

## ВИКОРИСТАННЯ ОНТОЛОГІЙ ДЛЯ РОЗШИРЕННЯ МОЖЛИВОСТЕЙ ПОШУКУ ТА НАВІГАЦІЇ В ОНЛАЙН-ВЕРСІЇ «ВЕЛИКОЇ УКРАЇНСЬКОЇ ЕНЦИКЛОПЕДІЇ»

Ціль роботи – аналіз засобів семантичного пошуку у Wiki-ресурсах та розробка напрямків розширення його можливостей на основі використання сучасних засобів керування знаннями. В статті аналізуються переваги, які забезпечує користувачам використання семантичної розмітки в інформаційних ресурсах Web для більш зручного та швидкого доступу до контенту. Розглядаються способи використання онтологій для пошуку та навігації у семантизованих Wiki-ресурсах. Наведено визначення Wiki-онтології, що формалізує відомості щодо структури бази знань інформаційного ресурсу (як звичайного, так і семантизованого), який базується на технології Wiki. Проаналізовано теоретичні засади визначення міри семантичної подібності між поняттями на основі онтологій та таксономій. Запропоновано методи побудови засобів навігації у Wiki-ресурсах, які базуються на знаходженні сторінок, що описують змістовно близькі поняття. Продемонстровано реалізацію запропонованого підходу в онлайн-версії «Великої української енциклопедії» (e-BUE), що забезпечує розширення функціоналу енциклопедичного видання шляхом надання засобів доступу до контенту на семантичному рівні. Обґрунтовано методологічні принципи організації бази знань e-BUE та категоризації її гасел, що забезпечують використання оцінок семантичної близькості між поняттями для пошуку подібних гасел.

Ключові слова: семантичні Wiki-ресурси, семантичний пошук, онтологія, семантична близькість, Велика українська енциклопедія.

### Вступ

Ефективність використання Web-ресурсів значним чином визначається ступенем їх інтелектуалізації – наявністю семантичної розмітки, можливістю пошуку за змістом, використанням онтологій для моделювання структури бази знань (БЗ) тощо. Це впливає на функціонал інформаційного ресурсу (ІР) та забезпечує можливість використання його контенту іншими інформаційними системами (ІС), що пов'язані з тією самою предметною областю (ПрО).

Значення семантизації збільшується для ІР із складною структурою, гетерогенним контентом та великим обсягом даних, що викликає потребу в застосуванні сучасних технологій керування розподіленими знаннями. Одна з таких поширених на сьогодні технологій є Wiki, яка забезпечує спільне створення та використання інформації у Web.

Інтелектуалізація ІР, що будується на основі Wiki-технології, має свою структуру. В першу чергу вона базується на

семантизації ресурсу, тобто на встановленні змістовних зв'язків між Wiki-сторінками. Такі можливості надає, наприклад, семантичне розширення Wiki – Semantic MediaWiki (SMW) [1], що підтримує використання як семантичних властивостей у розмітці, які визначають зміст відношень між сторінкою ІР та іншими сторінками та даними, так і виконання семантичних запитів, умовами яких можуть бути значення таких властивостей. Уся сукупність семантичних властивостей ІР та їх значень може бути формалізована за допомогою Wiki-онтології ресурсу.

Основні характеристики онлайн-версії «Великої української енциклопедії» (e-BUE [ue.gov.ua](http://ue.gov.ua)) дозволяють віднести її саме до таких ІР: ця енциклопедія містить велику кількість гасел з різних галузей знань з різноманітними наборами структурних елементів та мультимедійного ілюстративного матеріалу (рисунки, фотографії, відео, аудіо, карти, діаграми), між якими встановлюються різноманітні відношення.

Використання елементів онтологічного аналізу для створення структури БЗ е-ВУЕ обумовлено високою складністю та гетерогенністю знань, представлених у цьому інформаційному ресурсі: виразна здатність вбудованих можливостей що використовується як основа для створення порталу і підтримує такі механізми, як категорії, семантичні властивості та семантичні запити, достатня для реалізації такої системи знань, але не для її дослідження, вдосконалення та формалізації всіх різноманітних аспектів та характеристик.

Тому цей IP використовується в даній роботі для демонстрації можливостей розширення пошуку та навігації з використанням онтологічних знань. Онтологія ПрО може використовуватися як основа для знаходження семантично пов'язаних гасел е-ВУЕ, що надає ще один засіб навігації в IP: користувач отримує можливість перейти до тих понять ПрО, які змістовно пов'язані з тим гаслом, яке він розглядає.

Щоб автоматизувати використання розподіленого IP, виникає необхідність у побудові онтології його БЗ – формалізованої моделі, що дозволяє фіксувати характеристики його елементів, їх зв'язків, властивостей та відношень у формі, придатній для автоматичного оброблення, логічного виведення та аналізу. Така онтологія може бути основою для семантичної розмітки IP, яка уможлиблює інтелектуалізацію засобів навігації та пошуку відомостей на порталі, задовольняючи інформаційні потреби користувачів із урахуванням їх семантики.

Одним з напрямків використання Wiki-онтології IP є знаходження семантично близьких Wiki-сторінок: це дозволяє створити у ресурсі додаткові засоби навігації для швидкого доступу до сторінок зі спорідненим контентом. При цьому семантична близькість може визначатися відповідно до поточних потреб користувача.

### Постановка задачі

Ціль даної роботи – аналіз засобів семантичного пошуку у Wiki-ресурсах та напрямків розширення його можливостей

на основі використання онтологій. Надається теоретичне обґрунтування методологічних принципів використання міри семантичної близькості між поняттями онтології, що формалізує семантичну розмітку е-ВУЕ, для розширення функціоналу цього енциклопедичного видання шляхом створення засобів навігації між сторінками на змістовному рівні.

### Онтологічне подання знань

Онтології – БЗ спеціального виду, яка містить семантичну інформацію [2] з певної ПрО – застосовуються для подання знань різноманітних ПрО в Web-орієнтованих ІС [3]. Моделі онтологій зазвичай містять *поняття* (класи та екземпляри), *властивості* понять (атрибути, ролі), *відношення* між поняттями (залежності, функції) та додаткові *обмеження*, що визначаються аксіомами [4].

На формальному рівні онтологія – це система, що складається з наборів понять і тверджень про ці поняття, на основі яких можна будувати класи, об'єкти, відношення, функції та теорії.

Формальна модель онтології  $O$  – це впорядкована четвірка

$$O = \langle T, P, R, F \rangle,$$

де  $T$  – скінченна множина концептів (термінів) ПрО, яку описує онтологія  $O$ ;  $P$  – скінченна множина властивостей концептів;  $R$  – скінченна множина відношень між концептами заданої ПрО;  $F$  – скінченна множина аксіом, заданих на концептах, властивостях і/або відношеннях онтології  $O$ . Відношення представляють тип взаємодії між концептами ПрО. Приклад бінарного відношення – "є частиною". Аксіоми використовуються для моделювання тверджень, що завжди є істинними.

*Wiki-онтологія* – це окремих випадок комп'ютерної онтології, що відображає особливості технології Wiki та її семантичного розширення. Її виразні можливості обмежені в області використання аксіом ПрО та не припускають застосування характеристик для об'єктних властивостей та властивостей даних. Така онтологія

може бути побудована автоматизовано за семантично розміченим Wiki-ресурсом (набором Wiki сторінок, що містять семантичну розмітку) [5], але на практиці зазвичай спочатку розробляють Wiki-онтологію, а вже на її основі здійснюється реалізація самого Wiki-ресурсу [6].

