5. ОСВІТЯНСЬКІ ПРОБЛЕМИ ВИЩОЇ ШКОЛИ

Науковий вісник НЛТУ України
Scientific Bulletin of UNFU
http://nv.nltu.edu.ua
https://doi.org/10.15421/40270638
Article received 03.09.2017 p.
Article accepted 28.09.2017 p.
УДК 004.89

ISSN 1994-7836 (print)
ISSN 2519-2477 (online)

@ ✉ Correspondence author
N. E. Kunanets
nek.lviv@gmail.com

Н. Е. Кунанець[1], В. С. Ленько[1], В. В. Пасічник[1], Ю. М. Щербина[2]

[1] *Національний університет "Львівська політехніка", м. Львів, Україна*
[2] *Львівський національний університет ім. Івана Франка, м. Львів, Україна*

ПЕРСОНАЛЬНІ БАЗИ ДАНИХ ТА ЗНАНЬ ВІРТУАЛЬНИХ ДОСЛІДНИЦЬКИХ СПІЛЬНОТ

Досліджено процеси набуття та управління персональними даними і знаннями, а також комунікації в межах віртуальних дослідницьких спільнот. Розглянуто поняття віртуального наукового колективу та особливості захисту інформації у налагодженні процесів комунікації в ньому. Висвітлено проблеми в організації ефективної комунікації між учасниками віртуальних дослідницьких спільнот, що породжені відсутністю цілісного технологічного рішення, яке забезпечило б надійність, приватність, швидкодію та структурованість потоків інформації, даних і знань. Запропоновано принципово новий підхід до проектування платформ комунікації, що ґрунтується на ідеях децентралізації та крипто-безпеки. Проаналізовано методологічні відмінності між поняттями "інформація", "дані" та "знання". Висвітлено сучасний підхід до здійснення міркування в онтологічній моделі подання знань, що ґрунтується на використанні апарату теорії типів та асистента доведення теорем Coq. Структуровано подано відношення між логікою та теорією типів, а також наведено спосіб подання основних елементів онтології формальною мовою системи Coq. Розглянуто приклад здійснення міркування з використанням мови тактик Ltac над фрагментом онтології, що описує певні відношення з предметної області функціонування віртуальних дослідницьких спільнот. Вказано шляхи та інструменти здійснення подальшого дослідження.

Ключові слова: е-наука; комунікація; блокчейн; онтологія; теорія типів; логічне міркування.

Вступ. Стрімкий розвиток інформаційних технологій та обчислювальних ресурсів помітно впливає практично на всі сфери повсякденного суспільного життя, в якому чільне місце належить праці дослідника. Ретельне планування етапів наукового дослідження передбачає вивчення та аналіз значних за об'ємом та структурною складністю даних, що надходять з різноманітних джерел. У процесі їх опрацювання в дослідника формується знаннєвий потенціал, який за своєю суттю є особистісним відображенням інформації про конкретний об'єкт, методи, засоби та способи його дослідження та навколишній світ загалом.

Плин часу породжує проблему, яку зазвичай називають проблемою зберігання знань: якщо набуті знання активно не використовуються та не актуалізуються, то вони піддаються забуттю або ж частково втрачають свою первинну цінність. Оскільки набуття знань належить до категорії енергозатратних та часомістких видів діяльності, науковець-дослідник зацікавлений у подоланні означених вище проблем. Персональні бази знань дослідника, які зафіксовані у комп'ютерній формі, значною мірою дають змогу уникнути втрати знань, а також технічно полегшують процеси їх актуалізації.

Мета дослідження полягає в комплексному аналізі поточного стану та перспектив технологічної оптимізації функціонування віртуальних наукових колективів. Дослідження проблем ефективної та захищеної комунікації, набуття та управління знаннєвим потенціалом, розгляд основних понять та технологій, висвітлення сучасних тенденцій покликані створити міцне аналітичне підґрунтя для проектування надійної та ефективної платформи комунікації віртуальних дослідницьких спільнот. Конкретизація зазначеної мети передбачає застосування методів аналізу та синтезу до предметних областей, що відображають специфіку функціонування віртуальних наукових колективів, технологічні інструменти його забезпечення, ефективні моделі подання знань та міркування, понятійний апарат.

Стан дослідження проблеми. Сучасна комунікація віртуальних дослідницьких спільнот характеризується

Інформація про авторів:

Кунанець Наталія Едуардівна, д-р наук із соціальних комунікацій, професор кафедри інформаційних систем та мереж. **Email:** nek.lviv@gmail.com

Ленько Василь Степанович, аспірант кафедри інформаційних систем та мереж. **Email:** vs.lenko@gmail.com

Пасічник Володимир Володимирович, д-р техн. наук, професор кафедри інформаційних систем та мереж. **Email:** vpasichnyk@gmail.com.

Щербина Юрій Миколайович, канд. фіз.-мат. наук, професор кафедри дискретного аналізу та інтелектуальних систем. **Email:** yshcherbyna@yahoo.com.

Цитування за ДСТУ: Кунанець Н. Е., Ленько В. С., Пасічник В. В., Щербина Ю. М. Персональні бази даних та знань віртуальних дослідницьких спільнот. Науковий вісник НЛТУ України. 2017. Вип. 27(6). С. 185–191.

