меншу важливість і складність, ніж у бізнес-процесах, по-друге, повинна грати, як мінімум, роль керування в складних моделях наукових процесів.

Основним завданням системи керування потоками наукових робіт (Scientific Workflow System) є забезпечення необхідної послідовності дій, які пов'язані з етапами діяльності наукового персоналу і етапами взаємодії з ресурсами процесу. Процес характеризується набором параметрів: значенням споживаного ресурсу, часом життя та іншими. Саме поняття «процес» визначається з погляду завдань, що треба виконати, зокрема: роботи з приладами, складними технологічними системами, документами і персоналом.

3. Реалізація Scientific workflow у системі Kepler

Kepler – це програмний інструмент, який застосовується для аналізу та моделювання наукових даних. Kepler спрощує зусилля, необхідні для створення моделей, що виконуються за допомогою візуального представлення цих процесів. Ці представлення, або «наукові робочі процеси», відображують потік даних між дискретними компонентами аналізу та моделювання (рис. 1).

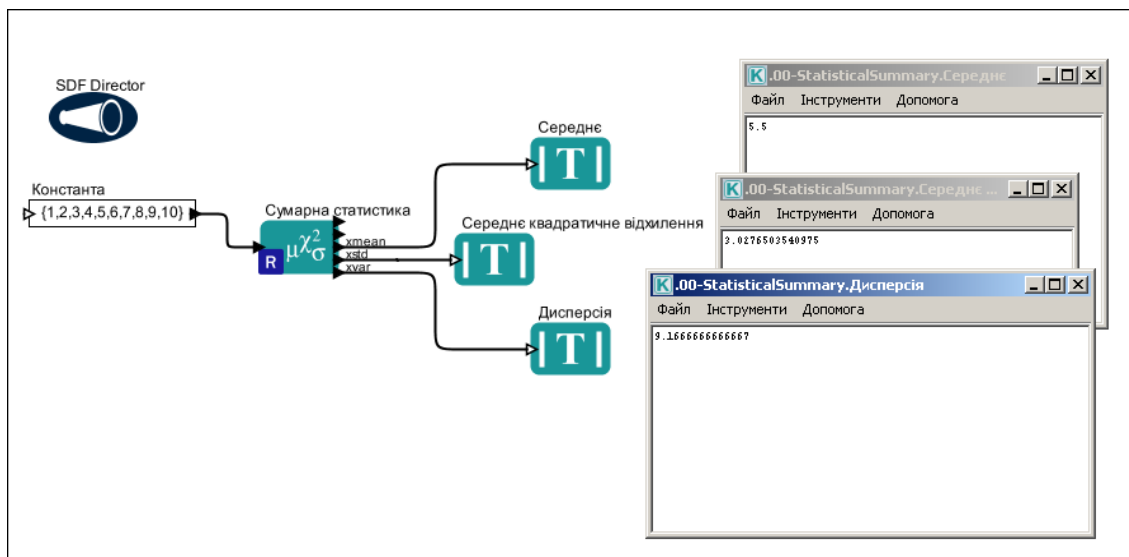


Рис. 1. Простий науковий робочий процес, розроблений у Kepler

Kepler дозволяє ученим створювати свої власні готові до виконання наукові технологічні процеси досить простим перетягуванням компонент до діаграми процесу й сполучати компоненти для побудови конкретного потоку даних, таким чином, створюючи візуальну модель аналітичної частини своїх досліджень. Kepler представляє загальний процес візуально, таким чином, що легко зрозуміти, які дані перетікають від одного компонента до іншого. Отриманий процес може бути збережений у текстовому форматі, відправлений по електронній пошті колегам, та/або опублікований для обміну з колегами по всьому світу.

Користувачі Kepler навіть з невеликим досвідом в інформатиці можуть створювати процеси зі стандартних компонент, або змінювати існуючі процеси для задоволення своїх потреб. Навіть досвідчені користувачі знайдуть, що Kepler має багато переваг, особливо коли мова йде про подання комплексних програм та аналізу у формі зрозумілій і зручній для обміну.