Можна вважати (з певним рівнем спрощення), що Wiki-онтологія та структура БЗ Wiki-ресурсу є різними формами подання тієї самої системи знань. Така онтологія містить тільки ті знання, які можна безпосередньо здобути із семантичної розмітки. Тому в цій онтології відсутні, приміром, такі характеристики класів та властивостей, як еквівалентність, відсутність перетину тощо.

Wiki-онтологія конкретизує основні компоненти онтологічного подання знань  $O = \langle X, R, F, T \rangle$  відповідно до технологій Semantic Web [7], розроблених W3C.

Для не семантизованого Wiki-ресурсу відповідна онтологія  $O_{wiki\_no\_semant}$  містить наступні компоненти:

- $X = X_{cl} \cup X_{ind}$  – множина понять онтології, де  $X_{cl}$  – множина класів, що співпадає з множиною категорій Wiki-ресурсу  $P_{categ}$ ;  $X_{ind}$  – множина екземплярів класів – це об'єднання множин Wiki-сторінок:  $P_{user}$  – множина сторінок, створених користувачами,  $P_{template}$  – множина сторінок, що описують шаблони,  $P_{spec}$  – множина інших спеціальних сторінок;

- $R = L \cup \{r_{ier\_cl}\}$  – множина відношень між елементами онтології, де  $L = \{ "link" \}$  – множина з одного елемента, що описує посилання однієї Wiki-сторінки цього ресурсу на іншу Wiki-сторінку цього ресурсу (посилання на інші види сторінок у рамках даної моделі не враховуються);  $r_{ier\_cl}$  – ієрархічне відношення між категоріями Wiki-ресурсу, яке визначається в процесі створення нових категорій;

- $F = \{f_{equ}\}$  – множина тих характеристик, що можуть використовуватися

для логічного виведення над онтологією, для Wiki-онтології обмежується єдиним відношенням еквівалентності між Wiki-сторінками (відсильні статті);

- інші елементи онтологічної моделі для цієї Wiki-онтології є порожніми множинами.

Формальна модель семантично збагачених Wiki-ресурсів є більш складною і включає ряд елементів, пов'язаних із семантичними властивостями. Онтологія  $O_{wiki\_semant}$  містить більше компонентів:

- множина відношень доповнюється  $L_{sem\_prop}$ , елементи якого – семантичні властивості Wiki-сторінок, областю значення яких є Wiki-сторінки;

- $T$  – множина типів даних (наприклад, рядок, ціле), значення з яких можуть приймати властивості даних класів онтології, – значення семантичних властивостей, що пов'язують не з іншими сторінками, а зі значеннями даних.

Шаблони типових ІО дозволяють визначати область значення та область визначення об'єктних властивостей Wiki-онтології, та навпаки, характеристики об'єктних властивостей з Wiki-онтології є основою для коректного створення шаблонів типових ІО. Семантична розмітка Wiki-ресурсу дозволяє автоматизувати процес побудови його Wiki-онтології, і тому для ІР на основі Semantic MediaWiki досить просто застосовувати семантичні технології, що базуються на використанні онтологічних знань.

## Семантичний пошук на порталі e-VUE

В ефективності використання сучасних розподілених баз знань одним з визначних чинників є організація пошуку, що дозволяє користувачеві отримати доступ до потрібної інформації. При цьому важливі як час доступу, так і зрозумілість і зручність користувацького інтерфейсу. Знання, представлені інформаційними ресурсами, що організовані на основі Wiki-технологій, мають певну специфіку [8].

Організація подання та збереження інформації на таких ресурсах визначає засоби та методи пошуку. Відомості, що представлені за допомогою Wiki-сторінок, з точки зору інформаційно-пошукових систем відносяться до частково структурованих: вони групуються за допомогою категорій Wiki, зазвичай містять посилання на інші Wiki-сторінки того самого ресурсу, та, якщо застосовуються семантичні Wiki, включають семантичну розмітку, тобто властивості інформації, що представлена на сторінці, охарактеризовані на семантичному рівні.

На сторінці категорії у Wiki-ресурсах представлено перелік її підкатегорій та усіх сторінок, що відносяться до цієї категорії. Це забезпечує досить зручну навігацію у ресурсі за змістовно поєднаними у категорії групами сторінок, але не дозволяє отримувати доступ до інформації, що відноситься до групи категорій.

Засоби пошуку у Wiki-ресурсах можна класифікувати на основні наступні групи:

- за *ключовими словами* (для енциклопедій найчастіше – за гаслом статті або за початковими літерами назви);
- за *тематикою предметної області* (для енциклопедій – за категоріями та підкатегоріями статті);
- за *типом інформаційного об'єкта* (для енциклопедій типи ІО теж формалізуються через апарат категорій, конкретні значення яких здобуваються із узагальнення типів статей енциклопедії);
- за *семантикою інформаційного об'єкта* (за його властивостями).

Перший варіант забезпечує найбільш швидкий доступ до інформації, другий базується на загальноновживаних класифікаторах (що дозволяє формалізувати класифікацію, але ускладнює пошук для пересічних користувачів), а третій – враховує семантику інформаційної потреби користувача та дозволяє ознайомитися із семантично близькими статтями.

Пошук за ключовими словами – назвою гасла, словами в природномовному контенті – є традиційним, його реалізують

практично всі електронні довідкові та енциклопедійні видання, але він потребує від користувача знань щодо точного найменування того, що він шукає. Пошук термінів, що вводяться користувачем з помилками, зазвичай не підтримується. Крім того, для такого пошуку має значення порядок слів у пошуковому рядку.

Пошук за тематикою є більш гнучким. Для е-ВУЕ найбільш вживаним є пошук за галузями знань та їх підкласами. Крім того, вводиться багато інших специфічних для окремих галузей категорій, до яких можна легко перейти з поточного гасла. Це забезпечує зручну навігацію для тих користувачів е-ВУЕ, що досліджують певну проблему, а не шукають визначення конкретного терміну.

Третій тип пошуку – за типом інформаційного об'єкта – підтримується в е-ВУЕ за допомогою великої кількості шаблонів для типових ІО. Кожен шаблон пов'язаний із відповідною категорією е-ВУЕ, і користувач може застосовувати ці категорії і їх ієрархію для навігації у ресурсі.

У багатьох випадках для користувачів найзручнішим виявляється пошук за типом інформаційного об'єкта. *Інформаційний об'єкт* (ІО) – це інформаційна модель об'єкта певної предметної області, що визначає структуру, атрибути, обмеження цілісності і, можливо, поведінку цього об'єкта.

Четвертий тип пошуку – це *семантичний пошук*, результатом якого має стати інформаційний об'єкт зі складною структурою, знання про яку використовуються в пошукових процедурах. Його можна розглядати як окремих випадок проблеми розпізнавання об'єктів, якщо у випадку енциклопедій повна реалізація потребує встановлення семантичних плагінів, призначених для оперування із семантичними властивостями статей.

У різних Wiki-ресурсах реалізуються різні підмножини видів пошуку. Це залежить від обсягу Wiki-ресурсу, складності його структури та від того, на яких саме користувачів орієнтовано цей ресурс.

