

ВИКОРИСТАННЯ ОНТОЛОГІЧНИХ ЗНАТЬ ДЛЯ БАГАТОКРИТЕРІАЛЬНОГО СПІВСТАВЛЕННЯ СКЛАДНИХ ІНФОРМАЦІЙНИХ ОБ'ЄКТІВ

Юлія Рогушина, Анатолій Гладун

Співставлення складних інформаційних об'єктів (CIO) розглядається в роботі як складова інтелектуального прийняття рішень. Особливістю запропонованого підходу є те, що співставляються не всі теоретично можливі CIO, а лише та їх підмножина, яка може бути обрана в поточній ситуації. Таким чином, проблема полягає не у пошуку оптимального (за певними критеріями) рішення, а у виборі прийняттого рішення з набору наявних.

Запропоновано формальну модель CIO як елемента інтелектуальних інформаційних систем, що базується на онтологічному поданні знань. Ця модель (на відміну від онтології Про) виокремлює позиції класів та елементів класів для позначення структурних елементів CIO, яким надаються унікальні імена. Розглянуто методи співставлення CIO, що базуються на використанні знань онтології відповідної предметної області. В якості метрик для кількісного оцінювання розглянуті різні методи визначення семантичної близькості та семантичної подібності, щоб визначити ті параметри інформаційних об'єктів, за якими обчислюються ці оцінки. Запропоновано алгоритм семантичного співставлення CIO, які базуються на одній онтології та мають подібну структуру, що дозволяє генерувати множину критеріїв співставлення та визначити їх ієрархію для поточної ситуації.

Це дозволяє оцінювати подібність на змістовному рівні елементів окремих CIO до певного еталону, який задається користувачем (як опис оптимального рішення або того CIO, що потрібно знайти або згенерувати). В результаті цього створюється підмножина CIO, які задовольняють вимоги користувача, але серед них потрібно обрати лише один CIO, який надалі буде застосовуватися для виконання задачі користувача. Тому виникає потреба у знаходженні множини критеріїв, за якими можна порівнювати такі CIO, та у методах визначення важливості кожного з цих критеріїв у поточний момент часу. Пропонується використовувати для цього метод ієрархічного аналізу, що базується на попарному порівнянні важливості окремих критеріїв.

Ключові слова: складний інформаційний об'єкт, онтологія, семантична подібність, семантична близькість, аналіз ієрархій.

In this work we consider comparison of complex information objects (CIO) as a component of intelligent decision-making. The specifics of proposed approach is that we compare not all theoretically possible CIOs but only their subset that is relevant for current situation and contains existing and available objects. Thus, we find an acceptable solution from the set of available ones that can be not optimal (according to certain criteria). We propose formal ontology-based model of CIO that considered as an element of intellectual information system. This model (in contrast to the domain ontology) defines unique names for positions of classes and class individuals to indicate the SIO structure. The methods of CIO comparison based on the use of knowledge from the relevant domain ontology are considered. Various approaches to determining of semantic proximity and semantic similarity are considered as metrics for quantitative evaluation to select parameters of information objects that can be used for calculation of these evaluations. We propose an algorithm for semantic comparison of CIOs which are based on the same ontology and have a similar structure. This algorithm allows generation of comparison criteria and determining hierarchy of this criteria for the current situation.

We propose to evaluate the semantic-level similarity of the elements of individual CIOs to certain reference CIO defined by the user (as a description of the optimal solution or generated CIO properties). As a result, a subset of CIOs that satisfy the user requirements is created, but we have to select only one CIO among them, which will be used in the future to fulfill the user task. Therefore, we need in a set of criteria for CIO comparison and methods to determine the importance of each of these criteria at the current moment in time. For this purpose, we propose to use the method of hierarchical analysis based on a pairwise comparison of the importance of individual criteria.

Keywords: complex information object, ontology, semantic similarity, semantic proximity, analysis of hierarchies.

Вступ

Співставлення складних інформаційних об'єктів (CIO) може розглядатися як складова інтелектуального прийняття рішень. Водночас проблеми генерації рішення як послідовності певних дій в динамічному інформаційному оточенні перебувають поза сферою розгляду цієї роботи. Основна увага приділяється багатокритеріальному порівнянню результатів (потенційних або реальних) обраного рішення. Особливістю запропонованого підходу є те, що співставляється відносно невелика кількість CIO, тобто не всі теоретично можливі, а лише ті, які можуть бути обрані в даній ситуації. Отож, проблема полягає не у пошуку оптимального (за певними критеріями) рішення, а у виборі прийняттого рішення з набору наявних. Наприклад, для виконання певного проекту потрібно обрати групу співробітників з певного підрозділу, а не взагалі з усіх людей. Тому припустима ситуація, коли всі можливі рішення є незадовільними і тільки змінюють ситуацію на гіршу. Зокрема, виконання проекту недостатньо компетентними співробітниками призведе до втрати часу та ресурсів, але потрібний результат не буде отримано. При цьому значущість критеріїв може змінюватися з часом через зміни в динамічному інформаційному оточенні. Найбільш поширеним прикладом зміни пріоритетів є вартість виконання робіт та швидкість отримання результатів.

Складні інформаційні об'єкти

Із точки зору онтологічного аналізу інформаційні об'єкти (IO) – це будь-які класи або екземпляри онтології. При цьому класи характеризуються своєю структурою – набором властивостей та їхніх характеристик, а також можливими відношеннями з іншими класами. У екземплярів класів онтології можуть також бути визна-

чені значення (всіх або деяких) властивостей. Але у багатьох практичних задачах потрібно аналізувати більш складні сукупності інформації, в яких ІО пов'язані один з одним певними відношеннями та обмеженнями. За приклад такої задачі можна навести семантичний пошук: знайти групу людей з певною кваліфікацією, що працюючій в одній організації, яка відповідає умовам конкурсу; визначити країни, в яких проводилися наукові дослідження певної тематики, результати яких були опубліковані в обраній множині журналів за певний період часу тощо. Але якщо результатом пошуку є здебільшого набір ІО одного або кількох класів, тоді як обмеження використовуються тільки для відбору прийнятних варіантів, то у багатьох інших задачах сам результат є множиною різних сукупностей ІО різних типів, кожна з яких відповідає певним умовам. Надалі такі сукупності ІО будемо називати складними інформаційними об'єктами.

СІО – це набір з більш ніж одного ІО, які пов'язані один з одним онтологічними відношеннями та відповідають вимогам щодо структури та значень властивостей ІО. Приклади СІО – організація, набір її співробітників та множина проектів, що виконуються в організації; кілька ієрархічно пов'язаних підрозділів, які виконують спільну задачу з використанням набору технічних засобів; інфраструктура населеного пункту, засоби її підтримки та особовий склад, що здійснює цю підтримку; навчальний заклад, набір спеціальностей, за якими проводиться навчання, та множина компетенцій, які цими спеціальностями забезпечуються [1].

Прикладом використання СІО є концепція ієрархічного агрегатного оцінювання (ІАО). Ця концепція є частиною інтеграції навчального процесу. Вона полягає в оцінюванні команд (такі команди можна розглядати як набори ІО типу “Особа”, що пов'язані відношеннями – як ієрархічними, так і специфічними для предметної області (ПрО)) з кількома рівнями ієрархії шляхом представлення іспитів або завдань оцінювання, які потрібно виконати в командах із кількома рівнями ієрархії. Індивідуальні оцінювання можуть виконуватися для членів команди, а інші оцінювання можуть бути застосовані до роботи всієї команди або її підмножин, а також до кінцевого результату роботи. Критерії оцінювання в ІАО відрізняються залежно від ієрархічної позиції особи: члени команди мають ієрархічні посади, такі як: член команди, лідер групи, а також менеджер групи або адміністратор (який очолює лідерів команди) тощо, які є також параметрами для визначення оцінок. Даний підхід передбачає оцінювання в одному завданні різних цілей, навичок, умінь, знань, продуктивності тощо. Оцінювання командної роботи передбачає колаборативну роботу та колаборативне проблемно-орієнтоване навчання за наявності комплексних завдань оцінювання з кількома рівнями ієрархії. Завдання для оцінювання містять ієрархію членів команди, завдання з оцінювання, ієрархію онтологій, а також результати оцінювання (оцінки, самооцінювання та експертні оцінки). Окремим випадком СІО є рішення, які ухвалюються в різних областях з використанням знань щодо структури ІО та інших закономірностей ПрО.