Kepler включає розподілені обчислювальні технології, які дозволяють вченим обмінюватися своїми даними і процесами з іншими вченими та використовувати процеси обробки даних та аналітичні процеси з іншими людьми по всьому світу. Kepler також забезпечує доступ до комплексу інформаційних сховищ, обчислювальних ресурсів, і бібліотек процесів, що постійно розширюється й географічно розподілений (наприклад, екологічні дані з польових станцій, зразки даних з музейних колекцій, дані про Землю й т. п.).

Kepler заснований на системі візуального моделювання Ptolemy II [23] з відкритим вихідним кодом, створює єдине робоче середовище для вчених. На сьогодні це зручна програма, яка дозволяє вченим створювати свої власні наукові робочі процеси без необхідності інтегрувати кілька різних програмних застосувань або заручатися допомогою програмістів.

Багато готових до використання компонент включено в стандартну поставку Kepler, у тому числі загальні математичні, статистичні компоненти і компоненти обробки сигналів та введення даних, обро-

бки та відображення. Статистичний аналіз на основі системи R або MATLAB, обробка зображень або ГІС-функціональність доступна через прямі посилання на ці зовнішні пакети. Ви також можете створити новий компонент або «обгорнути» існуючі компоненти з інших програм (наприклад, програм на мові C) для використання в рамках Kepler.

Kepler є крос-платформним програмним забезпеченням з відкритими початковими кодами, яке може працювати на Windows, Macintosh або Linux-платформі.

Компоненти Kepler: Наукові робочі процеси складаються з компонент, що налаштовуються (директори, актори і параметри), а також зв'язків і портів, які забезпечують взаємодію між компонентами (рис. 2).

Директор і актори: Kepler використовує метафору директор/актор для візуального представлення різних компонентів процесу. Директор контролює (або ж керує) виконання процесу так само, як режисер стежить за знімальною групою. Актори приймають на виконання інструкції з боку директора. Іншими словами, актори визначають, *що* буде відбуватись у процесі, водночас як директор указує, *коли* це буде відбуватись.

Кожен процес повинен мати директора, який контролює виконання процесу з використанням певної моделі обчислень. Кожну модель обчислень у Kepler представляє свій директор. Наприклад, процес виконання може бути синхронним з обробкою одного компонента в один момент часу, в заздалегідь розрахованій послідовності (SDF директор). Крім того, компоненти процесу можуть виконуватись паралельно, з одним або більшою кількістю компонентів, що працюють одночасно (що може бути у випадку з PN-директором). Невеликий набір найчастіше використовуваних директорів поставляється з інсталяцією Kepler, але ще більше їх є в базовому програмному забезпеченні Ptolemy II, яке можна отримати за необхідності.

Kepler надає великий набір акторів для створення і редагування наукових робочих процесів. Актори можуть бути