Для типових ІО доцільно виділити найбільш характерні параметри, які можна

відображати за допомогою семантичних властивостей Wiki. Для уніфікованого представлення таких властивостей доцільно розробити відповідні шаблони. Wiki-онтологія дозволяє формалізувати ці параметри та їх властивості, зв'язати їх з категоріями Wiki-ресурсу, що відповідають цим IO. Це дозволяє будувати семантичні запити для пошуку IO за їх властивостями та категоріями.

Саме наявність онтологічної моделі БЗ e-BUE – основа для виконання семантичного пошуку: Wiki-онтологія містить знання про семантичні властивості різних типових IO – їх назви, призначення, можливі значення тощо. У семантичному пошуку під IO будемо розуміти ту інформацію, яку користувач отримує в результаті виконання процесу пошуку. Стосовно Wiki-ресурсів, як окремі IO доцільно розглядати Wiki-сторінки як звичайні, так і спеціальні (приміром, сторінки категорій), які містять однаковий набір семантичних властивостей та відносяться до тих самих категорій.

Найчастіше використовуються запити, що сполучені з функцією *ask*. Ця функція використовується так само як і інші функції синтаксичного аналізатора MediaWiki: її виклик позначається подвійними фігурними дужками, перед іменем ставиться символ #, а після – двокрапка. Спочатку передається сам рядок запиту, що відбирає потрібну інформацію з Wiki, а потім усі параметри запиту, розділені символами вертикальної риси |. В умовах запиту можуть використовуватися назви категорій та семантичних властивостей, а також обмеження щодо значень цих властивостей. Крім того, можуть використовуватися деякі службові параметри, такі як поточний час або дата та назва поточної сторінки. Запити *ask* мають таку форму:

```
{{#ask:Критерії вибору сторінок
|?Запитувана властивість
|параметри
}}
```

Якщо користувачів цікавлять не стільки назви сторінок, що відповідають умовам запитів, але й інформація, яка

представлена на цих сторінках, тоді потрібно отримати значення семантичних властивостей цих сторінок. Наприклад:

```
{{#ask:
[[Категорія:Країна]]
[[Населення::>2000000]]
|?Столиця
|format=broadtable
}}
```

Слід звернути увагу, що у запиті необхідно використовувати точні назви категорій та семантичних властивостей. Саме для цього може бути використана Wiki-онтологія IP, що інтероперабельно формалізує ці відомості і дозволяє бачити зв'язки між властивостями та категоріями, їх характеристики тощо, які явно не подаються у середовищі Semantic Media-Wiki.

За допомогою семантичного пошуку користувач може отримувати доступ до інформації з різних Wiki сторінок e-BUE, а застосування значень семантичних властивостей як параметрів запиту дозволяє автоматизувати відбір саме тієї інформації, що потрібна йому. Це дозволяє економити час та зусилля користувачів, підвищує ефективність використання енциклопедії та забезпечує повніше задоволення персональних інформаційних потреб різних груп користувачів.

Семантичні запити можуть вбудовуватися до існуючих сторінок і розширювати можливості навігації (рис. 1) або викликатися за допомогою шаблонів. Це забезпечує можливість інтеграції відомостей з різних гасел, дозволяє запобігти повторному введенню інформації та гарантує автоматизоване внесення змін до усіх результатів запиту у випадку змін у тих сторінках, серед яких виконується пошук.

Наприклад, для сторінок категорій, що відповідають галузям знань e-BUE та їх підкатегоріям, за допомогою запитів, що вбудовані у шаблони, виконується пошук нових гасел відповідної категорії та гасел, для яких потрібні автори, виводиться інформація про модератора галузі тощо. У шаблоні “Нові\_сторінки\_категорії” міститься наступний запит:

**Технічні науки**

Це категорія Технічні науки.  
Кількість статей у цій категорії: 300.  
Технічні науки — сукупність наук, що досліджують техніку і вивчають явища, пов'язані з її створенням, розвитком і взаємодією з природою та людиною.

**Нові статті**

В цьому місяці опубліковано такі статті з цієї категорії:

**А**

- Алмазний інструмент (15 02 2019)
- Алідада (02 10 2019)

**Автори категорії**

До цієї категорії писали статті такі автори:

<b>А</b>	<b>А (прод.)</b>	<b>А (прод.)</b>
• АН (Снігир Г. В.)	• Автоколивальна система (Смирнов В. П.)	• Агломерат (металургія) (Вергельська Н. В.)
• Абдуллаєв, Гасан Мамедбагир огли (Храмов Ю. А.)	• Автомагістраль (Савенко В. Я.)	• Агломерація матеріалу (Нестеров О. С.)

Рис. 1. Сторінка категорії е-ВУЕ “Технічні науки”, контент якої будується в результаті виконання семантичних запитів

```

{{#ask:
[[Категорія:{{PAGENAME}}]]
[[Дата оприлюднення::>0]]
[[Місяць оприлюднення:{{CURRENTMONTH}}]]
[[Рік оприлюднення:{{CURRENTYEAR}}]]
|?Дата оприлюднення
|?Місяць оприлюднення
|?Рік оприлюднення
|format=category
|link=all
|order=descending
|sep=.
|intro=В цьому місяці опубліковано
|outro=
|default=В цьому місяці не оприлюднено нові
статті.
}}

```

Якщо з'являються нові гасла категорії або в таких гаслах інформація змінюється, контент сторінки категорії оновлюється автоматично.

Такі шаблони та вбудовані запити доцільно створювати лише для тих запитів, які мають виконуватися регулярно та можуть зацікавити велику кількість користувачів. Якщо інформаційні потреби користувача більш специфічні, йому потрібно побудувати семантичний запит самостійно.

Середовище Semantic MediaWiki підтримує просту, але досить потужну мову запитів *SMW-QL* для семантичного пошуку у Wiki-ресурсах, яка дозволяє, по-перше фільтрувати сторінки за заданими критеріями, і по-друге, виводити як результати запиту тільки цікавлячого користувача інформацію, а не весь текст Wiki-сторінки.

Результати таких запитів динамічно інтегрують актуальну інформацію з інших Wiki-сторінок і відображають її у зручному для користувача форматі – у вигляді діаграм, географічних карт, таблиць і схем

тощо. Якщо контент сторінок, з яких здобуваються потрібні дані, змінюється, то результати запитів також автоматично оновлюються, забезпечуючи цим несуперечність і погодженість даних. Це дуже зручно для IP зі складною структурою та розгалуженими зв'язками між сторінками.

На жаль, пошукова мова Semantic MediaWiki має певні обмеження:

- у семантичних запитах не обробляються ієрархічні відношення між категоріями, наприклад, немає засобів пошуку підкатегорій обраної категорії;

- категорії не можуть використовуватися як змінна у запиті (можна вводити тільки фіксоване значення), наприклад, неможливо знайти Wiki-сторінки, що відносяться до тих категорій, що й поточна сторінка;

- категорії Wiki-сторінок, що відповідають умовам запиту, не можуть виводитися в його результатах.

Це обмежує сферу застосування семантичних запитів та ускладнює виконання пошукових процедур, що пов'язані з поточною сторінкою, яку проглядає користувач. Тому виникає потреба у створенні додаткових засобів пошуку та навігації в IP, що базуються на Semantic MediaWiki, які дозволяють усунути ці обмеження. Але їх застосування потребує певного теоретичного обґрунтування.