Citation APA: Kunanets, N. E., Lenko, V. S., Pasichnyk, V. V., & Shcherbyna, Yu. M. (2017). Personal Data and Knowledge Bases of Virtual Research Communities. *Scientific Bulletin of UNFU, 27*(6), 185–191. https://doi.org/10.15421/40270638

використанням розрізнених інструментів, що здебільшого орієнтовані на широкий загал та не враховують особливі потреби конкретних груп. Формування методологічної та технологічної основи для проектування систем комунікації віртуальних наукових колективів є актуальною задачею. Деякі напрацювання в цьому напрямку сприяють її вирішенню. Зокрема, в роботі (Veretennikova et al., 2015) подано означення віртуального наукового колективу, його роль, структуру та особливості функціонування. Технологічні компоненти, що здатні забезпечити ефективне та надійне середовище для комунікації, описано в роботах (Romano & Schmid, 2017; Protocol Labs, 2016; Li et al., 2017). Понятійний апарат передбачає оперування концептами "дані", "інформація", "знання" та їхніми відношеннями (Wang, 2015; Liew, 2007). Онтологічний інжиніринг (Feilmayr, & Wöß, 2016; Noy, & McGuinness, 2001) та здійснення міркування з використанням теорії типів (Dapoigny, & Barlatier, 2008; Hafsi, Dapoigny & Bolon, 2015) у межах асистента доведення теорем Coq (Cog, 2017) пропонують сучасний підхід до управління базами знань.

Комунікація у віртуальних наукових колективах. Сучасні наукові дослідження, з огляду на свою складність, виходять за межі класичних інституцій та потребують експертизи дослідників-науковців, які працюють у різних фізичних локаціях. Інноваційну інформаційно-технологічну платформу для організації таких досліджень у Великобританії та в Європі трактують як e-Science, а в США як Cyberinfrastructure і перебуває вона на стадії швидкого формування та методологічного становлення.

Ефективно функціонуюча платформа вибудовується на базі телекомунікаційної інфраструктури, комп'ютерних та інформаційних технологій, сховищ та просторів даних. Одним зі сутнісних складників успішної реалізації проектів на базі інфраструктури е-науки (e-Science) та досягнення поставлених у проектах цілей є управління інформацією та ефективна комунікація між членами віртуальної проектної команди, що, зазвичай, передбачає формування віртуального наукового колективу (Veretennikova et al., 2015). Віртуальний науковий колектив – це ситуативно сформована група з представників різних наукових інституцій, які шляхом системного використання сучасних інформаційних та комунікаційних технологій віртуально об'єднуються для проведення спільних наукових досліджень та реалізації комплексних наукових проектів. Інтеграція відбувається по вертикалі з метою об'єднання ключових компетентностей окремих учасників віртуального територіально розподіленого колективу та організації роботи для вирішення задач як єдиної організаційно-цілісної науково-дослідницької одиниці.

Такі наукові колективи можуть об'єднувати географічно або інституційно розподілених дослідників, представників різних галузей знань, які проводять системне дослідження тієї чи іншої проблеми. Ефективність таких досліджень істотно підвищується в разі налагодження процесів комфортної співпраці та якісних комунікацій між членами наукового колективу, що складається з представників різних регіонів та установ.

Невідкладним завданням, яке потрібно вирішувати для успішної реалізації проекту, є забезпечення ефективного використання інформаційно-технологічної платформи і комунікаційних зв'язків між учасниками віртуального наукового колективу, який, зазвичай, формується із вузькопрофільних високо-фахових дослідників. Такий підхід забезпечує налагодження ефективної комунікації та співпраці у реальному масштабі часу із збереженням з генерованої інформації у формі даних та знань, виконання функцій управління спільною діяльністю учасників в рамках великомасштабних, розподілених мультидисциплінарних проектів.

Взаємодія у віртуальному науковому колективі повинна підкріплюватися архітектурними програмно-алгоритмічними та інформаційно-технологічними рішеннями, які сприяли б отриманню нових наукових результатів. Учасники наукових проектів, які реалізуються на платформі е-науки, стикаються з низкою проблем, які пов'язані з відсутністю якісних і верифікованих методів управління віртуальними проектними групами та зручних у використанні інструментальних програмно-алгоритмічних комплексів.

Сучасне інформаційне суспільство генерує потребу активного поступального розвитку окремої інноваційної галузі, якою є електронна наука як системно з інтегрований комплекс сучасних комп'ютерних інформаційних та комунікаційних технологій, які забезпечують реалізацію основних функцій і завдань науки. Такий підхід реалізує відкритий доступ до консолідованих інформаційних ресурсів, сформованих на використання технологій хмарних обчислень (Cloud computing) (Dhamdhere, 2014) та великих за обсягом даних (Big Data) (Kaisler et al., 2013), що містять результати наукових досліджень, які проводяться віртуальними проектними групами та науковими колективами.

Співпраця між науковцями, зазвичай, полягає в обговоренні ідей, координації спільних дій, обміні напрацюваннями, джерелами знань та отриманими результатами. Оскільки інформація передається здебільшого в електронному вигляді через мережу Інтернет, то постає необхідність в забезпеченні приватності, швидкісної передачі, безпеки, фіксації історії обміну повідомленнями в процесі комунікації.