Ми маємо досить великий досвід розробки інформаційних систем, які потребують співставлення різноманітних СІО. В даній роботі ми розглядаємо аспекти, які стосуються генерації критеріїв такого співставлення на основі онтології ПрО та визначення їхньої значущості для поточного стану інформаційного оточення.

Онтологічна модель СІО

Формальна модель СІО базується на онтології ПрО:

$$O_{\text{domain}} = \langle T, R, F \rangle \quad (1),$$

де T – скінченна множина термінів ПрО, яку описує онтологія O , що поділяється на множину класів T_{cl} та множину екземплярів класів T_{ind} ; R – скінченна множина відношень між термінами заданої ПрО; F – скінченна множина функцій інтерпретації, заданих на термінах і/або відношеннях онтології O . Вона містить підмножину елементів такої онтології, в якій виокремлюється набір пар екземплярів класів, між якими встановлено певні семантичні відношення.

Така модель СІО має наступну структуру:

$$C = \langle T_C = \{t_i, i = \overline{1, p}\} \subseteq T, N_C = \{n_i, i = \overline{1, p}\}, R_C \subseteq R, \{(t_j, t_k, r_m), t_j \in T_{ind}, t_k \in T_{ind}, r_m \in R\} \rangle \quad (2).$$

Важливо, що в моделі СІО на відміну від онтології ПрО виокремлюються позиції класів та елементів класів для позначення структурних елементів СІО, яким надаються унікальні імена $N_C = \{n_i, i = \overline{1, p}\}$. З кожним таким елементом може пов'язуватися набір характеристик та обмежень (наприклад, елемент обов'язково присутній, може мати єдиний екземпляр або кілька екземплярів, не може співпадати з екземпляром певного іншого елемента СІО тощо). У тих випадках, коли таке виокремлення не є важливим для ПрО, кілька формально різних СІО розглядаються як один. Наприклад, якщо в структурі СІО присутні ІО класу “Особа”, які позначаються назвами “Керівник проекту” та “Виконавець проекту”, то для елемента СІО “Керівник проекту” вказано, що елемент обов'язково має бути присутній та може мати лише єдиний екземпляр, а для елемента СІО “Виконавець проекту” вказано, що елемент може мати більше одного екземпляра, а їх порядок не є значущим, тобто між СІО, що відрізняються тільки порядком відомостей про виконавців, на змістовному рівні немає різниці. Слід звернути увагу, що такий підхід не прийнятний для СІО, де явно визначені окремі ІО – наприклад, вказано конкретні посади учасників проекту (“Програміст”, “Адміністратор” та “Дизайнер”).

Інтелектуальне прийняття рішень

Інтелектуальне прийняття рішень (Decision intelligence) є одним з актуальних напрямків наукових досліджень [2]. Це нова наукова дисципліна, яка вивчає різні аспекти вибору варіантів. Вона об'єднує приклад-

ні з дослідження даних (data science), штучного інтелекту, соціальних наук та теорії менеджменту в єдину. Застосування цих досліджень забезпечує засоби розробки цілей, метрик та критеріїв для оцінки рішень в різних сферах. У цьому контексті рішення – це будь-який вибір між варіантами для будь-якої сутності – людини, організації, програмного агента тощо. Основою для вибору рішення є дані. В такому розумінні рішення не є бінарною альтернативою щодо виконання однієї конкретної дії.

Особа, яка ухвалює рішення (ОПР), відповідає за архітектуру рішень і формування контексту, у тому числі – обирає ті набори даних, які використовуються для генерації та співставлення рішень.

Ухвалення рішень по-різному використовується різними дисциплінами, тому воно може стосуватися:

- виконання дії, якщо були альтернативні варіанти (людиною, твариною або комп'ютерною програмою);
- виконання такої функції особою, яка ухвалює рішення, що частиною цієї функції є відповідальність

за рішення (це може бути тільки людина або група людей, але не комп'ютерна система, оскільки за результати відповідатиме не така система, а ті хто її створив та використовував).

Не всі результати обчислень або пропозиції є рішеннями: рішення вважається ухваленим лише після розподілу ресурсів. Поки існує можливість змінити вибір без додаткових ресурсів, рішення ще не ухвалено.

В інтелектуальному ухваленні рішень можна виділити кількісні аспекти (що значною мірою збігаються з прикладною наукою про дані) та якісні аспекти (розроблені в основному дослідниками в соціальних і управлінських науках). Крім того, доцільно розуміти, чи існують факти, спроможні змінити зроблений вибір. До основних проблем в ухваленні рішень належать:

- вибір критеріїв ухвалення рішень;
- розробка сумісних метрик для їх кількісного оцінювання;
- оцінка якості та вартості даних, які використовуються для ухвалення рішення;
- вимоги до якості рішення;
- аналіз факторів, що впливають на процес ухвалення рішення;
- вибір засобів подання інформації, яка використовується для пошуку рішення;
- оптимізація результатів під час групового ухвалення рішень;
- ієрархічність та багатоетапність цілей в розробці контексту ухвалення рішень.

Багато з цих проблем потребують залучення зовнішніх джерел знань щодо ПрО рішення, поточних інтересів та потреб ОПР та середовища, в якому реалізуються ці рішення. Для цього можуть залучатися різні онтологічні структури, що різняться як обсягом, так і структурною складністю.

Це обґрунтовується тим, що стратегії ухвалення рішень, які базуються лише на кількісних оцінках без якісного розуміння ухвалення рішень і та закономірностей ПрО, як правило, менш ефективні порівняно зі стратегіями, які використовують також елементи семантичного аналізу.

Важливо розуміти, що ухвалення будь-яких рішень (так само, як і співставлення більш узагальнених СІО) залежить від наявної інформації. Якщо рішення ухвалюються у відкритому інформаційному середовищі (а саме це найбільш характерно для практичних задач), то інформація може бути неповною, нечіткою та суперечливою. Крім того, частина фактів може бути просто недостовірною. Але навіть в умовах повної інформованості велике значення має як форма подання інформації, так і наміри ОПР. Маємо розуміти, що практично для будь-якого припустимого рішення, яке розглядається, можна обрати такий набір критеріїв оцінювання й так визначити їхні пріоритети й джерела даних, що саме це рішення матиме найвищу оцінку.

Інженерія даних як окрема наукова дисципліна є базовою складовою аналізу рішень, але науки про ухвалення рішень містять також методи та засоби для розробки й контролю збору фактів. У багатьох випадках потрібно прогнозувати значення даних, що будуть отримані у майбутньому, або за наявною обмеженою вибіркою апроксимувати значення ширшого набору даних. Data science забезпечує таку інформацію, використовуючи статистичні методи.

Співставлення СІО

Співставлення СІО на семантичному рівні може використовувати онтологічні знання двох типів:

- щодо семантичної близькості між ІО, що займають певне місце в різних СІО;
- щодо семантичної подібності між відношеннями, які пов'язують ці ІО.

В обох випадках аналізуються як семантична відстань між відповідними концептами онтології, так і близькість значень властивостей ІО, що є екземплярами класів. Якщо порівняння кількісних значень є досить простою задачею, то співставлення якісних характеристик потребує також визначення семантичної відстані між відповідними об'єктами.