### Методи аналізу подібності понять

Для аналізу семантичної близькості між поняттями ПрО можуть використовуватися методи, що спрямовані на знаходження подібності між елементами різних онтологій. Оцінка подібності понять може базуватися на їх позиціях в ієрархії класів, для яких вже визначено подібність: якщо підкласи та надклас цих понять подібні, то самі такі поняття теж можуть бути подібними. Для кількісної оцінки подібності двох сутностей можуть враховуватися наступні параметри (ознаки):

- оцінка подібності їх прямих надкласів;
- оцінка подібності всіх їх надкласів;
- оцінка подібності підкласів понять;
- оцінка подібності екземплярів понять.

Аналогічно будуються оцінки подібності з урахуванням інших (не ієрархічних) зв'язків між поняттями. Наприклад, якщо поняття  $p_{O_{12}}$  та  $p_{O_{12}}$  пов'язані відношенням  $r_{O_1}$  в онтології  $O_1$ ; поняття  $p_{O_{21}}$  та  $p_{O_{22}}$  пов'язані відношенням  $r_{O_2}$  в онтології  $O_2$ ; поняття  $p_{O_{12}}$  подібне до поняття  $p_{O_{22}}$  (оцінка подібності вище за порогову); відношення  $r_{O_1}$  подібне до відношення  $r_{O_2}$  (оцінка подібності вище за порогову), тоді можна припустити, що поняття  $p_{O_{11}}$  теж подібне до поняття  $p_{O_{21}}$  (оцінка подібності вище за порогову).

Для оцінки близькості понять може застосовуватися також статистичний аналіз, який базується на порівнянні наборів екземплярів двох класів: подібність понять визначається шляхом порівняння екстенціоналів цих понять. Знаходження відповідності між поняттями базується на наступному емпіричному правилі:

- класи  $p_1$  та  $p_2$  онтології  $O_1$  *еквівалентні*, якщо для всіх екземплярів онтології  $P_i, \forall i = \overline{1, n}$  виконується  $P_i \in p_1 \Rightarrow P_i \in p_2$  і, навпаки,  $P_i \in p_2 \Rightarrow P_i \in p_1$ ;

- клас  $p_1$  онтології  $O_1$  є *підкласом*  $p_2$  цієї онтології, якщо для всіх екземплярів онтології  $P_i, \forall i = \overline{1, n}$  виконується  $P_i \in p_1 \Rightarrow P_i \in p_2$ .

Для Wiki-ресурсів, що розробляються без формалізації структури БЗ (прикладом такого IP є Вікіпедія), необхідно застосовувати саме такі підходи до знаходження СБП. Складність проблеми співставлення понять в онтології пов'язана з її поганим масштабуванням: збільшення кількості понять в онтології та ускладнення її структури значно збільшують простір пошуку. Тому в IP великого обсягу такі методи потребують значних обчислювальних ресурсів.

Якщо потрібно визначати подібність понять в одній онтології, то задача значно спрощується. Тому для тих IP, що мають заздалегідь розроблену систему класів та відношень (прикладом такого IP

є e-BUE), що може бути формалізована як Wiki-онтологія, доцільно застосовувати методи знаходження СБП, що базуються на семантичних мережах.

Використання онтологій у пошуку змістовно близьких понять дозволяє застосовувати знання Про для більш інтелектуальної навігації в IP, надаючи прямі переходи до змістовно пов'язаних сторінок.

Оцінювання семантичної спорідненості понять з використанням семантичних мереж для подання знань має довгу історію [9]. Семантична подібність є особливим випадком семантичної спорідненості [10]. Деякі дослідники припускають, що оцінку подібності в семантичних мережах потрібно розглядати із залученням лише таксономічних зв'язків [11], виключаючи інші типи зв'язків; але зв'язки між частинами також можна розглядати як атрибути, які впливають на визначення подібності [12]. В літературі визначено багато критеріїв подібності, але вони рідко супроводжуються незалежною характеристикою явища, яке вони вимірюють: їх цінність полягає в корисності для конкретного завдання.

Для окремого випадку онтології, де застосовується єдине відношення між поняттями – ієрархічне відношення типу IS-A, – таксономії – близькість двох термінів може бути оцінена за відстанню між поняттями в таксономії вершин, які відповідають цим термінам, в таксономії – окремому випадку онтології.

Семантична відстань між поняттями залежить від довжини найкоротшого шляху між вершинами та загальної специфічності двох вершин. Чим коротший шлях від одного вузла до іншого, тим більше вони подібні. Якщо між елементами існує кілька шляхів, використовують довжину найкоротшого з них [13, 14]. Довжина найкоротшого шляху в цій таксономії між відповідними поняттями, який визначається кількістю вершин (або ребер) в найкоротшому шляху між двома відповідними вершинами таксономії [11], з урахуванням глибини таксономічної ієрархії [15] (чим менше довжина шляху між вершинами, тим вони семантично ближчі):

$$S(a, b) = \log 2N/d(a, b),$$

де  $N$  – глибина таксономічного дерева,  $d(a, b)$  – довжина найкоротшого шляху між вершинами  $a$  та  $b$ .

Однак з цим підходом пов'язана складна проблема, спричинена уявленням про те, що всі зв'язки в таксономії являють собою однорідні відстані. На жаль, рівномірну відстань у таксономії важко визначити і ще важче контролювати.

У реальних таксономіях існує велика мінливість «відстані», що охоплюється єдиним таксономічним зв'язком, особливо коли деякі підмножини таксономії (наприклад, біологічні категорії) є набагато більш щільними, ніж інші. Наприклад, у WordNet [16] – широко відомій семантичній мережі для англійської мови – можна знайти прямі зв'язки як між досить подібними поняттями, так і між відносно далекими. Тому доцільно враховувати семантику відношень між поняттями для різних таксономічних відношень та враховувати кількість екземплярів у підкласах.

Деякі міри близькості [17] враховують тільки глибину вершин термінів:

$$S(a, b) = 2 \times N(p(a, b)) / (N(a) + N(b)),$$

де  $p(a, b)$  – найближчий спільний предок вершин  $a$  та  $b$ ,  $N(x)$  – глибина вершини  $x$ .

Подібність понять пов'язана також з їх інформаційним змістом (контентом). Нехай  $S$  – множина понять в таксономії  $is-a$ , що дозволяє множинне успадкування (multiple inheritance).

Одним із ключових факторів у подібності двох понять є ступінь, в якій вони поділяють інформацію, зазначену в таксономії IS-A високо специфічним поняттям, яке відноситься до обох цих понять. Метод підрахунку країв (edge-counting method) враховує це опосередковано, оскільки, якщо мінімальний шлях зв'язків IS-A між двома вузлами довгий, то це означає, що необхідно високо підійматися в таксономії до більш абстрактних понять, щоб знайти найменшу верхню межу – поняття, до якого відносяться обидва поняття, що аналізуються.



Інформаційний контент (*information content*) поняття с можна кількісно визначити як  $-\log p(c)$ : чим вище ймовірність використання поняття, тим нижче його інформативність. Таким чином, чим вище рівень абстракції поняття (тобто чим вище воно знаходиться в таксономії), тим менше його інформаційний контент. Якщо в таксономії існує унікальна верхня концепція, то її інформаційний контент дорівнює 0.

Така кількісна характеристика інформації забезпечує новий спосіб вимірювання семантичної подібності [18] на основі екстенсіоналу понять.