Як свідчать результати аналізу, комунікація у віртуальних наукових колективах здійснюється переважно за допомогою програмних продуктів, що реалізовані за принципом клієнт-серверної архітектури (Gmail, Skype, Viber, веб-платформи). Реалізація комунікаційних процесів покладається, зазвичай, на серверну компоненту. Недоліком цього підходу є те, що користувачі системи повністю покладаються на коректність функціонування віддаленого вузла. У разі його недоступності, несанкціонованого доступу, технологічної застарілості, цензорування віртуальний науковий колектив ризикує втратити конфіденційність повідомлень, не досягти ефективної комунікації. Нова форма організації дослідницької роботи потребує розроблення та запровадження інноваційних підходів і до комунікаційних процесів. Для удосконалення та підвищення ефективності інформаційно-технологічного супроводу науково-дослідної роботи віртуальних наукових колективів дослідники кафедри інформаційних систем та мереж Національного університету "Львівська політехніка" запропонували інноваційні підходи.

Для забезпечення режиму захищеності та збереження статусу конфіденційності отриманих результатів наукових досліджень та комунікативних процесів у віртуальних групах та наукових колективах планується вико-

ристання технологій блок-чейн (Blockchain) (Romano & Schmid, 2017). Такий технологічний вибір дає змогу ефективно реалізовувати процедури, пов'язані із захистом прав інтелектуальної власності авторів на результати, отримані в рамках виконання проекту, або конкретного наукового експерименту.

Відбувається бурхливий розвиток розподілених систем та технологій, що покликані вилучити централізовану серверну ланку з процесів обміну, збереження та доступу до інформації. Становлення мереж рівноправних вузлів peer-to-peer (P2P), вдосконалення криптогеш функцій, поява технології Blockchain та розподіленої файлової системи IPFS (Protocol Labs, 2016) підготували необхідне підґрунтя для проектування систем децентралізованої комунікації. Проведений авторами пошук продуктів для ефективної комунікації класу P2P зафіксував тільки проект системи обміну повідомленнями Echo (Bitshares Munich IVS, 2016). Згідно з описом, система шифрує повідомлення (Marshall, 2016) та зберігає їх в структурі Graphene Blockchain. Для того, щоб отримати та прочитати повідомлення користувача, децентралізований клієнт повідомлень повинен мати доступ до ланцюжка блоків. Обмін файловими та потоковими ресурсами здійснюється шляхом запису IPFS гешкоду в Blockchain полі "memo".

Технологію Blockchain було описано у 2008 р. (Nakamoto, 2008), а її перша практична реалізація з'явилася на рік пізніше в основі криптовалюти Bitcoin. Blockchain є розподіленою базою даних, яка підтримує постійно зростаючий перелік записів (блоків) та є криптографічно захищеною від підробки. Кожен блок містить групу нових "перевірених" транзакцій, хеш попереднього блоку, власний хеш та службову інформацію. У 2014 р. з'явилося поняття Blockchain 2.0 (Swan, 2015), яке позначає нові напрямки застосування розподіленої бази даних. Одним з таких застосувань є "розумні контракти" (Omohundro, 2014), що надають можливість програмованого обміну ціннісною інформацією без впливових посередників. Корпорації, своєю чергою, активно впроваджують "приватні" ланцюжки блоків (Li et al., 2017), що дають змогу надавати доступ визначеним вузлам.

Одним з недоліків Blockchain є постійне зростання обсягу розподіленої бази даних. Кожен учасник ланцюжка володіє повноцінною або частковою копією даних, яка регулярно синхронізується. Якщо спробувати зберігати в Blockchain великі дані чи мультимедійний контент, то в швидкій перспективі обсяг ланцюжка може спричинити для користувачів проблему, яка полягатиме у відсутності достатніх ресурсів для зберігання бази даних. Окрім цього, істотно зросте тривалість синхронізації стану бази між користувачами. Рішення цієї проблеми полягає у збереженні в Blockchain гешу об'єкта замість його контенту (рис. 1).

Розподілена файлова система IPFS початково проектувалася як система контролю версії об'ємних наукових файлів, проте згодом переросла у контентно-адресований, гіпермедійний P2P протокол. IPFS мають такі особливості:

- Кожному файлу та усім його блокам присвоюють унікальні крипто-геші;
- IPFS видаляє дублікати в мережі та зберігає історію версій кожного файлу;
- Вузли мережі зберігають тільки важливий для них контент та певний індекс;
- Пошук файлу здійснюється за унікальним гешом або через сервіс IPNS;
- Обмін контентом здійснюється в мережі рівноправних вузлів P2P.

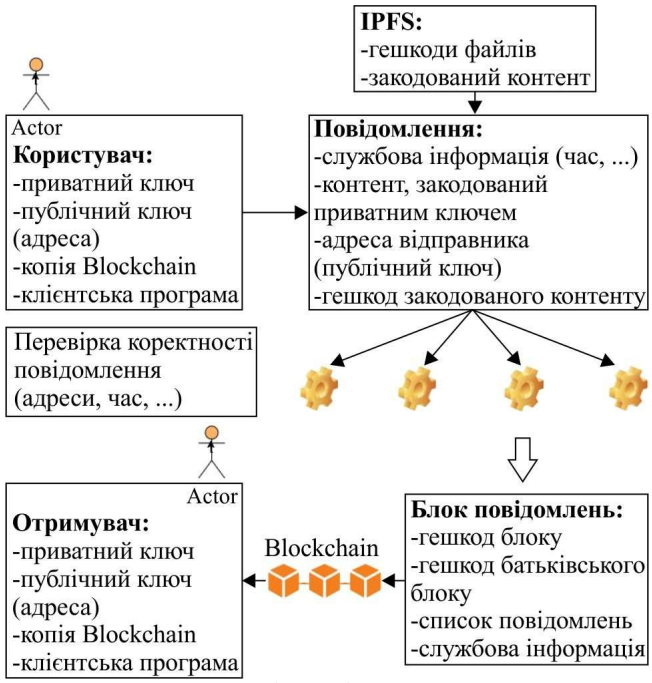

Рис. 1. Схема процесу обміну повідомленнями з використанням технологій Blockchain та IPFS