За інструмент для такого оцінювання можуть використовуватися різноманітні метрики для кількісного оцінювання семантичної близькості та семантичної подібності, які перетворюють якісне представлення знань в онтології ПрО у кількісні характеристики їх семантичної близькості та спорідненості.

Огляд оцінок семантичної близькості та семантичної подібності

Одним із основних інструментів пізнання, концептуалізації навколишнього світу є мова. Онтологія є ієрархічною семантичною мережею, вершинами якої є концепти (сміслові одиниці), а дугами – семантичні відношення між концептами. Семантика (смісл) концепту описується його смисловими відношеннями до інших концептів мережі [3]. Для співставлення СІО необхідно виявити семантичну подібність та семантичну близькість між концептами їхніх онтологічних моделей.

Є істотна різниця між поняттями “семантична подібність” і “семантична близькість”: останнє поняття значно ширше. Семантична подібність найбільше визначається за відношеннями синонімії та “клас-підклас” між поняттями, тоді як семантична близькість враховує усі характерні для ПрО відношення між цими поняттями (наприклад, відношення антонімії або меронімії) [4].

Семантична подібність пов’язана їх інформаційним змістом (контентом). Інформаційний контент поняття c можна кількісно визначити як $-\log p(c)$: чим вище ймовірність використання поняття, тим нижче його інформативність. Отож, чим вище рівень абстракції поняття (тобто, чим вище воно є в таксономії), тим менше його інформаційний контент. Подібність концептів оцінюється через знаходження максимального інформаційного контенту над усіма концептами, для яких обидва слова можуть бути екземпляром. Це дає змогу створювати множини семантично близьких слів (СБС), тобто слів, семантична відстань між якими менша за обрану порогову величину.

Аналіз робіт, що стосуються методів семантичної близькості та семантичної подібності можна розділити на 4 групи: 1) алгоритми семантичної подібності, які базуються на знаннях [5]; 2) методи, засновані на інформаційному контенті [6]; 3) методи семантичної близькості [7] на основі різних способів векторного подання слів; 4) комбіновані методи оцінки семантичної подібності [8] та узагальнені методи [9].

Методи обчислення семантичної подібності між концептами СЮ можна розділити на групи:

- на основі атрибутів;
- на основі контенту (вмісту);
- на основі семантичної відстані;
- гібридні методи [10],[11],[12].

У більшості методів, використовуваних для співставлення онтологій СЮ, семантична подібність оцінюється на основі пар “атрибут-значення” та застосовує кількісні показники подібності, які використовують ці значення атрибутів. Однак, простий вектор атрибутів недостатньо відображає складність СЮ, які з’являються на практиці, насамперед потрібно знати структуру онтологій СЮ [13].

Значна частина досліджень припускає, що семантична подібність має враховувати ієрархічну структуру онтологій СЮ. Однак застосування алгоритму подібності, що базується на атрибутах, часто призводить до певних обчислювальних помилок. Наприклад, у реальності еталонна онтологія та порівнювана онтологія можуть мати або не мати нульові значення атрибутів [14]. Алгоритм подібності на основі контенту визначає подібність двох класів шляхом порівняння інформації про контент, що міститься в загальному батьківському вузлі класів, і ігнорує інформацію контенту, який міститься в самому ключовому слові. Основна ідея алгоритму семантичної подібності на основі відстані полягає в обчисленні семантичної відстані між двома концептами в системі класифікації дерева онтологій [15]. Однак основним недоліком цього методу є припущення, що відстань усіх ребер у системі є однаково важливою. Резнік [16] висунув припущення, що оцінку подібності в семантичних структурах можна розглядати як таксономічні відношення, ігноруючи інші відношення, а це призводить до втрати деякої потенційно корисної інформації. В алгоритмах співставлення, за аналогією із семантичною подібністю, семантичною кореляцією, також ігнорують велику кількість корисної інформації.

Семантична подібність і семантична кореляція між двома об’єктами вже давно є основними проблемами в питанні аналізу даних і керування знаннями, але це два різні поняття [17].

Семантична кореляція має відношення до ступеня взаємозв’язку двох концептів. Часто-густо між двома концептами відсутня семантична кореляція, але кореляційні відношення можуть формуватися з інших причин. Семантична подібність – це агрегація понять і відношень, тоді як семантична кореляція – це комбінування понять. Пояснення різниці між подібністю та кореляцією приведемо на прикладі автомобілів, бензину та велосипедів, де автомобілі є залежними від бензину як палива, і, очевидно, тісніше пов’язані один з одним, ніж автомобілі та велосипеди. Але прийнято вважати автомобілі та велосипеди більш схожими одне на одне, ніж автомобілі та бензин.

В онтологічній структурі є подібність у відношенні між двома поняттями, які зазвичай асоціюють символом «is-a», і кореляційним відношенням між двома поняттями, які пов’язані іншими відношеннями, як-от, «part-of». Зазначимо, що семантичну подібність або кореляцію ґрунтують на інтерпретації або контексті, а поняття, які подібні або релевантні з одного боку, можуть не бути подібними або релевантними з іншого.

Методи семантичної подібності засновані на знаннях, обчислюють семантичну подібність між двома концептами на основі інформації, отриманої із основних джерел знань, таких як онтології/лексичні бази даних, тезауруси, словники тощо. База знань надає цим методам структуроване представлення понять, пов’язаних семантичними відношеннями, що забезпечує семантичну міру без двозначності, як фактичне значення термінів [18].

Методи підрахунку ребер. Метод розглядає основну онтологію як граф, що зв’язує терміни таксономічно, і підраховує кількість ребер між двома термінами для обчислення подібності між ними. Чим більша відстань між термінами, тим менше вони подібні. Цей метод визначає міру *path* і був запропонований [19], у ньому подібність обернено пропорційна найкоротшій довжині шляху між двома термінами. У цьому методі підрахунку ребер, слова, що знаходяться глибше внизу ієрархії мають конкретніше значення. Вони можуть бути більш схожими один на одного, навіть незважаючи на те, що мають однакову відстань. Ву та Палмер [20] запропонували міру *wup*, де важливим атрибутом є глибина слів в онтології. Міра *wup* підраховує кількість ребер між кожним терміном та їх найменшим загальним родовим об’єктом, що охоплює видові, тобто Least Common Subsumer (LCS) – найбільш конкретний спільний предок двох концептів в цій онтології. Лі (Li) та ін. [21] пропонують міру, яка враховує як мінімальну відстань шляху, так і глибину.

Однак методи підрахунку ребер ігнорують той факт, що в онтологіях ребра не повинні мати однакову довжину. Для подолання цього недоліку методів підрахунку ребер було запропоновано методи семантичної подібності, засновані на ознаках.

Методи на основі атрибутів обчислюють подібність як функцію властивостей слів, таких як «глос», сусідні поняття тощо [22]. «Глос» визначають як значення тлумачення слова в словнику; колекцію глосів називають глосарієм. Міри семантичної подібності, що базуються на глосі, використовують знання, що слова зі схожим значенням мають більш загальні слова у своєму глосі. Семантичну подібність вимірюють як ступінь перекриття глосів розглянутих слів. Міра подібності [23] обчислює значення подібності між двома словами на основі перекриття слів у їхньому глосі та глосів понять, з якими вони пов'язані, в онтології.