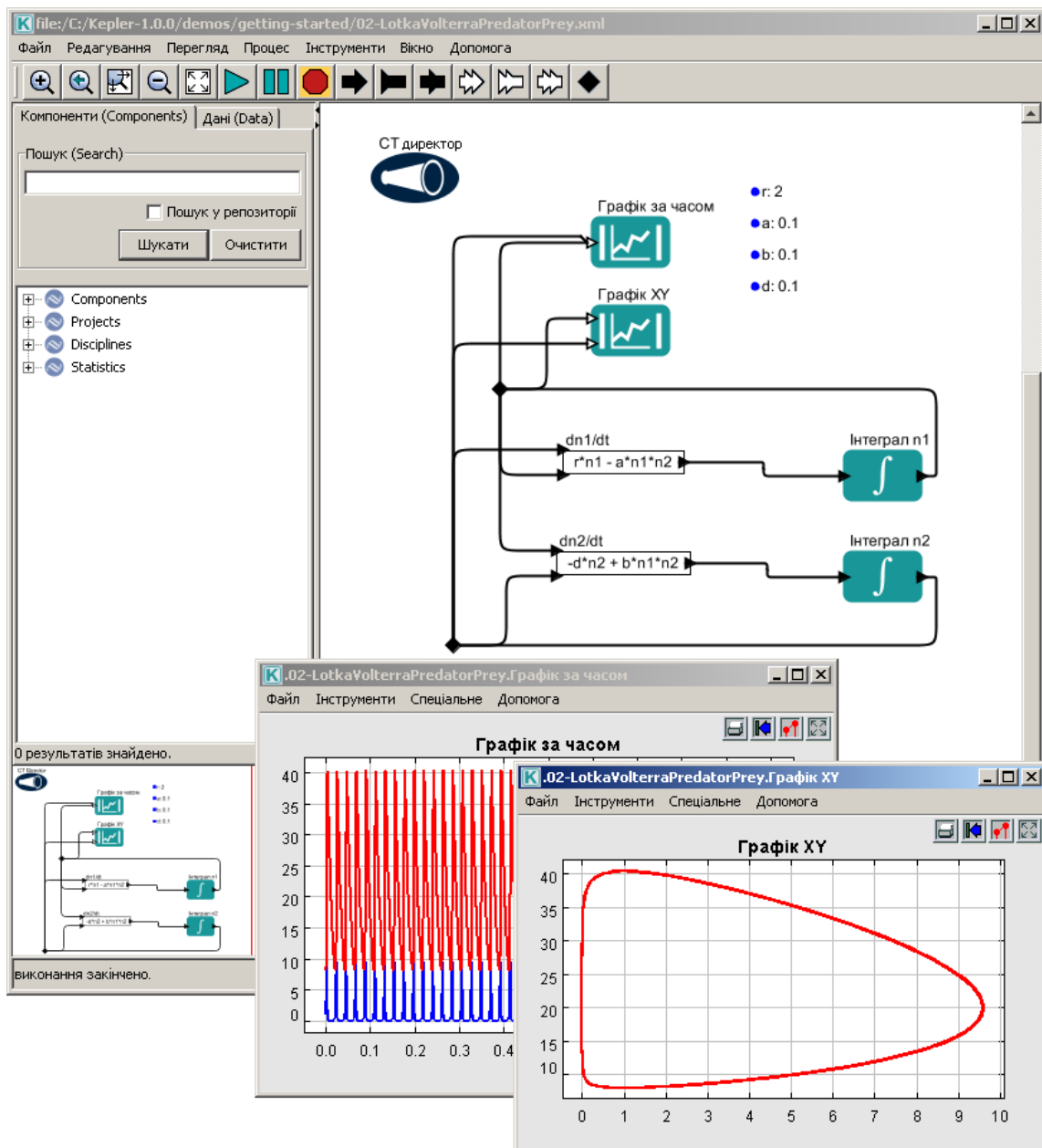


Рис. 2. Головне вікно Kerler з виділеними деякими з основних компонентів процесу. Вікно в правому нижньому кутку – це вікно, створене робочим процесом, щоб вивести результат у графічному вигляді

додані в Kerler для ексклюзивного використання та/або можуть бути зроблені доступними для інших.

Порти. Кожен актор у процесі може містити один або кілька портів, що використовуються для споживання або створення інформації і спілкування з іншими учасниками процесу. Актори пов'язуються в процес за допомогою своїх портів. Зв'язок, що представляє собою потік даних між одним портом актора та іншим, називається каналом.

Процеси можуть також використовувати зовнішні порти і параметри портів.

Зв'язки дозволяють користувачам «розгалужувати» потік даних. Розгалужені дані можуть бути передані до декількох місць у процесі. Наприклад, учений, можливо, побажає направити вихід актора процесу до іншого актора процесу для подальшої обробки, а також до актора відображення, щоб відобразити дані в цій конкретній точці. Розміщуючи зв'язок на каналі виведення даних, користувач може

направити інформацію одночасно в обидва місця.

Параметри – це значення, закріплені за процесом, окремими директорами або акторами. Параметри можуть бути налаштовані. Наприклад, актор Integrator (інтегратор) має параметр під назвою InitialState (початковий стан), який має встановити перше значення функції інтегрування. Параметри акторів імітаційної моделі можуть бути налаштовані на контроль деяких аспектів моделювання (наприклад, початкові значення моделі). Параметри директора керують кількістю ітерацій процесу й відповідними критеріями для кожної ітерації.

Інтерфейс користувача Kepler. Наукові робочі процеси будуються і редагуються в інтерфейсі користувача Kepler, який пропонує легку навігацію і інтерфейс «перетягування». Основні розділи вікна застосування Kepler (рис. 2) складаються з наступного:

- рядок меню – забезпечує доступ до всіх функцій Kepler;
- панель інструментів – забезпечує доступ до функцій Kepler, що найбільш часто використовуються;
- область компонент і доступу до даних – складається із закладки Компоненти, а також із закладки Дані. Обидві закладки містять функцію пошуку та відображують бібліотеку доступних компонентів або результати пошуку;
- діаграма процесу (workflow canvas) – надає робочий простір для відображення та створення процесів;
- область навігації – відображає повний процес.