Дані, інформація, знання. Персональні бази даних та знань дослідників, науковців, фахівців виконують специфічну роль порівняно із традиційними базами даних та знань, які слугують для зберігання та опрацювання емпіричних даних, масивів документів та відомостей у їх висхідному поданні. Вони, зокрема, дають змогу накопичувати та опрацьовувати дані, які є важливими для фахівця-дослідника. Такі бази даних виконують роль персоналізованого інформаційного ресурсу, що створюється та впорядковується відповідно до індивідуальних потреб дослідника. Технологічно персональна база даних може реалізовуватися у вигляді хмарного сховища, на основі реляційної чи постреляційної моделі даних, або ж поєднувати ряд відомих сучасних підходів. Вибір тієї чи іншої архітектури зазвичай здійснюється з урахуванням функціональних та безпекових вимог її власника.

В процесі проектування персональних баз даних та знань потрібно обов'язково враховувати методологічні відмінності, які існують між поняттями "інформація", "дані" та "знання". У роботах (Wang, 2015; Liew, 2007) подано означення, функції, відмінності і зв'язки, які існують між цими поняттями. З формально гносеологічних позицій, дані трактуються як абстрактне подання кількісних властивостей об'єктів реального світу або абстрактних об'єктів, відповідно до конкретних кількісних шкал (Wang, 2015). Дані по відношенню до кількісної властивості X за шкалою міри σ, $D(X,\sigma)$, отримуються шляхом застосування перетворення $\Gamma_\sigma : X \to \Re$, яке породжує дійсне число $I_X^\sigma.R_X^\sigma$ в одиницях $[\sigma]$, де I_X^σ є цілою частиною, а R_X^σ є залишком (Wang, 2015):

$$D(X,\sigma) = \Gamma_\sigma(X) = \frac{X}{\sigma} = I_X^\sigma.R_X^\sigma[\sigma],$$

$$\sigma \in \Re,\ I_X^\sigma \in Z,\ 0 \le R_X^\sigma \le \sigma,\ \{I_X^\sigma = \mathrm{mod}_\sigma(X), R_X^\sigma = \mathrm{mod}_\sigma(X)\}.$$

Більш інтуїтивне означення трактує дані як збережені символи чи покази сигналу, що представляють певні факти чи події (Shen, 2014). Варто зазначити, що окрім кількісних, тут враховуються ще й і якісні властивості об'єктів. Обидва означення передбачають, що дані є "сирою" інформацією, яка отримана в результаті абстракції зовнішніх сутностей та їх відношень в певні величини: $f_D : O \to Q$, де O – об'єкт, Q – величина, D – дані. Основна роль даних полягає в записі дій чи ситуацій з метою фіксування істинної картини чи реальної події.

Інформація – це повідомлення, що має відповідний зміст, висновок або вхідні дані для прийняття рішення чи здійснення дії (Liew, 2007). Інформація є інтерпретацією даних та передбачає наявність семантики: $f_I : D \to S$, де D – дані, S – семантика, I – інформація. У n-символьній системі інформація I визначається за формулою ентропії:

$$H = -\sum_i p_i \log_2(p_i),$$

де p_i – ймовірність появи значення i в повідомленні. Окрім цього, в роботі (Wang, 2015) подано модель та поняття "сучасної інформації", яка дає змогу розширити певні твердження, що присутні в теорії інформації К. Шеннона. Основна роль концепту "інформація" полягає у забезпеченні відомостями процесів прийняття рішень щодо розв'язання актуальних задач чи реалізації потенційних можливостей.

Теорія інформації за К. Шенноном ґрунтується на оцінці кількості інформації у випадкових величинах, яку інтуїтивно можна описати як "визначення середньої кількості бітів, що є необхідною для кодування випадкової величини" (Shen, 2014). Проте існує багато джерел інформації, які не є випадковими величинами, зокрема книги, структура ДНК тощо. Ідея А. М. Колмогорова, який є автором алгоритмічної теорії інформації, полягала в переході від визначення кількості інформації у випадкових величинах до визначення кількості інформації в індивідуальних об'єктах (Shen, 2014). Нехай $F : p \mapsto F(p)$ – це функція (мова програмування, інтерпретатор), яка обчислює результат роботи програми p. Тоді складність (кількість інформації) слова x при інтерпретаторі F – це мінімальна довжина l програми p, яка необхідна, щоб це слово створити:

$$K_F(x) = \min\{l(p) \mid F(p) = x\}.$$

Варто зазначити, що функція $x \mapsto K_F(x)$ в загальному випадку є нерозв'язною, тобто не існує ефективного алгоритму, який для $\forall x$ обчислює $K_F(x)$.