Методи на основі інформаційного контенту (ІК) концептів визначають ІК як інформацію, що походить від концепту, коли він з'являється в певному контексті [24]. Високе значення ІК вказує на те, що слово є більш конкретним і чітко описує поняття з меншою двозначністю, тоді як низьке значення ІК вказує, що слова абстрактніші за значенням. Унікальність слова визначають за допомогою інверсної частотності документа (Inverse Document Frequency – IDF), яка спирається на принцип, що, чим конкретніше слово, тим рідше воно зустрічається в документі. Такі методи вимірюють подібність між термінами, використовуючи значення ІК, що пов'язане з цими термінами. Резнік Ф. [16] пропонує міру семантичної подібності *res*, базуючись на ідеї, що якщо два концепти мають спільний загальний родовий об'єкт, який охоплює видові, то вони поділяють більше інформації. Лін [15] пропонує покращення міри *res* з урахуванням як значення ІК термінів, що стосуються окремої інформації або опису термінів, так і значення ІК їх LCS, що забезпечує уніфіковану спільність між термінами. Цзян і Конрат [25] обчислюють міру відстані на основі різниці між сумою окремих значень ІК термінів і величиною ІК їх LCS. ІК може бути обчислено за допомогою основних корпусів або з внутрішньої структури самої онтології припускаючи, що онтології структуровано значущим чином. Деякі терміни можуть входити до різних онтологій, що дає змогу використовувати кілька онтологій для обчислення їхніх відношень.

Комбіновані методи поєднують різні комбінації методів, розглянутих вище. Наприклад, Гао та ін. [26] запропонував метод обчислення семантичної подібності на основі онтології WordNet, де для додавання ваг до ребер використовують три різні стратегії, а для обчислення семантичної подібності використовують найкоротший зважений шлях. Згідно з першою стратегією, глибину всіх термінів у WordNet вздовж шляху між двома розглянутими термінами додають як ваги до найкоротшого шляху. У другій стратегії як ваги було додано лише глибину LCS термінів, а в третій стратегії як ваги додано значення термінів ІК. Потім обчислюють найкоротшу зважену довжину шляху та нелінійно трансформують для отримання міри семантичної подібності.

Комбіновані методи подібності враховують як інформацію щодо положення (відстань, ієрархію) класу в дереві онтологій, так і інформацію щодо контенту (вміст) і забезпечують знаходження СІО, що найбільш схожі до еталонного СІО. Цей процес можна представити у вигляді п'яти кроків:

- 1) обчислення семантичної подібності атрибутів,
- 2) обчислення семантичної кореляції атрибутів,
- 3) обчислення агрегованої семантичної подібності атрибутів і семантичної кореляції,
- 4) виведення ваг атрибутів моделей,
- 5) обчислення зваженої подібності моделей.

Оцінки семантичної подібності між поняттями Про дозволяють формалізувати інформаційні потреби користувачів та описувати структуру й властивості бажаних рішень [27]. Вони можуть використовуватися для побудови формалізованого тезаурусу задачі, що стає джерелом інформації щодо структури СІО [28].

Постановка задачі

Розглянуті вище метрики дозволяють кількісно оцінювати семантичну близькість між елементами різних СІО та між різними СІО в цілому. Це дозволяє оцінювати подібність на змістовному рівні елементів окремих СІО до певного еталону, який задається користувачем (як опис оптимального рішення або того СІО, що його потрібно знайти чи згенерувати). В результаті створюється підмножина СІО, які задовольняють вимоги користувача, але серед них потрібно обрати лише один СІО, який надалі буде застосовуватися для виконання задачі користувача. Тому виникає потреба у знаходженні множини критеріїв, за якими можна порівнювати такі СІО, та у методах визначення важливості кожного з цих критеріїв у даний момент часу.

Етапи співставлення СІО

У загальному випадку задача співставлення СІО, які мають різну структуру та базуються на різних онтологіях, потребує вирівнювання цих онтологій та пошуку подібності між їхніми структурними елементами – ІО та їх поєднаннями. В даній роботі ми розглядаємо окремий випадок такої задачі, коли всі СІО, які потрібно співставити, базуються на єдиній онтології та мають однакову (або подібну) структуру і різняться між собою значеннями властивостей та наявністю цих значень. Така ситуація характерна для пошукових задач, для задач ухвалення рішень та для різноманітних рекомендаційних та дорадницьких систем (на практиці в таких системах співставляються СІО з різною структурою, такі як резюме та вакансії, але семантичне співставлення виконується для елементів, що мають подібну структуру – наприклад, для набору компетенцій або описів проєктів). В тих задачах, де потрібно співставляти СІО з різною структурою, дана ситуація обробляється на останньому етапі порівняння та є складовою більш комплексного процесу співставлення.

- Задача співставлення СІО з подібною структурою підрозділяється на такі незалежні етапи;
- створення еталонної формалізованої структурної моделі СІО, яка відображає основні вимоги та обмеження, що визначаються задачею користувача;
 - генерація наявних СІО, які структурно відповідають еталонній моделі та відомості про екземпляри реального інформаційного середовища на певний момент часу;
 - вибір серед наявних СІО тих, що відповідають вимогам користувача на семантичному рівні, тобто розміщені на семантичній відстані від еталонної моделі не більше, ніж на визначене значення;
 - генерація набору критеріїв співставлення СІО з використанням знань з онтології ПрО;
 - визначення рівня значущості кожного окремого критерію на даний момент співставлення на основі експертних оцінок та евристик ПрО (важливо враховувати, що значущість критеріїв у динамічному інформаційному середовищі може значно змінюватися у часі для тієї самої задачі користувача, але набір критеріїв, що будуватиметься для певної задачі на основі обраної онтології ПрО, як правило, не змінюється взагалі або тільки поповнюється додатковими критеріями);
 - визначення кількісної оцінки кожного СІО на основі сукупності критеріїв його подібності до еталонної моделі та вибір найбільш придатного СІО (потрібно враховувати, що такий вибір СІО не є оптимальним у глобальному розумінні, і зміни рівня значущості критеріїв впливають на такий вибір).

Генерація набору критеріїв співставлення СІО

Як показує огляд методів визначення семантичної подібності між поняттями ПрО, у найбільш загальному вигляді основними параметрами визначення семантичної близькості між двома екземплярами класів онтології є: 1. семантична відстань між класами, до яких належать ці екземпляри (для її визначення можуть враховуватися лише відношення з певними характеристиками або з обраної множини, досить часто використовуються лише ієрархічні та синонімічні відношення, але відповідно до специфіки задачі можуть бути обрані будь-які специфічні для ПрО онтологічні відношення); 2. семантична близькість між значеннями властивостей одного типу цих екземплярів.

Задача значно спрощується, якщо порівнюються екземпляри одного класу. Тоді першу групу параметрів можна не враховувати, а для другої не виникає потреби у вирівнюванні властивостей різних класів, яка потребує використання їхньої семантики (наприклад, властивість типу «Рік» може характеризувати як рік народження особи, так і рік початку навчання). У співставленні СІО в даній роботі ми будемо аналізувати саме таку ситуацію, коли співставляються СІО з однаковою (або близькою) структурою. Проблема співставлення СІО з різною структурою є більш комплексною і потребує додаткових етапів обробки інформації.

Таким чином, для попередньої генерації набору критеріїв K_0 для співставлення СІО з подібною структурою пропонуємо наступний алгоритм:

- для кожного ІО $t_i, i = \overline{1, p}$ з формальної моделі СІО (2) визначається множина його властивостей, які поділяються на властивості даних $t_{d_{im}}, i = \overline{1, p}, m = \overline{0, y_i}$ (їхні значення є константи різних типів – число, текст, дата тощо), $T_{data_{0i}} = \{t_{d_{im}}, m = \overline{0, y_i}\}$ та об'єктні властивості $t_{o_{ik}}, i = \overline{1, p}, m = \overline{0, z_k}$ (їхні значення є екземпляри різних класів О), $T_{o_{0i}} = \{t_{o_{ik}}, m = \overline{0, z_k}\}$ що визначені в онтології ПрО О для класу, до якого належить ІО: $K_{0i} = T_{d_{0i}} \cup T_{o_{0i}}$.