Чим більше інформації поділяють (сумісно використовують) два поняття, тим більше вони подібні, а інформація, що сумісно використовується двома поняттями, визначається інформаційним контентом понять, що входять до їх складу в таксономії. Формально така семантична подібність визначається наступним чином:

$$\text{sim}(c_1, c_2) = \max_{c \in S(c_1, c_2)} [-\log p(c)],$$

де  $S(c_1, c_2)$  – множина понять, що входять до складу як  $c_1$ , так і  $c_2$ .

Хоча подібність обчислюється з урахуванням всіх верхніх меж для двох понять, інформаційна міра дозволяє ідентифікувати мінімальну верхню межу, оскільки жоден клас не є менш інформативним, ніж його надкласи.

На практиці часто виникає потреба у вимірюванні подібності слів, а не понять. Використовуючи для представлення слів з множини  $W$  через сукупності понять в таксономії, які є значеннями (змістами) слова  $w$ , функцію  $s(w)$ , таку, що  $s: W \rightarrow C$ , тобто

$$s(w \in W) = \{c_k \in C, k = \overline{1, m}\},$$

можна визначити

$$\text{sim}_w(w_1, w_2) = \max \text{sim}(c_i, c_j),$$

де  $c_i \in s(w_1), c_j \in s(w_2)$ .

Це узгоджується з визначенням «диз'юнктивних понять», що використовує підрахунок країв: вони визначають відс-

тань між двома диз'юнктивними наборами понять як мінімальну довжину шляху від будь-якого елемента з першого набору до будь-якого елемента з другого. Подібність слів оцінюється через знаходження максимального інформаційного контенту над усіма поняттями, для яких обидві слова можуть бути екземпляром.

Це дозволяє створювати множини *семантично близьких слів* (СБС), тобто слів, семантична відстань між якими менша за обрану порогову величину.

*Семантично близькі поняття* (СБП) – це нечітка множина, яка включає набір понять, для яких кількісне значення семантичної близькості з обраним поняття вище заданого порогу [19]. Міри визначення семантичної близькості понять на основі онтологій використовують різноманітні семантичні характеристики цих понять – їх властивості (атрибути і відношення з іншими поняттями), взаємне положення в онтологічних ієрархіях.

В основу багатьох мір близькості на основі онтологій покладений теоретико-множинний підхід Тверської [20], який визначає міру близькості двох об'єктів шляхом співставлення властивостей (*feature matching*). Міра близькості  $S(a, b)$  між об'єктами  $a$  і  $b$  – це функція трьох наборів властивостей цих об'єктів  $A$  і  $B$ , їх перетину  $A \cap B$ , доповнень  $A - B$  та  $B - A$ .

Недоліком більшості мір, які основані на онтологічних структурах, є симетричність (експертні оцінки показують, що міра близькості не завжди симетрична), незалежність від контексту і чутливість до структури ієрархії. Цю проблему дозволяє вирішити застосування асиметричної міри семантичної близькості [21], яка залежить від напрямку руху між ребрами (наприклад, нащадок більш подібний до предка, ніж предок до нащадка).

В інших роботах пропонуються міри близькості, які базуються на неієрархічних («горизонтальних») відношеннях і атрибутах. Оцінювання близькості понять, яке використовує горизонтальні відношення, спирається на припущенні, що якщо два поняття мають одне й те саме відношення з третім поняттям, тоді вони ближче, ніж два поняття, які мають це відношення з різними поняттями, тобто

близькість двох понять залежить від близькості тих понять, з якими вони мають відношення. Така міра близькості обчислюється рекурсивно.

Атрибутивна міра близькості базується на близькості значень спільних атрибутів понять, діапазони яких є літералами, числами, рядками та іншими типами даних. Як міри близькості для рядкових даних можна використовувати нормовану редакторську відстань [22], для чисел – інверсію різниці, що нормована максимальним значенням атрибуту.

Використання цих мір близькості дозволяє оцінювати подібність значень параметрів понять (властивості даних екземплярів класів онтології), яким в семантизованих Wiki-ресурсах відповідають значення семантичних властивостей, що не є посиланнями на інші Wiki-сторінки.

Аналіз існуючих підходів до кількісного оцінювання семантичної подібності понять показує доцільність використання для цього таксономій та відстані в цих таксономіях між поняттями, близькість яких оцінюється, та їх спільним надкласом. Урахування інших типів онтологічних відношень між поняттями та порівняння їх семантичних властивостей дозволяють уточнювати ці оцінки відповідно до специфіки ПрО.

Розглянуті вище методи знаходження СБП можуть використовуватися в е-ВУЕ, де реалізується кілька незалежних таксономій для категоризації гасел, а інші зв'язки між гаслами встановлюються за допомогою семантичних відношень. Ця інформація щодо структури БЗ е-ВУЕ представлена у вигляді Wiki-онтології та може застосовуватися як користувачами енциклопедії, так і іншими інтелектуальними ІС.

### Структура БЗ е-ВУЕ

Основний інструмент організації БЗ порталної версії е-ВУЕ – засоби, що надаються Semantic MediaWiki (категорії та семантичні властивості). Для структурованого подання інформації використовується набір незалежних таксономій, з якими пов'язані відповідні набори кате-

горій, між якими встановлені об'єктні відношення типу «клас-підклас» – кожне гасло може бути віднесено до довільного набору існуючих категорій, і наявність ієрархічних зв'язків між цими категоріями не викликає суперечностей в обробці інформації. Це дозволяє відображати різні аспекти, за якими можна класифікувати гасла енциклопедії [23]. Для інтеграції набору семантичних властивостей виділяються типові *інформаційні об'єкти* (ІО) [24] – гасла, що відносяться до визначеного набору категорій та мають фіксований набір характеристик. Більш детально типові ІО для е-ВУЕ проаналізовано в [25].

Онтологічна модель БЗ е-ВУЕ дозволяє автоматизовано знаходити змістовно близькі поняття серед гасел енциклопедії, аналізуючи їх структуру, семантику та зв'язки з іншими гаслами. Це дозволяє пропонувати користувачам нові шляхи навігації у Wiki-ресурсі та розширені функціональні можливості у пошуку.

Для визначення близькості між поняттями можуть використовуватися як знання з відповідних онтологій ПрО, наприклад, кількість посилань-зв'язків між поняттями та рівень значущості цих відношень, так і відомості з таксономій, що фіксують відношення між категоріями, та враховуються як кількість, наприклад, кількість рівнів між категоріями, до яких віднесені поняття, що порівнюються, та обсяг цих категорій.

БЗ е-ВУЕ підтримує довільні змістовні відношення між поняттями, водночас як у Вікіпедії можуть використовуватися лише два типи відношень – ієрархічне відношення «Клас-елемент класу» (категоризація) та відношення зв'язку (посилання між сторінками).

Наявність довільних, семантично визначених відношень значно розширює як виразність подання знань у ресурсі, так і можливості навігації у ньому.

В е-ВУЕ Wiki-онтологія дозволяє явно визначати наявність або відсутність ієрархічних зв'язків між категоріями е-ВУЕ, припустимість або неприпустимість їх перетину; відображати зв'язки

між категоріями та типовими Ю; вказувати, які саме семантичні властивості Wiki-сторінок, що відображають змістовні відношення між різними сторінками енциклопедії, і для яких Ю припустимі ті або інші відношення, притаманні кожному з цих Ю; описувати властивості категоризованих гасел e-ВУЕ та визначати характеристики цих властивостей.