Знання – це внутрішнє особистісне відображення об'єктивної дійсності, що формується суб'єктом у процесі її пізнання. Знання є набутою та осмисленою інформацією, що породжена інтелектуальною розумовою діяльністю та втілена у певному концепті-понятті: $f_K : I \to C$, де I – інформація, C – формальне поняття, K – знання. Формальна модель знань "об'єкт-атрибут-властивість" задається кортежем $\langle O, A, R \rangle$, де O – множина об'єктів з унікальними іменами, A – множина атрибутів, які відносяться до того чи іншого об'єкта, R – множина відношень між об'єктами та атрибутами (Wang, 2015). Формулюючи узагальнене подання трьох означених вище понять "дані", "інформація", "знання" можна зафіксувати такі трактування: "дані" – це матеріалізована на певному носієві інформація, а "знання" – це суб'єктизована "інформація", яка є особистісним відображенням, що формується у суб'єкта внаслідок пізнання ним зовнішнього світу (рис. 2).

Рис. 2. Співвідношення понять "знання", "інформація" та "дані" (Liew, 2007)

Робота дослідників у складі віртуальних наукових колективів передбачає одночасове оперування як даними, так і інформацією та знаннями, що отримані учасниками колективу чи надходять з різноманітних різнотипових інформаційних джерел. Дані можуть подавати результати натурних експериментів, покази давачів, статистики певного дослідження тощо. Інформація здебільшого подається у формі наукових статей, матеріалів доповідей і дискусій, монографій, посібників і т. ін. Знання подаються через призму внутрішніх особистісних відображень об'єктивної дійсності учасниками віртуального дослідницького колективу. Реалізація інформаційно-технологічної підтримки пошукової діяльності дослідника та оперування знаннями потребує їх явного подання.

Подання персональних знань дослідника. У теорії подання знань та міркування є кілька класів моделей, найпопулярнішими серед яких є логічні моделі, семантичні мережі, фрейми та продукційні правила. Кожен із наведених вище класів моделей характеризується експресивністю, тобто здатністю представити довільне за природою знання, розвиненим формально-математичним апаратом для здійснення міркування, що полягає у виведенні нових несуперечливих знань на основі висхідних знань, а також зручністю у використанні кінцевим користувачем. Широке різноманіття пропонованих моделей пояснюється наявністю груп явних переваг, які надає та чи інша модель за певними критеріями. Водночас, проектування великих та гіпервеликих баз знань потребує вибору гнучкої та верифікованої моделі подання знань, яка б різнопланово та ефективно задовольнила потреби користувача. На думку широкого фахового загалу, однією з найприйнятніших для цього є саме онтологічна модель.

Онтологія – це формальна, явна специфікація спільної концептуалізації, яка характеризується високою семантичною виразністю (Feilmayr, & Wöß, 2016). Узагальнено онтологія подається кортежем $O = \langle C, R, F \rangle$, де: C – скінченна множина понять-"концептів" предметної області; $R : C \to C$ – скінченна множина відношень між поняттями; F – скінченна множина функцій інтерпретації (обмеження, аксіоми) (Lytvyn, 2013). Особливістю онтологій є те, що вони поєднали переваги цілого спектра моделей подання знань в узагальненій структурі, що уможливлює її використання як цілісного, уніфікованого формального апарату. Поняття та відношення подаються у вигляді семантичної мережі, структура поняття відповідає фреймовій моделі, обмеження запису-

ються як продукційні правила, а формальна логіка використовується для аксіоматизації предметної області.

Однією з потужних додатних особливостей онтологій є уможливлення поширення загального розуміння структури інформації, багаторазового повторного використання знань про предметну область, явного подання гіпотез предметної області, розділення предметних і операційних знань. Спектр застосування тієї чи іншої онтології визначається ступенем загальності предметної області, яку вона описує. І як наслідок, базова онтологія містить таксономію понять загального призначення. На противагу, орієнтовані на конкретні завдання онтології описують зазвичай тільки ті поняття, які відіграють важливу роль при виконанні певного завдання. Предметно орієнтовані онтології мають ширше практичне застосування та ґрунтуються на експертних знаннях у певній предметній області.

Методи та засоби проектування онтологій розлого подано в роботах (Feilmayr, & Wöß, 2016; Noy, & McGuinness, 2001). Процес їх формування, зазвичай, зводиться до виконання таких кроків:

1. Визначення предметної області та обсягу онтології;
2. Аналіз наявних онтологій на предмет їх повторного використання;
3. Створення переліку важливих "концептів" (понять) в онтології;
4. Означення класів та їхньої ієрархії;
5. Означення властивостей класів та слотів;
6. Означення аспектів (*facets*) слотів;
7. Створення екземплярів класів.

Кроки 3-4 полягають у здійсненні концептуалізації предметної області з метою створення таксономії важливих понять. Кроки 5-6 деталізують властивості понять, щоб забезпечити можливість здійснення логічного міркування.

Важливою властивістю баз знань та онтологій зокрема є несуперечність збережених фактів. Незабезпеченість цієї властивості призводить до можливості логічного виведення абсурдних тверджень, які є джерелом потенційних помилок. Несуперечність онтологічних знань забезпечується шляхом створення їхнього подання в межах певної формальної системи та дедуктивним обчисленням істинності певного твердження. При виборі відповідної формальної системи необхідно враховувати її ключові характеристики, а саме: експресивність мови, алгоритмічна розв'язність та обчислювальна складність розв'язання основних логічних проблем. Значною популярністю серед розробників онтологій користується технологія OWL 2 DL, що заснована на базі формальної системи описової логіки. Оскільки вона є підмножиною логіки предикатів, то її виразність є обмеженою. Актуальним напрямком сучасних досліджень є спроби застосування логік вищих порядків та теорії типів до міркування в онтологіях (Dapoigny, & Barlatier, 2008; Hafsi, Dapoigny & Bolon, 2015).