- Для властивостей даних на цьому аналіз закінчується, а для об'єктних властивостей за потреби може повторюватися рекурсивно для кожного класу ІО, що є значеннями властивостей для t_i . Водночас множина K_{0i} поповнюється відповідними властивостями, які надалі можуть розглядатися як критерії для співставлення СІО;

- На основі об'єднання інформації множин K_{0i} будується множина K_0 , в якій явно фіксується, до якого саме ІО з СІО належить кожен із критеріїв оцінювання: $K_0 = \{(t, i), t \in K_{0i}, i = \overline{1, p}\}$;

- Користувач аналізує побудовану множину K_0 та явно видаляє з неї ті критерії, які вважає несуттєвими для його задачі.

На цьому робота алгоритму можна закінчити, або продовжити вдосконалення множини критеріїв із використанням онтології ПрО та різноманітних мір семантичної близькості та семантичної подібності, розглянутих вище.

Множина K_{sem} будується наступним чином:

- Визначається підмножина класів онтології О, для яких в K_0 міститься відповідний критерій: $T_{C_{cl}} \subseteq T_C, T_{C_{cl}} \subseteq K_0$;

- Для кожного з елементів $t_j \in T_{C_{cl}}, t = \overline{1, x}$ визначається множина семантично близьких або семантично подібних (відповідно до вибраної міри $f(t_a, t_b)$ та константи L) понять онтології ПрО: $S_j = \{t \in T_{C_{cl}}, f(t, t_j) \leq L\}$;

- За кожною множиною $S_j, j = \overline{1, x}$ будується множина критеріїв K_{sem_j} за таким самим алгоритмом, як будуються множини K_{0i} ;

- Надалі аналогічно тому, як будується множина K_0 , ці множини об'єднуються в множину K_{sem} , в якій теж явно фіксується, до якого саме ІО з СІО належить кожен з критеріїв оцінювання: $K_{sem} = \{(t, i), t \in K_{sem_j}, j = \overline{1, x}\}$.

Відтак користувач аналізує побудовану множину K_{sem} та явно видаляє з неї критерії, які він вважає несуттєвими для його задачі. Крім того, користувач може вручну додати ті критерії, які він вважає важливими, але які не відображені в онтології ПрО або не ввійшли до K_{sem} внаслідок виконання запропонованого вище

алгоритму. Причиною цього може бути невдало обрана онтологія або міра семантичної близькості, а також недостатня кількість ітерацій алгоритму. Інша причина може бути у наявності у користувача глибоких експертних знань щодо ПрО, коли йому легше явно вказати важливі критерії, ніж шукати їх у структурі онтології. Але слід зважати на те, що досить часто такі експертні знання стосуються лише окремих аспектів задачі, і тоді використання запропонованого алгоритму забезпечить урахування й інших елементів знань щодо ПрО.

Наступним кроком співставлення СЮ є визначення ієрархії критеріїв з K_{sem} , які відображають потреби користувача на даний момент. Це дозволить визначити відносну значущість кожного з критеріїв для побудови комплексної оцінки близькості між парами СЮ (в тому числі – між СЮ, що потрібно оцінити, та еталонним СЮ, що будується за описом користувача).

Генерація еталонної моделі СЮ

Генерація еталонної моделі СЮ полягає у тому, щоб за вимогами користувача побудувати формалізований опис цих вимог, який може розглядатися як створення такого екземпляру СЮ, де для кожного Ю вказано конкретні значення, набір значень або діапазон прийнятних значень, які є бажаними для користувача. Основою для побудови еталонного СЮ є співставлення природномовного опису задачі з онтологією ПрО, на якій базується структура СЮ. В деяких випадках бажана ситуація, яка задовольняє користувача, може описуватися більш ніж одним СЮ (водночас інші поєднання таких СЮ будуть менш бажаними або взагалі незадовільними). Найпростіший приклад такої ситуації – для виконання певних будівельних робіт можна використовувати цвяхи й молоток або шурупи та шуруповерт, але не можна застосувати поєднання молотка й шурупів.

У такому випадку задача в постановці користувача трансформується на набір задач-варіантів для кожного з еталонних СЮ, і кожне зі знайдених рішень пропонується користувачеві для їх порівняння. Основою для цього може стати побудова тезаурусу задачі за природномовним описом та обраною онтологією, як це описано в [29].

Методи аналізу ієрархій

Побудова ієрархії, що відображає функціональні відношення між елементами системи, є важливою складовою аналізу її функціонування. Для простих систем з невеликою кількістю елементів така ієрархія може бути безпосередньо пов'язана із функціями системи. Але у складних системах досить складно виявити ієрархічну структуру, що відповідає її функціям. Моделюючи складну систему, потрібно певним чином розбити її на простіші підсистеми і аналізувати вже їх складові. Ці групування служили рівнями ієрархії. В даній роботі набір критеріїв для оцінювання СЮ може розглядатися як складна система, тому що набір, який будується за онтологією ПрО, може містити велику кількість взаємопов'язаних класів та їхніх характеристик. Але не всі ці критерії є значущими для даної задачі, а в різні моменти часу їхня важливість може різнитися. Тому виникає проблема визначення тих базових компонентів, котрі впливають на ухвалення рішень щодо оцінки СЮ, властивостей цих компонентів та відношень між ними.

ОПР зазвичай зіштовхується зі складною системою взаємозалежних компонентів (ресурси, бажані цілі, інтереси різних груп осіб тощо). Чим більше факторів враховується в такому аналізі, тим точніші будуть прогнози чи прийняті рішення. Але здатність людини до безпосереднього аналізу обмежена, і тому на практиці для складних задач система поділяється на підсистеми більш простих задач та наборів компонентів, що пов'язані ієрархічно. Дуже часто під час аналізу структури кількість елементів і їхніх взаємозв'язків настільки значна, що перевищує здатність дослідника сприймати інформацію в повному обсязі.

Одним із прикладів такого підходу є метод ієрархічного аналізу (МІА) [30] Сааті, в якому дослідження складних систем зводиться до послідовності попарних порівнянь відповідно до визначених компонентів. Ця теорія почала розвиватися з 1971 року як інструмент розв'язання проблем планування в непередбачених обставинах в сфері економіки, політики та воєнних задач. МІА є потужним засобом класифікації, що дозволяє експертам упорядковувати досвід, спостереження і факти. Однією з таких задач є задача багатокритеріального вибору (ухвалення рішення) в тій ПрО, яку описує онтологія. МІА дозволяє визначити ієрархію окремих критеріїв ухвалення рішення, які порівнюються попарно одним або кількома експертами ПрО. Рішення – це екземпляр класу онтології Р, для якого визначено набір семантичних властивостей C_1, C_2, \dots, C_n , що визначаються в цій онтології. Це можуть бути як об'єктні властивості (зі значеннями з екземплярів інших класів), так і властивості даних (з кількісними, якісними значеннями-константами різних типів). Ці поняття онтології ПрО розглядаються як критерії для оцінювання відповідності екземплярів класу Р до визначеної цілі. Відповідно до цього значення всіх критеріїв мають оцінюватися в діапазоні від “зовсім неприйнятне” до “дуже вдале”.

Ухвалення рішення полягає у виборі найбільш придатного рішення серед обмеженого непорожнього набору Х екземплярів класу Р: $X \subseteq P$. Важливо, що сам набір Х створюється за зовнішніми критеріями, що знаходяться поза сферою розгляду даного підходу (приміром, за ціною придбання, наявністю доступу або часовим інтервалом використання). Потрібно враховувати, що у більшості практичних задач Х складає дуже малу підмножину від Р, і тому не виникає потреба у пошуку узагальнених правил класифікації екземплярів Х (на категорії “зовсім неприйнятне”, “прийнятно” “дуже вдале” тощо), як це виконується у різноманітних методах машинного навчання (МН). Замість цього розв'язується задача зі значно меншою обчислювальною складністю, яка замість індуктивного правила “від окремого – до загального” будує традиційне правило “від окремого – до окремого”. На відміну від МН, де екземпляри Р класифікуються за довільним набором категорій, між якими

можуть бути відсутні формалізовані семантичні зв'язки (наприклад, рекомендована посада або ймовірний діагноз), у МІА екземпляри X належать до впорядкованого набору категорій з фіксованою кількістю значень.

