

ОНТОЛОГОКЕРОВАНІ КОГНІТИВНІ СЕРВІСИ ЯК СКЛАДОВА ІНФОРМАЦІЙНО-ТЕЛЕКОМУНІКАЦІЙНИХ СИСТЕМ АНАЛІЗУ СТАТИСТИЧНОЇ ІНФОРМАЦІЇ

Приходнюк В. В., Горборуков В. В.

*Національний центр «Мала академія наук України»,
tangens91@gmail.com, slavon07@gmail.com*

ONTOLOGICAL COGNITIVE SERVICES AS A COMPONENT OF INFORMATION AND TELECOMMUNICATION SYSTEMS ANALYSIS OF STATISTICAL INFORMATION

In the operation of most information and telecommunication systems, there is a need to use various data, in particular, represented by natural language documents. To effectively operate with them, it is advisable to use modern cognitive services, especially ontology-controlled ones. This paper proposes the use of such a cognitive toolkit that implements the recursive reduction method and is an effective modern solution.

Згідно українського законодавства [1], інформаційно-телекомунікаційною системою вважається організаційно-технічна сукупність, що складається з автоматизованої системи та мережі передачі даних. Створення інформаційно-телекомунікаційних систем є надзвичайно складним процесом, що часто може бути зв'язаним з використанням великої кількості різнорідних засобів, модулів та інструментів.

В залежності від призначення інформаційно-телекомунікаційної системи, при її функціонуванні може виникати потреба у внесенні різноманітних даних, представлених природномовними документами. Часто об'єми таких даних є настільки великими, що внесення їх вручну може бути недоцільним або навіть неможливим. В таких випадках можуть використовуватись сучасні когнітивні сервіси, зокрема – онтологокеровані [2, 3]. Особливо гостро стає така задача при аналізі статистичної інформації, що часто представлена великими масивами слабкоструктурованих документів. Такі масиви в загальному випадку досить легко піддаються автоматизованій обробці, зокрема – за допомогою розробленого методу рекурсивної редукції [2].

Даний метод передбачає використання спеціалізованих онтологій (онтологічних дескрипторів структуризації) для опису перетворень, що повинні застосовуватись до вхідного слабо структурованого масиву даних. Метод застосовується до кожного з документів в рамках вхідного масиву даних і передбачає [2] виконання послідовності перетворень (1).

$$T \rightarrow T_{sn} \rightarrow T_{mk} \rightarrow \langle X, T_{mk} \rangle \rightarrow \langle X, R, T_{mk} \rangle \rightarrow \langle X, R, F, T_{mk} \rangle \rightarrow O \quad (1)$$

де T – текст вхідного природномовного документу; T_{sn} – лексична структура тексту; T_{mk} – розширена лексична структура; X, R, F – множини об'єктів, зв'язків і функцій інтерпретації вихідної онтології O .

Основою перетворення (1) є база правил рекурсивної редукції, що отримується в результаті інтерпретації онтологічного дескриптора структуризації. Кожне з правил в базі має вигляд (2), і складається з двох функцій – ідентифікації та перетворення.

$$d = \langle f_{id}^d, f_{tr}^d \rangle, \quad (2)$$

де f_{id}^d – функція ідентифікації лексем; f_{tr}^d – функція перетворення лексем.

$$F_{rd}(\beta_L, D) = \begin{cases} f_{tr}^d(\beta_0), \exists d \in D, f_{id}^d(\beta_0) \\ F_{rd}(\beta_L \setminus \beta_0, D) \end{cases} \quad (3)$$

де β_L – множина всіх підмножин, сформована для певної множини лексем L ; D – база правил виду (2); f_{id}^d, f_{tr}^d – елементи певного правила d .

Прикладом дескриптора структуризації є онтологія, що задає процес створення обробки статистичних даних щодо кадрового забезпечення установ НАН України (Рис. 1).