Основою формальних логік вищих порядків є організація логічних понять – висловлювань, індивідів, пропозиційних функцій – в ієрархію типів таким чином, що сутності можуть оперувати тільки тими поняттями, які розташовані нижче в ієрархії. Таким чином логіки вищих порядків дають змогу у природний та зручний спосіб явно описувати зв'язок між універсалами та індивідами без необхідності введення додаткових спеціалізованих конструкцій (Hafsi, Dapoigny & Bolon, 2015). Хоча логіка предикатів володіє хорошими мета-властивостями компактності та Ловенгейма-Сколема, її обмежена експресивність є перешкодою для втілення багаторівневого міркування.

Відкриття ізоморфізму Каррі-Говарда-Ламбека дало змогу встановити зв'язки між формальними системами теорії доведень, теорії типів та теорії категорій. Таким чином логічні висловлювання можуть бути записаними за допомогою елементів теорії типів (табл. 1). Більше того встановлено, що процес логічного доведення може бути реалізований в межах просто типізованої системи λ-числення. Це спричинило появу нових концепцій "висловлювання як типи", "доведення як програми" та "спрощення доведень, як оцінювання програми". Програмним втіленням цих концепцій є асистенти доведення теорем Coq (Cog, 2017), Lean (Lean, 2017), Nuprl (Cornell University, 2017) та інші, які в напівавтоматизованому режимі дають змогу обчислити несуперечність певних висловлювань задекларованим положенням.

Табл. 1. Деякі співвідношення між логікою висловлювань та теорією типів (Homotopy type theory, 2013)

Логіка висловлювань	Теорія типів	Опис
висловлювання	A	тип A
доведення	$a : A$	індивід a типу A
$\bot$, $\top$	0, 1	нульовий, одиничний тип
$A \vee B$	$A + B$	тип кодобутку
$A \wedge B$	$A \times B$	тип добутку
$A \Rightarrow B$	$A \to B$	функційний тип
$\neg A$	$A \to 0$	функційний тип

У фундаментальних роботах відомих вчених (Dapoigny, & Barlatier, 2008; Hafsi, Dapoigny & Bolon, 2015) послідовно викладено метод подання онтологій з використанням теорії типів у межах системи "асистент доведення теорем" Coq. У зазначеній системі реалізована формальна мова з високою експресивністю, засобами напівавтоматичного доведення теорем, механізмом наслідування та гарантованим закінченням обчислень. Опис онтологій у теорії типів передбачає наявність конструкцій для подання відношень, понять, екземплярів, атрибутів, ієрархічних та мереологічних зв'язків (табл. 2).

Табл. 2. Подання елементів онтології у системі Coq

Елемент онтології	Конструкція Coq
Поняття	Class C: Type.
Екземпляр	Instance X: C.
Атрибут	Class C: Type:= { attr: Prop; }.
Відношення	Parameter Relation: $C \to C \to$ Prop.
Ієрархічне наслідування	Coercion D1: SubClass_G >→ C.
Мереологічне відношення "частина-ціле"	Definition Part_of(x y: C):= Relation x y. Axiom A1: Reflexive Part_of. Axiom A2: Asymmetric Part_of. Axiom A3: Transitive Part_of.
квантори $\exists, \forall$	exists X: C, forall X: C

Приклад 1. Розглянемо частину предметної онтології $O = \langle C, R, F \rangle$, що відображає особливості комунікації дослідників у межах віртуального наукового колективу. Нехай

$C = \{Віртуальна\_Наукова\_Мережа, Повідомлення, Учасник\}$;

$R = \{Власник\_повідомлення : Повідомлення \to Учасник \to \Pr op,$

$Існує\_контакт : Учасник \to Учасник \to \Pr op\}$;

F = {Повідомлення _ Належить _ До _ Колективу,
Редукція _ Існує _ Зв'язок,
Учасник _ Належить _ До _ Колективу}

Тоді подання онтології в Coq та міркування запишеться наступним чином:

1. Оголошення поняття відношення та класу для наслідування:

```
Definition Kind:= Type.
Parameter Relation: Kind → Kind → Prop.
```

2. Оголошення таксономії понять онтології:

```
Class VirtualScientificNetwork: Type.
Parameter D2: VirtualScientificNetwork → Kind.
Coercion D2: VirtualScientificNetwork >→ Kind.
Class VSNMessage: Type.
Parameter D3: VSNMessage → Kind.
Coercion D3: VSNMessage >→ Kind.
Class VSNMember: Type.
Parameter D4: VSNMember → Kind.
Coercion D4: VSNMember >→ Kind.
```

3. Означення відношень онтології та їх властивостей:

```
Definition Owned_by(x: VSNMessage)(y: VSNMember):= Relation x y.
Definition In_touch(x y: VSNMember):= Relation x y.
Axiom s_of_in_touch: Symmetric In_touch.
Axiom t_of_in_touch: Transitive In_touch.
Definition Part_of(x y: Kind):= Relation x y.
Axiom r_of_part_of: Reflexive Part_of.
Axiom a_of_part_of: Asymmetric Part_of.
Axiom t_of_part_of: Transitive Part_of.
```