Таблиця 1. Порівняння основних характеристик МІА та МН

	МІА	МН
Об'єкт аналізу	Властивості ІО ПрО та їхні значення	Властивості ІО ПрО та їхні значення
Дані для аналізу	Експертні оцінки властивостей	Дані про екземпляри ІО ПрО певного класу та їх характеристики
Результат аналізу	Ієрархія значущості властивості об'єктів ПрО	Класифікаційне правило для екземплярів ІО ПрО обраного класу
Подання результатів аналізу	Вектор ваг розмірності n	Дерево рішень, лінійний або нелінійний вираз, нейронна мережа
Область застосування	Впорядкування фіксованої підмножини екземплярів ІО ПрО обраного класу за їхньою відповідністю цілі	Віднесення всіх екземплярів об'єктів ПрО обраного класу до однієї з категорій
Обчислювальна складність	Залежить від розмірності C та обсягу X	Залежить від розмірності C та обсягу P
Складність простору рішень	Впорядкована множина оцінок щодо фіксованого класу рішень	Довільний набір категорій, що не перетинаються, до якого можуть бути віднесені екземпляри класу рішень
Роль онтології ПрО	Джерело C та структури класу P	Джерело C та структури класу P , їхніх характеристик

Основною задачею в побудові ієрархії є оцінка вищих рівнів, виходячи із взаємодії різних рівнів ієрархії, а не тільки з безпосередньої залежності між елементами на цих рівнях. Точні методи побудови ієрархічних з'являються в природних і суспільних науках, і особливо в задачах загальної теорії систем, пов'язаних із плануванням і побудовою соціальних систем.

Кожен елемент ієрархії функціонально може належати до декількох інших різних ієрархій. Наприклад, та сама людина може розглядатися разом з іншими співробітниками організації, учасниками навчання або пацієнтами медичного закладу. Елемент може бути на певному рівні однієї ієрархії керуючим компонентом, але в іншій ієрархії бути просто елементом, розкриваючим функції нижнього чи вищого порядку.

Ієрархія може розглядатися як модель реальної ситуації. Вона відображає аналіз найбільш важливих елементів і їх взаємовідношень, однак вона – не досить потужний засіб у процесі ухвалення рішень (наприклад, порівняно з онтологіями). Але саме розробка ієрархічної підсистем онтології може підтримуватися засобами МІА. Необхідний метод визначення сили, з яким різні елементи одного рівня впливають на елементи попереднього рівня, щоб можна було обчислювати величину впливів елементів найнижчого рівня на загальну ціль.

Важливою складовою МІА є перехід від кількісних та символічних значень до якісних оцінок. Слід зазначити, що, попри використання стандартизованих значень, для різних застосованих задач однієї ПрО функції перетворення кількісних та символічних значень на якісні можуть значно відрізнятися. Такі функції визначаються семантикою самої задачі та специфікою, пов'язаною саме зі знаннями ПрО. Наприклад, компетенція тієї самої особи в певній галузі може оцінюватися як висока для студента, але задовільна для викладача і незадовільна для експерта.

Припустимо, що n видів дії чи об'єктів розглядаються групою експертів. Тоді група має наступні цілі: 1) висловити твердження щодо відносної важливості цих об'єктів; 2) гарантувати такий процес отримання тверджень, який дозволить кількісно інтерпретувати ці твердження для всіх об'єктів.

Очевидно, що для досягнення другої цілі буде потрібно розробити відповідний метод для одержання з кількісних суджень групи (тобто з відносних величин, пов'язаних із парами об'єктів) множини ваг, що асоціюються з окремими об'єктами.

Завдяки такому підходу інформація перетворюється в зручну для обробки кількісну форму без інформаційних втрат, властивих якісним судженням.

Після отримання кількісних оцінок для всіх пар (C_i, C_j) задача зводиться до того, щоб n можливим діям поставити у відповідність множину з n кількісних оцінок w_1, w_2, \dots, w_n . Для цього потрібно більш чітко формалізувати саму задачу в термінах абстрактної математичної структури. В даній задачі бажано чітко визначити основні етапи процесу її формулювання і якомога докладніше описати кожен етап, щоб потенційний користувач міг оцінити доцільність його застосування до певної практичної проблеми. Потрібно чітко визначити, як саме залежать ваги $w_i, i = 1, n$ від суджень $a_{ij}, i, j = 1, n$.

Отримані в результаті застосування МІА до набору СІО кількісні оцінки дозволяють порівняти якість кожного з них, враховуючи відносну важливість кожного з критеріїв, визначених на попередньому етапі.

Висновки та перспективи подальшої роботи

Співставлення СЮ, що мають подібну структуру, є необхідною складовою в аналізі та порівнянні СЮ, побудовані на основі різних онтологій та мають різну структуру. В цьому випадку спочатку потрібно: 1. виконати вирівнювання базових онтологій Про й знайти відповідності між їх концептами та відношеннями; 2. потім знайти в СЮ, що співставляються, підмножини з подібною структурою; 3. співставити такі підмножини за розглянутим вище алгоритмом.

Було виявлено, що семантична подібність і семантична кореляція є невід'ємними компонентами аналізу СЮ, до того ж семантична подібність охоплює три складові: 1) подібність на основі атрибутів, де основою є інформація щодо атрибутів двох концептів; 2) подібність на основі контенту, що стосується можливості заміни двох концептів без втрати сенсу їх значень; 3) подібність заснована на відстані, де, чим менше семантична відстань між двома концептами, тим ближче буде їхня семантична подібність, і навпаки. Саму ж семантичну кореляцію можна віднести до комбінаторних відношень між двома концептами.

Дослідження методів семантичної подібності та близькості мають обмеження щодо класів та атрибутів в моделі онтології, оскільки не в змозі охопити всю інформацію щодо понять. Тож розглянуті концепти також обмежені, оскільки розроблення практичної онтологічної моделі та системи виведення висновку є тривалим і складним процесом, а реалізація на практиці вимагає багаторазового модифікування та оцінювання. Тому необхідні подальші інтегровані дослідження для покращення онтологічної моделі та лексичної бази даних і алгоритмів обчислення подібності з використанням інтелектуального аналізу даних, машинного навчання і статистичних методів обробки даних, щоб забезпечити ще точніші оцінки подібності.

Використання таких метрик дозволяє співставляти екземпляри СЮ, враховуючи на кожен поточний момент відносну важливість (ієрархію) тих критеріїв, що є значущими з точки зору задачі користувача.

Передбачається, що запропонований підхід може бути інтегровано із методами семантичної обробки метаданих Big Data з метою пошуку в них інформації щодо потрібних користувачам СЮ [31].