Основні операції в Kepler. До них відносяться:

- відкриття існуючих наукових робочих процесів;
- запуск існуючого наукового робочого процесу;
- редагування існуючого наукового робочого процесу;
- пошук доступних даних;
- пошук доступних компонент обробки даних;

- створення наукового робочого процесу.

Більш детальна інформація про операції з прикладами їх виконання, а також деякі методики для редагування, проектування і створення власних наукових процесів міститься у звіті Інституту програмних систем НАН України за темою «Розробка загальносистемних проектних рішень з питань побудови технологій та систем підтримки науково-дослідної діяльності».

Висновки

Розглянуті у даній статті проблеми стосуються важливих особливостей опису потоків наукових робіт, що по-перше, відрізняють ці потоки від потоків робіт бізнес-процесів, по-друге, за своєю актуальністю з часом зрівнялись з потребами потоків бізнес-процесів, та потребують повторного формулювання, дослідження та порівняння.

При подальшому розгляді існуючих систем керування потоками наукових робіт для їх практичного використання у підтримці наукових досліджень розглянуті питання надають змогу висунути функціональні вимоги до таких систем та виконати зважений вибір системи для впровадження.

Очевидно, що рівень відкритості системи, можливість вибору функціоналу, та навіть можливість модифікації механізмів підтримки потоків робіт буде важливим аргументом при виборі такої системи для практичного використання. У такому випадку, наприклад, по-перше, з'являється можливість перевірки зроблених припущень або розроблених методів на практиці на базі тієї ж самої системи, що використовується для інформаційної підтримки; по-друге, полегшується процес аналізу і усунення виявлених проблем.

Ми також провели огляд проблем керування науковими процесами, мотивованих реальними прикладами. Спектр того, до чого може звернутися науковий процес широкий і включає наукові технологічні процеси відкриття, технологічні процеси, які автоматизують ручні процедури або спеціалізовані інструменти реінженірингу і

інтенсивної обробки даних, а також інтенсивні обчислювальні процеси. Наукова підтримка процесу потрібна практично для всіх інформаційно-орієнтованих наукових дисциплін, у тому числі біоінформатики, хіміоінформатики, екоінформатики, геоінформатики, фізики, і т.п.

Система Кеплер забезпечує підтримку процесів, що базуються на Веб-послугах і Грід. Початковий код Кеплера вільно доступний [24].

Унікальна особливість Кеплера успадкована від основної системи Птолемей II: актор-орієнтований підхід моделювання. Цей підхід полегшує моделювання і проектування складних систем і тому забезпечує також багатообіцяючий напрям для невідкладних проблем, як, наприклад, компонування і оркестровка Веб-послуг. Шлях поліморфізму даних і динамічний поліморфізм підтримує актор-орієнтований підхід, який “концентрує” взаємодію компонент в окремому об’єкті директора, може також висвітлити інші зусилля, необхідні для створення компонентної архітектури, що легко повторно використувати, як наприклад ССА [25].