Онтологічна модель структури БЗ e-ВУЕ використовується для підтримки семантичної навігації на порталі. Однією з істотних переваг e-ВУЕ як семантичного порталу є можливість знаходження СБП. Такий пошук базується на наступних припущеннях:

- поняття, що відповідають гаслам, які відносяться до однакового набору категорій, семантично ближчі одне до одного, ніж інші поняття, відображені на порталі;

- поняття, що відповідають гаслам, які мають однакові чи близькі значення семантичних властивостей, семантично ближчі одне до одного, ніж поняття, що відповідають гаслам, які відображені на порталі, в яких семантичні властивості мають інші значення або взагалі не визначені;

- поняття, які визначені як семантично близькі за обома попередніми критеріями, є більш семантично близькими, ніж ті, для яких виконується тільки одна умова близькості або не виконується жодної.

Для e-ВУЕ потребу у знаходженні СБП може викликати нездатність користувача коректно обрати галузь знань, до якої відноситься потрібне йому гасло, або без помилок ввести потрібний термін. В таких випадках користувач може отримати доступ до гасла, що він шукає, не безпосередньо, а через ті семантично близькі гасла, які він спроможний знайти.

Наприклад, користувач хоче знайти відомості про письменника або художника, прізвище якого він точно не пам'ятає, та не здатний точно визначити жанр його творів, але може вказати прізвище його більш відомого сучасника, який працював

у тому ж жанрі. В деяких випадках проблема знаходження СБП вирішується за допомогою знаходження СБС: це дозволяє оцінювати семантичну близькість природно-мовних текстів, що характеризують гасла e-ВУЕ.

Для розширення функціоналу e-ВУЕ, що стосується пошуку та навігації у ресурсі, виникає потреба у засобах знаходження семантично близьких Ю – як *глобально* близьких (за всією сукупністю ознак), так і *локально* близьких (тільки за певною підмножиною ознак).

Для визначення кількісних оцінок семантичної близькості довільних гасел та категорій e-ВУЕ, які при цьому розглядаються як екземпляри класів та класи відповідної БЗ, проведено аналіз наукових досліджень цієї проблематики, розглянуто переваги та недоліки існуючих підходів, враховано можливість їх масштабування для великих даних.

Використання онтологій як основи формалізації структури бази знань є ефективним засобом для встановлення семантичної близькості між елементами цієї БЗ. Особливу увагу викликає обробка окремих випадків онтологічних відношень – таких, як відношення «is-a», синонімії та ієрархічних відношень «Клас-підклас» та «Клас-екземпляр класу», що реалізуються базовими засобами технологічного середовища Semantic MediaWiki.

Для класифікації сторінок e-ВУЕ гасел застосовується набір незалежних таксономій, які дозволяють знаходити семантично подібні гасла енциклопедії та можуть використовуватися для класифікації гасел як окремо, так і інтегровано;

- поділ на три основні групи – Персоналії, Цивілізація та Природа;

- категоризація за тематичною спрямованістю: поділ на 27 базових напрямків знань відповідно з класифікацією ВАК із підкатегоріями різного рівня;

- таксономія типових Ю, для яких розроблено шаблони з основними семантичними властивостями та уніфікованою формою подання на порталі;

– поділ за формою публікації: сторінки, що представлено у паперовій версії, та сторінки, що подано виключно на порталі;

– поділ за ступенем готовності: завершені сторінки; сторінки, що шукають авторів; сторінки, над якими ведеться робота;

– поділ за часом публікації, що враховує такі семантичні властивості Wiki-сторінки, як Рік публікації, Місяць публікації та Дата публікації;

– поділ за територіальною спорідненістю, що враховує семантичні властивості Wiki-сторінки з областю значення типу Місто, Країна, Регіон тощо (наприклад, Місце народження, Місце розташування);

– поділ за часовою спорідненістю, що враховує семантичні властивості Wiki-сторінки з областю значення типу Дата (наприклад, Рік народження, Рік створення);

– поділ за параметрами, специфічними для певної наукової галузі або Про (наприклад, Кількість учасників, Площа).

В е-ВУЕ підтримуються всі основні типи пошуку – як за формальними ознаками (за ключовими словами) та за категоріями, так і на змістовному рівні за семантичними властивостями. Семантичний тип пошуку базується на оригінальній структурі БЗ е-ВУЕ, що розроблена з урахуванням досвіду експертів прикладних галузей та інженерів зі знань, і саме цей засіб доступу до інформації складає одну з основних переваг даного енциклопедійного ресурсу.

### Використання СБП для навігації в е-ВУЕ

Для пошуку СБП в е-ВУЕ можуть використовуватися наступні джерела знань щодо понять:

– таксономії категорій, що використовуються в е-ВУЕ;

– природномовні описи понять, що містяться у відповідних Wiki-сторінках;

– гіперпосилання між сторінками;

– семантичні властивості, що пов'язують між собою Wiki-сторінки е-ВУЕ;

– семантичні властивості, що пов'язують сторінки понять із значеннями даних.

Відповідно до специфіки енциклопедичного IP, недоцільно шукати сторінки, що збігаються за усіма доступними параметрами – деякі групи параметрів є унікальними (наприклад, прізвище та ім'я разом з роком народження), а інші є функціонально залежними від інших параметрів, хоча мають і самостійну цінність (наприклад, прізвище особи мовою оригіналу). Тому пропонується знаходити *локальні* СБП, обираючи ті характеристики, які можуть бути корисними для досить широкого кола досліджень.

Для демонстрації можливостей описаного вище підходу пропонується наступні локальні СБП, пошук яких вбудованими засобами Semantic MediaWiki [26] не може бути реалізований:

1) СБП, визначені за *належністю до набору категорій*;

2) СБП, визначені за *значеннями обраних семантичних властивостей*;

3) СБП, визначені за *комбінацією категорій та значень семантичних властивостей*.

Семантична близькість шуканих понять визначається відносно поточної Wiki-сторінки, яку розглядає користувач, тобто категорії та властивості цієї сторінки аналізуються як параметри такого обчислення.

У першому випадку для поточної Wiki-сторінки потрібно знайти гасла, які віднесені одночасно до тих категорій, до яких віднесена поточна сторінка. Нині реалізовано пошук за категоріями та підкатегоріями галузей знань та за типовими Ю, а службові категорії, що пов'язані з формою публікації матеріалу (наприклад, "ВУЕ") не враховуються (рис. 2).

Рис. 2. Пошук СБП для сторінки е-ВУЕ «Авіація»

Пошук може бути рекурсивно розширено для пошуку груп СБП, для яких спільною є тільки певна підмножина категорій поточної сторінки. Предметом дослідження залишається питання щодо ваги окремих категорій та врахування кількості підкатегорій та екземплярів у кожній з категорій, що досліджуються.

Другий варіант – пошук може здійснюватися не для всього контенту е-ВУЕ, а лише для сторінок, що відносяться до типових Ю одного типу – персоналій, міст, країн тощо. У такому пошуку враховуються значення деяких (обраних при створенні запиту) семантичних властивостей, що характерні для цього типу Ю, але значення цих властивостей залежатимуть від їх значень для поточної сторінки. Наприклад, для типового Ю “Персоналія” можна шукати осіб, що народи-

лися (рис. 3) в тому самому місці, сучасників тощо.