4. Означення аксіом онтології:

```
Axiom Msg_Part_Owned_by:
forall
(message: VSNMessage) (member: VSNMember)(network: VirtualScientificNetwork),
Owned_by(message)(member) /\ Part_of(message)(network) →
Part_of(member)(network).
Axiom Mem_Part_Owned_by:
forall (message: VSNMessage)(member: VSNMember)(network: VirtualScientificNetwork),
Owned_by(message)(member) /\ Part_of(member)(network) →
        Part_of(message)(network).
Axiom intro_of_in_touch:
forall (m1 m2: VSNMember)(message: VSNMessage),
Owned_by(message)(m1) /\ Owned_by(message)(m2) →  In_touch(m1)(m2).
Axiom elim_of_in_touch:
forall (m1 m2: VSNMember), In_touch(m1)(m2) →
        exists message: VSNMessage, Owned_by(message)(m1)
        /\ Owned_by(message)(m2).
```

5. Доведення леми з використанням мови тактик:

```
Lemma Dist_Part_In_touch:
forall (m1 m2: VSNMember)(network: VirtualScientificNetwork),
In_touch(m1)(m2) /\ Part_of(m1)(network) → Part_of(m2)(network).
Proof.
intros m1 m2 net H1.
destruct H1 as [H1 H2].
apply elim_of_in_touch in H1.
elim H1; intros msg H3; clear H1.
destruct H3 as [H3 H4].
assert (H5: Owned_by(msg)(m1) /\ Part_of(m1)(net)).
split; assumption.
apply Mem_Part_Owned_by in H5.
assert (H6: Owned_by(msg)(m2) /\ Part_of(msg)(net)).
split; assumption.
apply Msg_Part_Owned_by in H6; assumption.
Qed.
```

Висновки. Отже, системно проаналізовано процеси управління персональною інформацією дослідника, що здійснює свою діяльність у межах віртуального наукового колективу. Розглянуто технологічні можливості побудови принципово нових, децентралізованих систем комунікації віртуального наукового колективу, з відмінними характеристиками приватності, безпеки, надійності та швидкодії. Висвітлено особливості функціонування розподілених баз даних Blockchain та IPFS, що є наріжними каменями P2P комунікаційних систем. Проаналізовано роль та особливості понять "дані", "інформація" та "знання", а також технологічні особливості їхнього адміністрування.

Значну увагу зосереджено на проектуванні персональних баз знань та здійсненні логічного міркування. Наведено принципово новий підхід до міркування в онтологічній моделі подання знань, що ґрунтується на використанні апарату теорії типів у межах системи асистента доведення теорем Coq. У вигляді таблиць висвітлено основні співвідношення між логікою висловлювань та теорією типів, а також структурними елементами онтологій та їх поданням у формальній мові системи Coq. Розглянуто приклад типізованого подання та міркування над фрагментом онтології, що описує предметну область функціонування віртуальної наукової спільноти.

Подальше дослідження охоплює ретельний аналіз структури та властивостей компонент проекту системи комунікації віртуальних наукових колективів, зокрема, подання онтологій та міркування з використанням теорії типів, функціонування P2P мереж (Homotopy type theory, 2013), застосування технології Blockchain у поєднанні з алгоритмами мульти-підпису та ієрархічного детермінованого створення ключів (Wuille, 2017), використання технології IPFS для приєднання великих даних до ланцюжка блоків, дослідження алгоритмів консенсусу, розроблення зручного інтерфейсу користувача тощо. Аналіз зазначених компонент здійснюватиметься як на теоретичному, так і емпіричному рівнях.

Перелік використаних джерел

Bitshares Munich IVS (2016). ECHO – Free Encrypted Private Chat. Retrieved from: https://my-echo.com/#messenger

Coq (2017). French Institute for Research in Computer Science and Automation. The Coq Proof Assistant. Retrieved from: https://coq.inria.fr

Cornell University. (2017). Nuprl Proof Development System. Retrieved from: http://www.nuprl.org

Dapoigny, R., & Barlatier, P. (2008). *Towards a Conceptual Structure based on Type theory*. Retrieved from: http://ceur-ws.org/Vol-354/p63.pdf

Dhamdhere, S. N. (2014). *Cloud computing and virtualization technologies in libraries*. Hershey, PA: Information Science Reference.

Feilmayr, C., & Wöß, W. (2016). An analysis of ontologies and their success factors for application to business. *Data & Knowledge Engineering, 101,* 1–23. https://doi.org/10.1016/j.datak.2015.11.003

Hafsi, M., Dapoigny, R., & Bolon, P. (2015). Toward a Type-Theoretical Approach for an Ontologically-Based Detection of Underground Networks. *Knowledge Science, Engineering and Management Lecture Notes in Computer Science*, 90–101. https://doi.org/10.1007/978-3-319-25159-2_8

Homotopy type theory: univalent foundations of mathematics. (2013). Princeton, NJ: Univalent Foundations Program.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big Data: Issues and Challenges Moving Forward. *2013 46th Hawaii International Conference on System Sciences.* https://doi.org/10.1109/hicss.2013.645

Lean (2017). Microsoft Research. The Lean Theorem Prover. Retrieved from: https://leanprover.github.io

Li, W., Sforzin, A., Fedorov, S., & Karame, G. O. (2017). Towards Scalable and Private Industrial Blockchains. *Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts – BCC 17*. https://doi.org/10.1145/3055518.3055531

Liew, A. (2007). Understanding Data, Information, Knowledge and Their Inter-Relationships. *Journal of Knowledge Management Practice, 8*(2), 1–10.

Lytvyn, V. (2013). Pidkhid do pobudovy intelektualnykh system pidtrymky pryiniattia rishen na osnovi ontolohii. *Problemy prohramuvannia, 4*, 43–52.