1. Твердохліб Є.М., Перконос П.І. Побудова інтегрованої е-інфраструктури підтримки наукових досліджень в грід-середовищі // Проблеми програмування (матеріали конф. УкрПрог-2012). – 2012. – № 2–3. – С. 143–149
2. Твердохліб Є.М. Інформаційна система в управлінні підприємством. Міжнародна науково-практична конференція «Розробка систем програмного забезпечення (software system engineering): Виклик часу та роль у інформаційному суспільстві», Тезиси докладів, Київ, 2005.
3. Твердохліб Є.М., Перконос П.І. Автоматизация управления: информационный конвейер в действии. КПД (Компьютерное проектирование и технический документооборот), Киев, 08.2009.
4. On Scientific Workflow. Dr. Jinjun Chen, Swinburne University of Technology, Australia; Prof. dr. ir. W.M.P. van der Aalst, Eindhoven University of Technology, The Netherlands.
5. Grid, Scientific and Business Workflow Management Systems. Dr. Jinjun Chen, CITR – Centre for Information Technology Research Swinburne University of Technology, Australia, August 2007.
6. Scientific Data Management Framework Workshop, Argonne National Labs, August 2003. <http://sdm.lbl.gov/~arie/sdm/SDM.Framework.k.wshp.htm>.
7. e-Science Workflow Services Workshop, e-Science Institute, Edinburgh, Scotland, December 2003. <http://www.nesc.ac.uk/esi/events/303/index.html>.
8. e-Science Grid Environments Workshop, e-Science Institute, Edinburgh, Scotland, May 2004. <http://www.nesc.ac.uk/esi/events/>.
9. GRIST Workshop on Service Composition for Data Exploration in the Virtual Observatory, California Institute of Technology, July 2004. <http://grist.caltech.edu/sc4devo/>.
10. LINK-Up Workshop on Scientific Workflows, San Diego Supercomputer Center, October 2004. <http://kbis.sdsc.edu/events/link-up-11-04/>.
11. Workflow in Grid Systems Workshop, GGF10, Berlin, Germany, March 2004. <http://www.extreme.indiana.edu/groc/Workflow-call.html>.
12. Michener W.K., Beach J.H., Jones M.B., Ludäscher B., Pennington D.D., Pereira R.S., Rajasekar A. and Schildhauer M. A Knowledge Environment for the Biodiversity and Ecological Sciences // Journal of Intelligent Information Systems, 2004.
13. Bowers S. and Ludäscher B. An Ontology Driven Framework for Data Transformation in Scientific Workflows. In International Workshop on Data Integration in the Life Sciences (DILS), LNCS 2994, Leipzig, Germany, March 2004.
14. Bowers S. and Ludäscher B. Actor-Oriented Design of Scientific Workflows. submitted for publication, 2005.
15. Alonso G. and Mohan C. Workflow Management Systems: The Next Generation of Distributed Processing Tools // In S. Jajodia and L. Kerschberg, editors, Advanced Transaction Models and Architectures. 1997.
16. W. van der Aalst and K. van Hee. Workflow Management: Models, Methods, and Systems (Cooperative Information Systems) // MIT Press, 2002.
17. M. zur Muehlen. Workflow-based Process Controlling. Logos Verlag, Berlin, 2004.

18. *Curbera F., Goland Y., Klein J., Leyman F., Roller D., Thatte S. and Weerawarana S.* Business Process Execution Language for Web Services (BPEL4WS), Version 1.0, 2002.
<http://www.ibm.com/developerworks/library/ws-bpel/>.
19. *The Taverna Project.* <http://taverna.sf.net/>.
20. *The Triana Project.*
<http://www.trianacode.org/>.
21. *Kierpuszewski B.* Expressiveness and Suitability of Languages for Control Flow Modelling in Workflows // PhD thesis, Queensland University of Technology, 2002.
22. *Нестеренко А.К., Сысоев Т.М., Бездушный А.Н., Бездушный А.А., Яроцук И.О.* «Интеграция научных информационных систем при помощи механизма рабочих процессов».
23. *Ptolemy II project and system.* Department of EECS, UC Berkeley, 2004.
<http://ptolemy.eecs.berkeley.edu/ptolemyII/>.
24. *Kepler: A System for Scientific Workflows.*
<http://kepler-project.org>.
25. *Armstrong R., Gannon D., Geist A., Keahey K., Kohn S., McInnes L., Parker S. and Smolinski B.* Toward a Common Component Architecture for High-Performance Scientific Computing // In 8th IEEE Intl. Symposium on High Performance Distributed Computation, August 1999.

Про авторів:

Твердохліб Євген Миколайович,
кандидат технічних наук,
старший науковий співробітник відділу,

Перконос Петро Іванович,
старший науковий співробітник відділу,

Лой Анатолій Іванович,
Провідний консультант
Департаменту бізнес-аналізу
ТОВ «БМС Консалтінг».

Місце роботи авторів:

Інститут програмних систем
НАН України
03680, Київ,
Проспект Академіка Глушкова, 40,
Тел. +38(044) 526 6408,
E-mail: Eugene@nas.gov.ua ,
Perkonos@nas.gov.ua

Одержано 03.01.2013