Література

1. Pryima S. M., Strokana O. V., Rogushina J. V. et al. Ontological Analysis of Outcomes of Non-formal and Informal Learning for Agro-Advisory System AdvisOnt. International Conference on Technologies and Innovation, Springer, Cham. 2020, P. 3-17. Available from: https://link.springer.com/chapter/10.1007/978-3-030-62015-8_1. [Accessed 12/07/2022].
2. Kozyrkov C. Introduction to Decision Intelligence, 2019. Available from: <https://towardsdatascience.com/introduction-to-decision-intelligence-5d147ddab767> [Accessed 12/07/2022].
3. Анісімов А.В., Марченко О.О., Никоненко А.О., Алгоритмічна модель асоціативно-семантичного контекстного аналізу текстів природною мовою Проблеми програмування. 2008., № 2-3, С.379-384. Available from: <https://core.ac.uk/reader/38468700>. [Accessed 12/07/2022].
4. Крюков К.В., Панкова Л.А., Пронина В.А. та інші Меры семантической близости в онтологиях. Труды научных сессий МИФИ-2010. Информационно-телекоммуникационные системы. Проблемы информационной безопасности, том. 5. 2010. С. 75-78.
5. Lastra-Díaz J. J., Goikoetxea J., Taieb M.A.H. et al. A reproducible survey on word embeddings and ontology-based methods for word similarity: linear combinations outperform the state of the art. Engineering Applications of Artificial Intelligence, 85, 2019, P.645-665. Available from: <https://doi.org/10.1016/j.engappai.2019.07.010>. [Accessed 11/07/2022].
6. Lastra-Díaz J., García-Serrano A. A new family of information content models with an experimental survey on WordNet. Knowledge-Based Systems, 89, 2015, P.509-526.
7. Camacho-Collados J., Pilehvar M. T. From word to sense embeddings: A survey on vector representations of meaning. Journal of Artificial Intelligence Research 63, 2018, P.743-788.
8. Taieb M. A. H., Zesch T., Aouicha M. B. A survey of semantic relatedness evaluation datasets and procedures. Artificial Intelligence Review 53(6). 2019. P. 4407-4448. Available from: <https://doi.org/10.1007/s10462-019-09796-3>. [Accessed 11/07/2022].
9. Altašnel B., Ganiz, M. C. 2018. Semantic text classification: A survey of past and recent advances. Information Processing & Management 54, 6 (2018), P.1129 - 1153. <https://doi.org/10.1016/j.ipm.2018.08.001>. [Accessed 4/07/2022].
10. Gan M., Dou X., Jiang R. From Ontology to Semantic Similarity: Calculation of Ontology-Based Semantic Similarity. Sci. World J. 2013, P.1-11.
11. Schickel-Zuber V., Faltings B. OSS: A semantic similarity function based on hierarchical ontologies. Proc.of the IJCAI-07, India, 2007, P. 551-556.
12. Maedche A., Staab S. Measuring similarity between ontologies. Proc.of the International Conference on Knowledge Engineering and Knowledge Management, 2002, P. 251-263.
13. Cunningham P. A Taxonomy of Similarity Mechanisms for Case-Based Reasoning. IEEE Trans. Knowl. Data Eng. 2008, 21, P.1532-1543.
14. Liao T.W., Zhang Z., Mount C.R. Similarity measures for retrieval in case-based reasoning systems. Appl. Artif. Intell. 1998, 12, P.267-288.
15. Lin D. An Information-Theoretic Definition of Similarity. Proc.of the International Conference on Machine Learning, USA, 1998.
16. Resnik P. Using information content to evaluate semantic similarity in a taxonomy. Proc. of the 14th international joint conference on Artificial intelligence, V. 1. 1995, P.448-453.
17. Yuhua L., Bandar Z.A., McLean D. An approach for measuring semantic similarity between words using multiple information sources. IEEE Transactions on knowledge and data engineering, 15, 4, 2003, P.871-882.
18. Sanchez D., Batet M., Isern D. et al. Ontology-based semantic similarity: A new feature-based approach. Expert Systems with Applications 39, 9, 2012, P.7718 - 7728. Available from: <https://doi.org/10.1016/j.eswa.2012.01.082>. [Accessed 4/07/2022].
19. Rada R., Mili H., Bicknell E. et al Development and application of a metric on semantic nets. IEEE transactions on systems, man, and cybernetics 19, 1, 1989, P.17-30.
20. Wu Z., Palmer M. . Verbs semantics and lexical selection. Proc. of the 32nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 1994, P.133-138.
21. Li Y., Bandar Z. A., McLean D. An approach for measuring semantic similarity between words using multiple information sources. IEEE Transactions on knowledge and data engineering 15, 4, 2003, P.871-882.
22. Banerjee S. Pedersen T. Extended gloss overlaps as a measure of semantic relatedness. IJCAI, Vol. 3. 2003, P.805-810.
23. Jiang Y., Zhang X., Tang Y. et al. Feature-based approaches to semantic similarity assessment of concepts using Wikipedia. Information Processing & Management 51, 3, 2015, P.215-234.
24. Zhu G., Iglesias C. A.. Computing Semantic Similarity of Concepts in Knowledge Graphs. IEEE Transactions on Knowledge and Data Engineering 29, 2017, P.72-85. Available from: <https://doi.org/10.1109/TKDE.2016.2610428>.

25. Jiang J., Conrath D. W. Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. Proc. of the 10th Research on Computational Linguistics International Conference. 1997, P.19–33.
26. Gao J.-B., Zhang B.-W., Chen X.-H. A WordNet-based semantic similarity measurement combining edge-counting and information content theory. Engineering Applications of Artificial Intelligence 39, 2015, P.80 – 88. Available from: <https://doi.org/10.1016/j.engappai.2014.11.009>. [Accessed 12/07/2022].
27. Рогущина Ю.В. Кількісні оцінки семантичної подібності між поняттями Про як засіб моделювання інтересів користувача. Прикладні системи та технології в інформаційному суспільстві, Київський нац. ун-т ім. Тараса Шевченка, 2020. С.183-187. (in Ukrainian).
28. Rogushina Y.V., Gladun A.Y. Development of domain thesaurus as a set of ontology concepts with use of semantic similarity and elements of combinatorial optimization. Проблеми програмування, 2021, № 2. С.4-15. Available from: <http://dspace.nbuv.gov.ua/handle/123456789/180661>. DOI: <https://doi.org/10.15407/pp2021.02.003>. [Accessed 4/07/2022].
29. Rogushina J., Gladun A. Task Thesaurus as a Tool for Modeling of User Information Needs, Chapter 7 in New Perspectives on Enterprise Decision-Making Applying Artificial Intelligence Techniques, 2021, Springer Verlag . 358 p. DOI:10.1007/978-3-030-71115-3_17.
30. Saaty T. L. How to make a decision: the analytic hierarchy process. European journal of operational research, 48(1), 1990, P.9-26.
31. Rogushina J., Gladun A. Semantic processing of metadata for Big Data: Standards, ontologies and typical information objects. CEUR Vol-2859, ITS 2020, Information Technologies and Security, 2020, P.114-128. Available from: <http://ceur-ws.org/Vol-2859/paper10.pdf>. [Accessed 25/07/2022].