Третій варіант пошуку дозволяє шукати СБП обраної категорії з набором семантичних властивостей або категорій, що притаманні обраній сторінці. Наприклад, можна знайти видатних осіб (категорія “Персоналія”), що спеціалізувалися в тій сфері, до якої відноситься певне поняття або пристрій (рис. 4).

На сьогодні пошук СБП для е-ВУЕ знаходиться на стадії досліджень, і тому певні його елементи реалізовано лише на тестовій версії порталу.

Слід відмітити, що пошук СБП (локально та глобально) неможливо виконувати вбудованими засобами Semantic MediaWiki. Для кожного такого пошуку потрібно писати окремі запити, що аналізують код сторінок.

Семантичні властивості Ю «Персоналія» за якими можна шукати СБП

Абрамович, Всеволод Михайлович (Abramowitsch, Wsewolod)	
<b>Рік народження</b>	1890
<b>Місце народження</b>	Одеса, Україна
<b>Рік смерті</b>	1913
<b>Місце смерті</b>	Берлін, Німеччина
<b>Alma mater</b>	Берлінський технічний університет, Берлін
<b>Напрями діяльності</b>	літакобудування, випробування літальних апаратів

Семантично близькі сторінки

Пошук ВУЕ

CountrymanPages

Абрамович, Всеволод Михайлович

Земляки

Місце народження=Одеса, Україна

- 1 - Аксентьева, Зінаїда Миколаївна
- 2 - Авдєєнко, Геннадій Валентинович
- 3 - Авілов, Микола Вікторович
- 4 - Азрікан, Арнольд Григорович
- 5 - Алігер, Маргарита Йосипівна
- 6 - Абалакін, Віктор Кузьмич
- 7 - Аблов, Антон Васильович
- 8 - Аганін, Марк Абрамович

Пошук виконується за категорією Персоналія.

Рис. 3. Пошук земляків для сторінок е-ВУЕ категорії «Персоналія»


Search persons by page categories

**Авіація**


Семантичний пошук персоналій в множині категорій:

Технічні науки  
Авіаційна та ракетно-космічна техніка  
Проектування, виробництво та випробування літальних апаратів


Антонов, Олег Костянтинович




Сікорський, Ігор Іванович




Адер, Клеман Агнес




Абрамович, Всеволод Михайлович




Авдєєв, Сергій Васильович




Айвінс, Марша Сью




Айзлі, Донн Фултон




Адлер, Георгій Петрович (1886–1965)



Акїяма, Тойохіро



Анатра, Артур Антонович (1919–1943)



Категорії: Цивілізація Технічні науки Авіаційна та ракетно-космічна техніка Проектування, виробництво та випробування літальних апаратів

Пошук сторінок за набором категорій

Semantic Similarity Search

CountrymanPages Search

Persons By Category Search

Категорії гасла

Семантично близькі персоналії

... (ідів), набула актуальності тема екологізації авіації: двигунів та двигунів зі зменшеним викидом вуглекислого газу на біопаливі тощо.

... (ев А. Н. Развитие самолетов мира. Москва : Издательство ЦАГИ, 1919–1945. Москва : Российская политехническая академия наук, 1943).

... (вільної авіації в Україні. Київ : Аеробізнес, 2004).

Категорії: Цивілізація Технічні науки Авіаційна та ракетно-космічна техніка Проектування, виробництво та випробування літальних апаратів

f поділитися
tвітнути
поділитися

Рис. 4. Пошук спеціалістів (за набором категорій сторінки) для сторінок е-ВУЕ

## Висновки

В результаті проведених досліджень розроблено структурну схему БЗ е-ВУЕ та обумовлено наявність у цій структурі окремих елементів та зв'язків між ними. На основі аналізу специфіки знань, що представлені в е-ВУЕ, обґрунтовано потребу в використанні онтологічної моделі Вікі-ресурсу та визначено специфіку Вікі-онтології, що відповідає особливостям е-ВУЕ, визначено сферу застосування цієї моделі.

Використання цієї моделі забезпечує підвищення ефективності пошуку та навігації в інформаційному контенті портальної версії енциклопедії, підтримує семантичний пошук.

Проаналізовано існуючі підходи та методи знаходження СБП, розглянуто роль онтологій в цьому процесі, досліджено специфіку знаходження семантично близьких Вікі-сторінок та можливості застосування для цього онтологічної моделі БЗ е-ВУЕ, наведено практичні приклади застосування запропонованого підходу.

## Література

1. Semantic MediaWiki. – [https://www.semantic-mediawiki.org/wiki/Semantic\\_MediaWiki](https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki)
2. Guarino N. Formal Ontology in Information Systems. Formal Ontology in Information Systems. Proc. of FOIS'98. 3–15. 1998.
3. Ushold M., Gruninger M. Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review*. 1996. Vol. 11, N 2.
4. Asuncion G. Ontological Engineering: with Examples from Areas of Knowledge Management, e-Commerce and the Semantic Web (Advanced Information and Knowledge Processing). – [http://www.amazon.com/gp/reader/1852335513/ref=sib\\_rdr\\_toc/](http://www.amazon.com/gp/reader/1852335513/ref=sib_rdr_toc/).
5. Rogushina J. Semantic Wiki resources and their use for the construction of personalized ontologies. CEUR Workshop Proceedings 1631. 2016. P. 188–195.
6. Рогушина Ю.В. Теоретичні засади застосування онтологій для семантизації ресурсів Web. *Проблеми програмування*. 2018. № 2-3. С. 197–203.
7. W3C Semantic Web Activity. <http://www.w3.org/2001/sw/Activity/>.
8. Рогушина Ю.В., Прийма С.М., Строкань О.В. Створення та використання семантичних Вікі-ресурсів: навчальний довідник. Мелітополь, ФОП Однорог Т.В. 2017. 169 с.
9. Quillian M.R. Semantic memory. In Minsky, M. (Ed.), *Semantic Information Processing*. MIT Press, Cambridge, MA, 1968.
10. Collins, A., Loftus, E. A spreading activation theory of semantic processing. *Psychological Review*. 1975. 82. P. 407–428.
11. Rada R., Mili H., Bicknell E., Blettner M. Development and application of a metric on semantic nets. *IEEE Transaction on Systems, Man, and Cybernetics*. 1989. 19(1). P. 17–30.
12. Richardson R., Smeaton A.F., Murphy J. Using WordNet as a knowledge base for measuring semantic similarity between words. Working paper CA-1294, Dublin City University, School of Computer Applications, Dublin, 1994. <ftp://ftp.compapp.dcu.ie/pub/w-papers/1994/CA1294.ps.Z>
13. Lee J.H., Kim M.H., Lee Y.J. Information retrieval based on conceptual distance in IS-A hierarchies. *Journal of Documentation*. 1993. 49(2). P. 188–207.
14. Rada R., Bicknell E. Ranking documents with a thesaurus. *JASIS*. 1989. 10(5). P. 304–310.
15. Leacock C., Chodorow M. WordNet: An electronic lexical database. Cambridge. 1998. P. 265.
16. Miller G. WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4). (Special Issue). 1990.
17. Wu Z., Palmer M. Verb semantics and lexical selection. Proc. 32nd Annual Meeting of the Association for Comput. Linguistics. Las Cruces. 1994. P. 133–138.
18. Resnik P. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. *Journal of Artificial Intelligence Research*. 1999. 11. P. 95–130.
19. Крюков К.В., Панкова Л.А., Пронина В.А., Шипилина Л.Б. Меры семантической близости в онтологиях. Труды научных сессий МИФИ, Научная сессия МИФИ-2010. Т. 5. *Информационно-телекоммуникационные системы. Проблемы информационной безопасности*. С. 75–78.
20. Tversky A. Features of Similarity. *Psychological Rev*. 1977. Vol. 84. P. 327.
21. Bulskov H., Knappe R., Andreasen T. On Measuring Similarity for Conceptual