Marshall, A. (2016). Powered by Blockchain, New Decentralized Messenger to Save Data, Battery and Time. Retrieved from: https://cointelegraph.com/news/powered-by-blockchain-new-decentralized-messenger-to-save-data-battery-and-time

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Retrieved from: https://bitcoin.org/bitcoin.pdf

Noy, N., & McGuinness, D. (2001). Ontology Development 101: *A Guide to Creating Your First Ontology*. Retrieved from: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html

Omohundro, S. (2014). Cryptocurrencies, smart contracts, and artificial intelligence. *AI Matters, 1*(2), 19–21. https://doi.org/10.1145/2685328.2685334

Protocol Labs (2016). IPFS is the Distributed Web. Retrieved from: https://ipfs.io

Romano, D., & Schmid, G. (2017). Beyond Bitcoin: A Critical Look at Blockchain-Based Systems. *Cryptography, 1*(2), 15. https://doi.org/10.3390/cryptography1020015

Shen, A. (2014). *Algoritmicheskaia teoriia informatcii i sluchainost individualnyh obektov*. Retrieved from: https://www.youtube.com/watch? v=X0Lo5IWLjko

Swan, M. (2015). *Blockchain: blueprint for a new economy*. Sebastopol, CA: OReilly Media, Inc.

Veretennikova, N., Pasichnyk, V., Kunanets, N., & Gats, B. (2015). E-Science: New paradigms, system integration and scientific research organization. *2015 Xth International Scientific and Technical Conference "Computer Sciences and Information Technologies" (CSIT)*. https://doi.org/10.1109/stc-csit.2015.7325436

Wang, Y. (2015). Formal Cognitive Models of Data, Information, Knowledge, and Intelligence. *WSEAS Transactions on Computers, 14*, 770–781.

Wuille, P. (2017). Hierarchical Deterministic Wallets. Retrieved from: https://github.com/bitcoin/bips/blob/master/bip-0032.mediawiki.

Н. Э. Кунанец[1], В. С. Ленько[1], В. В. Пасичник[1], Ю. Н. Щербина[2]

[1] *Национальный университет "Львовская политехника", г. Львов, Украина*
[2] *Львовский национальный университет им. Ивана Франко, г. Львов, Украина*

ПЕРСОНАЛЬНЫЕ БАЗЫ ДАННЫХ И ЗНАНИЙ ВИРТУАЛЬНЫХ ИССЛЕДОВАТЕЛЬСКИХ СООБЩЕСТВ

Исследованы процессы приобретения и управления персональными данными и знаниями, а также коммуникации среди членов виртуальных исследовательских сообществ. Рассмотрено понятие виртуального научного коллектива и особенности защиты информации при организации процессов коммуникации в нем. Освещены проблемы коммуникации между участниками виртуальных исследовательских сообществ, возникшие из-за отсутствия целостного технологического решения, которое бы обеспечило конфиденциальность, быстродействие и структурированность потоков информации, данных и знаний. Предложен принципиально новый подход к проектированию платформ коммуникации, основанной на идеях децентрализации и крипто-безопасности. Проанализированы методологические различия между понятиями "данные", "информация" и "знание". Освещён современный подход к логическому мышлению в онтологической модели представления знаний, на основе использования аппарата теории типов и ассистента доказательства теорем Coq. Структурированы отношения между логикой и теорией типов, а также рассмотрен способ представления основных элементов онтологии с помощью формального языка системы Coq. Представлен пример рассуждения на языке тактик Ltac над фрагментом онтологии, который описывает определенные отношения предметной области функционирования виртуальных исследовательских сообществ. Указаны пути и инструменты проведения дальнейших исследований.

Ключевые слова: е-наука; коммуникация; блокчейн; онтология; теория типов; логическое рассуждение.

N. E. Kunanets[1], V. S. Lenko[1], V. V. Pasichnyk[1], Yu. M. Shcherbyna[2]

[1] *Lviv Polytechnic National University, Lviv, Ukraine*
[2] *Ivan Franko National University of Lviv, Lviv, Ukraine*

PERSONAL DATA AND KNOWLEDGE BASES OF VIRTUAL RESEARCH COMMUNITIES

The paper researches the processes of acquisition and management of personal data and knowledge, as well as the communication within the virtual research communities. It presents the concept of a virtual research group and the specifics of information protection during the establishment of communication processes within it. The paper highlights the existing issues in organization of an effective communication between the participants of virtual research communities, which are caused by the lack of a unified technological solution that ensures reliability, privacy, speed and common structure of an information, data and knowledge. A fundamentally new approach to the architecture of communication platforms based on the ideas of decentralization and crypto-security is proposed. It emphasizes the utility of a Blockchain solution, which combined with the distributed encrypted file system IPFS, InterPlanetary File System, provides the security of a next generation. The methodological differences between the concepts "data", "information" and "knowledge" are emphasized. The study provides a description of a modern approach to the reasoning in ontologies that is based on the use of a type-theoretical approach and Coq proof assistant. It highlights the relationship between the logic and type theory and defines the way of representing the core structural elements of ontology with the formal language of Coq system. An application of the Ltac tactics language for the reasoning over an ontology fragment, which describes some relations in the domain of the virtual research communities functioning, is considered. It shows the feasibility and advantages of the hierarchical type system of concepts, and provides a strong evidence of the higher-order reasoning in ontology. Finally, the paths and tools for further research are indicated.

Keywords: e-Science; communication; blockchain; ontology; type theory; logical reasoning.