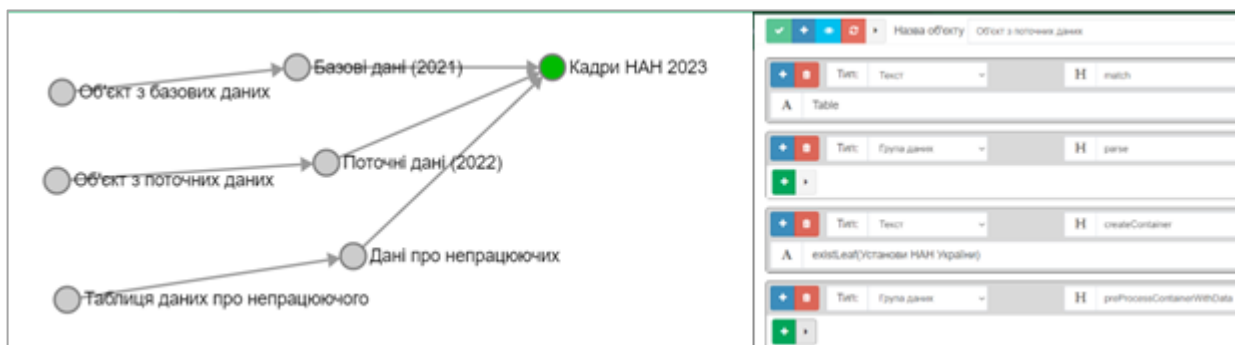


Рис. 1. Приклад онтологічного дескриптора структуризації.

Такий дескриптор може використовуватись для наповнення даними автоматизованих робочих місць в рамках певного мережевого середовища (Рис. 2), призначеного для роботи експертів.

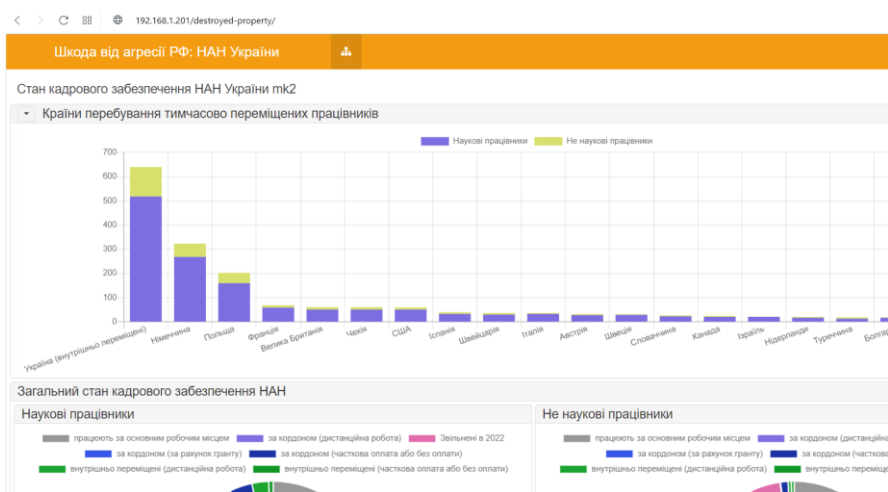


Рис. 2. АРМ відображення результатів в рамках локального мережевого середовища

Інструментарій, що реалізує метод рекурсивної редукції, може легко інтегруватись в різноманітні інформаційно-телекомунікаційні системи, як відкриті (з доступом з мережі Інтернет) так і закриті. В залежності від

розгорнутої в рамках системи інфраструктури інструментарій може працювати як локально (на робочому місці оператора системи) так і в рамках певного виділеного серверу (серверу додатків). Онтології, що керують роботою інструментарію, також можуть розміщуватись як локально (в файловій системі робочого місця або серверу), так і в рамках виділеного серверу онтологій (серверу КІТ «ПОЛІЕДР»).

Запропонований інструментарій є ефективним сучасним рішенням, що дозволяє суттєво зменшити витрати часу операторів інформаційно-телекомунікаційних систем, зокрема, таких, що призначені для аналізу статистичної інформації.

Література

1. Наказ “Про затвердження Порядку захисту державних інформаційних ресурсів у інформаційно-телекомунікаційних системах” від 24.12.2001 N 76. URL: <https://zakon.rada.gov.ua/laws/show/z0027-02#Text>
2. Stryzhak O., Prykhodniuk V., Popova M., Nadutenko M., Haiko S., Chepkov R. Development of an Oceanographic Databank Based on Ontological Interactive Documents. Lecture Notes in Networks and Systems. Cham : Springer. 2021. С. 97–114. DOI: https://doi.org/10.1007/978-3-030-80126-7_8
3. Nadutenko M., Prykhodniuk V., Shyrokov V., Stryzhak O. Ontology-Driven Lexicographic Systems. Advances in Information and Communication. FICC 2022. Lecture Notes in Networks and Systems. Cham : Springer. 2022. С. 204–215. DOI: https://doi.org/10.1007/978-3-030-98012-2_16