References

1. PRYIMA, S. M. & STROKAN, O. V. & ROGUSHINA, J. V. et al. (2020). Ontological Analysis of Outcomes of Non-formal and Informal Learning for Agro-Advisory System AdvisOnt. International Conference on Technologies and Innovation, Springer, Cham. P. 3-17. Available from: https://link.springer.com/chapter/10.1007/978-3-030-62015-8_1. [Accessed 12/07/2022].
2. KOZYRKOV C. (2019). Introduction to Decision Intelligence. Available from: <https://towardsdatascience.com/introduction-to-decision-intelligence-5d147ddb6767> [Accessed 12/07/2022].
3. ANISIMOV A.V. & MARCHENKO O.O. & NIKONENKO A.O. (2008). Algorithmic model of associative-semantic contextual analysis of texts in natural language. Problems in Programming. No. 2-3, P. 379-384. Available from: <https://core.ac.uk/reader/38468700>. [Accessed 12/07/2022]. (in Ukrainian).
4. KRYUKOV K.V. & PANKOVA L.A. & PRONINA V.A. and others (2010). Measures of semantic similarity in ontologies. Proc. of Scientific sessions MIFI-2010. Information and telecommunication systems. Problems of information security, vol. 5. S. 75-78. . (in Russian).
5. LASTRA-DÍAZ J. J. & GOIKOETXEA J. & TAIEB M.A.H. et al. (2019). A reproducible survey on word embeddings and ontology-based methods for word similarity: linear combinations outperform the state of the art. Engineering Applications of Artificial Intelligence, 85, P.645-665. Available from: <https://doi.org/10.1016/j.engappai.2019.07.010>. [Accessed 11/07/2022].
6. LASTRA-DÍAZ J. & GARCÍA-SERRANO (2015). A new family of information content models with an experimental survey on WordNet. Knowledge-Based Systems, 89, P.509–526.
7. CAMACHO-COLLADOS J. & PILEHVAR M. (2018). T. From word to sense embeddings: A survey on vector representations of meaning. Journal of Artificial Intelligence Research 63, P.743–788.
8. TAIEB M. A. H. & ZESCH T. & AOUICHA M. B. (2019). A survey of semantic relatedness evaluation datasets and procedures. Artificial Intelligence Review 53(6). P. 4407–4448. Available from: <https://doi.org/10.1007/s10462-019-09796-3>. [Accessed 11/07/2022].
9. Altašnel B. & Ganiz M. C. (2018). Semantic text classification: A survey of past and recent advances. Information Processing & Management 54, 6 (2018), P.1129 – 1153. <https://doi.org/10.1016/j.ipm.2018.08.001>. [Accessed 4/07/2022].
10. GAN M. & DOU X. & JIANG R. (2013). From Ontology to Semantic Similarity: Calculation of Ontology-Based Semantic Similarity. Sci. World J. , P.1–11.
11. SCHICKEL-ZUBER V. & FALTINGS B. (2007). OSS: A semantic similarity function based on hierarchical ontologies. Proc. of the IJCAI-07, India, P. 551–556.
12. MAEDCHE A. & STAAB S. (2002). Measuring similarity between ontologies. Proc. of the International Conference on Knowledge Engineering and Knowledge Management, P. 251–263.
13. CUNNINGHAM P. (2008). A Taxonomy of Similarity Mechanisms for Case-Based Reasoning. IEEE Trans. Knowl. Data Eng., 21, P.1532–1543.
14. LIAO T.W. & ZHANG Z. & MOUNT C.R. (1998). Similarity measures for retrieval in case-based reasoning systems. Appl. Artif. Intell., 12, P.267–288.
15. LIN D. (1998). An Information-Theoretic Definition of Similarity. Proc. of the International Conference on Machine Learning, USA, .
16. RESNIK P. (1995). Using information content to evaluate semantic similarity in a taxonomy. Proc. of the 14th international joint conference on Artificial intelligence, V. 1, P.448–453.
17. YUHUA L. & BANDAR Z.A. & MCLEAN D. (2003). An approach for measuring semantic similarity between words using multiple information sources. IEEE Transactions on knowledge and data engineering, 15, 4, P.871–882.
18. SANCHEZ D. & BATET M. & ISERN D. et al. (2012). Ontology-based semantic similarity: A new feature-based approach. Expert Systems with Applications 39, 9, P.7718 – 7728. Available from: <https://doi.org/10.1016/j.eswa.2012.01.082>. [Accessed 4/07/2022].
19. RADA R. & MILI H. & BICKNELL E. et al. (1989). Development and application of a metric on semantic nets. IEEE transactions on systems, man, and cybernetics 19, 1, , P.17–30.
20. WU Z. & PALMER M.. (1994). Verbs semantics and lexical selection. Proc. of the 32nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, P.133–138.
21. LI Y. & BANDAR Z. A. & MCLEAN D. An approach for measuring semantic similarity between words using multiple information sources. IEEE Transactions on knowledge and data engineering 15, 4, 2003, P.871–882.
22. BANERJEE S. & PEDERSEN T. (2003). Extended gloss overlaps as a measure of semantic relatedness. IJCAI, Vol. 3, P.805–810.
23. JANG Y. & ZHANG X. & TANG Y. et al. (2015). Feature-based approaches to semantic similarity assessment of concepts using Wikipedia. Information Processing & Management 51, 3, , P.215–234.
24. ZHU G., IGLESIAS C. A. (2017). Computing Semantic Similarity of Concepts in Knowledge Graphs. IEEE Transactions on Knowledge and Data Engineering 29, P.72–85. Available from: <https://doi.org/10.1109/TKDE.2016.2610428>.
25. JIANG J., CONRATH D. W. (1997). Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. Proc. of the 10th Research on Computational Linguistics International Conference , P.19–33.
26. GAO J.-B. & ZHANG B.-W. & CHEN X.-H. (2015). A WordNet-based semantic similarity measurement combining edge-counting and information content theory. Engineering Applications of Artificial Intelligence 39, P.80-88. Available from: <https://doi.org/10.1016/j.engappai.2014.11.009>. [Accessed 12/07/2022].
27. ROGUSHINA Y.V. (2020). Quantitative evaluations of semantic similarity between the domain concepts as a means of modeling user interests. Applied systems and technologies in the information society, Kyiv National University. University named after Taras Shevchenko, P.183-187.
28. ROGUSHINA Y.V. & GLADUN A.Y. (2021). Development of domain thesaurus as a set of ontology concepts with use of semantic similarity and elements of combinatorial optimization. Проблеми програмування, № 2. С.4-15. Available from: <http://dspace.nbuv.gov.ua/handle/123456789/180661>. DOI: <https://doi.org/10.15407/pp2021.02.003>. [Accessed 4/07/2022].
29. ROGUSHINA J., GLADUN A. (2021). Task Thesaurus as a Tool for Modeling of User Information Needs, Chapter 7 in New Perspectives on Enterprise Decision-Making Applying Artificial Intelligence Techniques, Springer Verlag . 358 p. DOI:10.1007/978-3-030-71115-3_17.

30. SAATY T. L. (1990). How to make a decision: the analytic hierarchy process. European journal of operational research, 48(1), P.9-26.
31. ROGUSHINA J. & GLADUN A. (2020). Semantic processing of metadata for Big Data: Standards, ontologies and typical information objects. CEUR Vol-2859, ITS 2020, Information Technologies and Security, 2020, P.114-128. Available from: <http://ceur-ws.org/Vol-2859/paper10.pdf>. [Accessed 25/07/2022].

Одержано 01.08.2022

Про авторів:

Рогущина Юлія Віталіївна,
кандидат фізико-математичних наук,
старший науковий співробітник
Інституту програмних систем НАН України,
публікації в українських виданнях – 207,
публікації в іноземних журналах – 61,
ORCID <http://orcid.org/0000-0001-7958-2557>.

Гладун Анатолій Ясонович,
кандидат технічних наук, старший науковий співробітник.
Кількість публікацій в українських виданнях – 185.
Кількість зарубіжних публікацій – 55.
<http://orcid.org/0000-0002-4133-8169>

Місце роботи авторів:

Інститут програмних систем НАН України,
03181, Київ-187,
проспект Академіка Глушкова, 40,
e-mail: ladamandraka2010@gmail.com,
066 550 1999.

Міжнародний науково-навчальний центр інформаційних технологій та систем
Національної Академії наук України і Міністерства науки і освіти України.
проспект Академіка Глушкова, 44, 03680 м.Київ, Україна.
Тел.: +38-044-502-63-66
E-mail: glanat@yahoo.com

Прізвища та ініціали авторів і назва доповіді англійською мовою:

Rogushina Ju. V., Gladun A. Ya.
The use of ontological knowledge for multi-criteria comparison
of complex information objects

Прізвища та ініціали авторів і назва доповіді українською мовою:

Рогущина Ю. В., Гладун А. Я.
Використання онтологічних знань для багатокритеріального
співставлення складних інформаційних об'єктів

Контакти для редактора: Рогущина Ю., старший науковий співробітник
Інституту програмних систем НАН України,
e-mail: ladamandraka2010@gmail.com, тел.: (38)(066) 5501999