- Querying. Proc. 5th Int. FQAS Conf. LNCS. V. 2522. Berlin: Springer. 2002. P. 100.
22. Levenshtein I.V. Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory*. 1966. Vol. 10. P. 707.
  23. Методичні рекомендації з підготовки, редагування та оформлення статей до Великої української енциклопедії / За ред. Киридон А.М. К.: ДНУ «Енциклопедичне видавництво». 2015. 120 с.
  24. Rogushina J.V. The Use of Ontological Knowledge for Semantic Search of Complex Information Objects. Proc. of OSTIS-2017. 2017. P. 127–132.
  25. Рогушина Ю.В. Використання семантичних властивостей вікі-ресурсів для розширення функціональних можливостей «Великої української енциклопедії». Енциклопедичні видання в сучасному інформаційному просторі: колективна монографія / За ред. Киридон А.М. – К.: Державна наукова установа «Енциклопедичне видавництво». 2017. С. 104–115.
  26. Гришанова І.Ю., Рогушина Ю.В. Адаптація технологічних засад semantic mediawiki до потреб онлайн-версії великої української енциклопедії ВУЕ. Енциклопедистика в Україні: люди, ідеї, поступ: колективна монографія / За ред. Киридон А.М. К.: Державна наукова установа «Енциклопедичне видавництво». 2018. С. 240–253.
  - ontologies. CEUR Workshop Proceedings 1631. 2016. P. 188–195.
  6. Rogushina J. (2018) Theoretical means of ontology use for the Web resources semantization. Problems in Programming. N 2-3. P. 197–203. [in Ukrainian]
  7. W3C Semantic Web Activity. – <http://www.w3.org/2001/sw/Activity/>.
  8. Rogushina Y.V., Priyma S.M., Stokan O.V. (2017) Creating and use of the Semantic Wiki resources: tutorial. Melitopol, FOP Odiorog T.V. 169 p. [in Ukrainian]
  9. Quillian M.R. Semantic memory. In Minsky, M. (Ed.), *Semantic Information Processing*. MIT Press, Cambridge, MA, 1968.
  10. Collins, A., Loftus, E. A spreading activation theory of semantic processing. *Psychological Review*. 1975. 82. P. 407–428.
  11. Rada R., Mili H., Bicknell E., Blettner M. Development and application of a metric on semantic nets. *IEEE Transaction on Systems, Man, and Cybernetics*. 1989. 19(1). P. 17–30.
  12. Richardson R., Smeaton A.F., Murphy J. Using WordNet as a knowledge base for measuring semantic similarity between words. Working paper CA-1294, Dublin City University, School of Computer Applications, Dublin, 1994. <ftp://ftp.compapp.dcu.ie/pub/w-papers/1994/CA1294.ps.Z>
  13. Lee J.H., Kim M.H., Lee Y.J. Information retrieval based on conceptual distance in IS-A hierarchies. *Journal of Documentation*. 1993. 49(2). P. 188–207.
  14. Rada R., Bicknell E. Ranking documents with a thesaurus. *JASIS*. 1989. 10(5). P. 304–310.
  15. Leacock C., Chodorow M. WordNet: An electronic lexical database. Cambridge. 1998. P. 265.
  16. Miller G. WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4). (Special Issue). 1990.
  17. Wu Z., Palmer M. Verb semantics and lexical selection. Proc. 32nd Annual Meeting of the Association for Comput. Linguistics. Las Cruces. 1994. P. 133–138.
  18. Resnik P. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. *Journal of Artificial Intelligence Research*. 1999. 11. P. 95–130.
  19. Kriukov K.V., Pankova K.V., Pronina V.A., Shipilina L.B. (2010) Measures of semantic proximity in ontologies. Proc. of MIFI-2010, Vol. 5. Information- telecommunication systems. Problems of information security. P. 75–78. [in Russian]

## References

1. Semantic MediaWiki. – [https://www.semantic-mediawiki.org/wiki/Semantic\\_MediaWiki](https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki)
2. Guarino N. Formal Ontology in Information Systems. *Formal Ontology in Information Systems*. Proc. of FOIS'98. 3–15. 1998.
3. Ushold M., Gruninger M. Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review*. 1996. Vol. 11, N 2.
4. Asuncion G. Ontological Engineering: with Examples from Areas of Knowledge Management, e-Commerce and the Semantic Web (Advanced Information and Knowledge Processing). – [http://www.amazon.com/gp/reader/1852335513/ref=sib\\_rdr\\_toc/](http://www.amazon.com/gp/reader/1852335513/ref=sib_rdr_toc/).
5. Rogushina J. Semantic Wiki resources and their use for the construction of personalized



20. Tversky A. Features of Similarity. *Psychological Rev.* 1977. Vol. 84. P. 327.
21. Bulskov H., Knappe R., Andreasen T. On Measuring Similarity for Conceptual Querying. *Proc. 5th Int. FQAS Conf. LNCS.* V. 2522. Berlin: Springer. 2002. P. 100.
22. Levenshtein I.V. Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory.* 1966. Vol. 10. P. 707.
23. Methodical recommendations for preparing, editing and design of articles for Great Ukrainian Encyclopedia (2015) / Ed. Kyrydon A.M., Kyiv, Kyiv, 120 p. [in Ukrainian]
24. Rogushina J.V. (2017) The Use of Ontological Knowledge for Semantic Search of Complex Information Objects // *Proc. of OSTIS-2017.* P. 127–132.
25. Rogushina J.V. (2017) Use of semantic properties of the Wiki resources for expansion of functional possibilities of “Great Ukrainian Encyclopedia”. *Encyclopaedias in the modern information space: collective monograph* / Ed. Kyrydon A.M., Kyiv. P. 104–115. [in Ukrainian]
26. Grishanova I.Y., Rogushina J.V. (2018) Adaptation of technological means of Semantic Mediawiki for needs of online version of Great Ukrainian Encyclopedia // *Encyclopaedias in Ukraine: people, ideas, steps: collective monograph* / Ed. Kyrydon A.M., Kyiv. P. 240–253. [in Ukrainian]

**Про авторів:**

*Рогушина Юлія Віталіївна,*  
кандидат фізико-математичних наук,  
старший науковий співробітник.  
Кількість наукових публікацій в  
українських виданнях – 150.  
Кількість наукових публікацій в  
зарубіжних виданнях – 31.  
<http://orcid.org/0000-0001-7958-2557>,

*Гришанова Ірина Юріївна,*  
науковий співробітник.  
Кількість наукових публікацій в  
українських виданнях – 18.  
Кількість наукових публікацій в  
зарубіжних виданнях – 3.  
<http://orcid.org/0000-0003-4999-6294>.

**Місце роботи авторів:**

Інститут програмних систем  
НАН України,  
03181, Київ-187,  
проспект Академіка Глушкова, 40.  
Тел.: 066 550 1999.  
E-mail: [ladamandraka2010@gmail.com](mailto:ladamandraka2010@gmail.com),  
[i26031966@gmail.com](mailto:i26031966@gmail.com)

Одержано 16.10.2019