

МЕТАДАНИ ЯК ЗАСІБ СЕМАНТИЧНОГО АНАЛІЗУ СКЛАДНИХ КОНТЕНТІВ ВЕЛИКИХ ДАНИХ. ЗОБРАЖЕННЯ

Метою дослідження є визначення ефективних підходів щодо вдосконалення семантичного аналізу графічних контентів великих даних, а саме таких як зображення або відеосцени. Сутність запропонованого підходу полягає в урахуванні особливих характеристик складних контентів та створенні гібридної моделі анотування, що розширює текстову модель більш специфічними елементами, наприклад, для візуальних даних, характеристиками візуалізації. Визначення подібності інформаційних контентів є критичною проблемою для вирішення цілої низки задач великих даних. В тому числі, є основою для категоризації цих контентів. Семантичний аналіз інформаційних контентів зазвичай базується на їхніх метаданих, які складають основу семантичних анотацій та є елементами структурованого семантичного опису контенту й базисом для його автоматизованої обробки. В основу підходу покладено використання онтологій для визначення семантичних анотацій. Онтології надають різноманітні джерела знань для вимірювання семантичної подібності, містять багато інформації про тлумачення понять та інші семантичні зв'язки з ієрархічною структурою, що базується на відносинах гіпонімії. Але останні роки разом зі швидким зростанням кількості зображень та відеоресурсів, спостерігається суттєве збагачення доступної візуальної інформації. З візуальної точки зору легше зрозуміти, чи є подібними два поняття. Тому інтеграція семантичної та візуальної інформації зображення забезпечує оптимізацію методів оцінювання подібності, що засновані на онтологіях, та дозволяє отримати більш узгоджені з уявленням людини метрики подібності. Поєднання ознак візуалізації з семантичними та онтологічними характеристиками контенту у формуванні оцінок подібності й становлять центральну ідею даного дослідження.

Ключові слова: складний контент, семантична подібність інформації, візуальні дані, дескриптор, дескрипторний простір, ключові точки зображення, вектор візуалізації, онтологічні показники подібності, текстова модель анотування, гібридна модель анотування, словник візуальних слів, метадані великих даних, характеристики візуалізації.

Вступ

Нагальною задачею сучасної інформаційної спільноти є створення методів, що уможлиблювали б швидкий аналіз інформаційних джерел та представлених в них даних. Найголовнішою макроціллю такого аналізу є семантична категоризація інформаційних контентів. Водночас, слід враховувати, що інформація, як правило, різномірдна, містить дані різних типів (часто не текстові) і форматів представлення та великі об'єми. Але з цими інформаційними елементами, окрім саме їхнього контенту, можуть бути асоційовані такі об'єкти даних, як назва, опис, мітки тощо. Та їхній зміст є джерелом текстової семантичної інформації і становить базис для формування семантичних метаданих, які є елементами структурованого семантичного опису контенту та основою для його автоматичної обробки. Також існують метадані, що не завжди по-

мітні користувачеві, та можуть бути розкидані по всьому контенту. Але вони дуже необхідні для пошуку, структуризації, категоризації та обробки інформації. Метадані можуть використовуватись для вирішення цілої низки задач великих даних, включно із категоризацією, композицією документів, конвертацією неструктурованого контенту в структури релевантних знань, візуалізацію інформації на основі онтологій. Семантично анотований контент стає джерелом інформації, яку простіше інтерпретувати, комбінувати та повторно використовувати автоматизовано.

Це породжує цілу низку задач метаданих, які потребують вирішення. Перш за все, витягнення метаданих з контенту, аналіз їхніх зв'язків, створення й формалізація метаданих зі збереженням їхніх зв'язків, їх класифікація, приєднання додаткової семан-

тичної інформації до документу великих даних, розробка схем використання метаданих тощо. Зрозуміло, що, міркуючи про набори метаданих, які описують контенти різних типів, треба враховувати, що кожний з них має власні характеристики. У [1] вже були визначені спільні аспекти та підходи до семантичного анотування контенту великих даних за допомогою метаданих та запропонована їхня загальна класифікація. Метою даного дослідження є вдосконалення запропонованої загальної класифікації з урахуванням особливих характеристик складних контентів та створення гібридної моделі анотування, що розширює тестову модель більш специфічними елементами, як-от, для візуальних даних, характеристиками візуалізації.

Текстова модель анотування великих даних

Текстова модель анотування охоплює всі можливі типи текстових метаданих, пов'язані з документом великих даних. Усі анотації за призначенням можна поділити на дві основні групи: метадані, що описують загальні характеристики документу, та метадані прикладного контенту.

Метадані для анотування загальних властивостей об'єктів великих даних є контекстно-незалежними і для їх визначення розроблено чимало спеціальних стандартів. Насамперед, це загальні характеристики документу, такі як автори, дата та час створення, формат тощо. Або загальні категорії об'єктів, як особа, організація, локація тощо. Вони можуть бути застосовані до будь-якого документа, не залежно від формату представлення. Найбільш розповсюдженим стандартом, що специфікує множину загальних характеристик документу, є відомий стандарт Dublin Core [2].

Наступна група – це метадані, що визначають загальні характеристики контентів певних типів: відео, зображень, аудіо, веб-сторінок, постів (твітів) тощо, та враховують їхні специфічні характеристики. Для специфікації таких метаданих на сьогодні розроблено безліч стандартів та онтологій. Наприклад, онтології FOAF (Friend of a Friend) [3], OntoWeb [4], KnowledgeWeb [5] для створення анотованої мережі домашніх сто-

рінок людей, груп або компаній, стандарти EXIF [6], DCF [7], IPTC, що визначають метадані для опису зображень, тощо.

Такі групи онтологій та стандарти дозволяють автоматично визначити велику кількість технічних характеристик документа великих даних та деяку загальну його семантику, але не охоплюють деталей його контенту (можливо, лише деякі ключові слова та описи природною мовою в таких характеристиках, як тема або опис).

Справжній семантичний опис, який викриває сутність та особливості контенту даних, вимагає визначення різних типів контекстно-залежних метаданих, що не можливо без застосування прикладних онтологій домену. Це можуть бути загальні прикладні онтології, призначені для анотування документів або вирішення проблем у певному широкому домені. Так, наприклад, онтології Esperanto Cultural Tour [8] та Fund Finder дозволяють анотувати документи в прикладних доменах культура та фінансування, відповідно. Такі онтології, як правило, визначають загальні характеристики обраної прикладної області. А можуть бути більш специфічні прикладні онтології, які дозволяють точніше визначити семантику даних. Але їх застосування у повністю автоматизованому режимі є проблематичним з огляду на те, що вони є досить специфічними для конкретного об'єкта даних та задачі, що вирішується.

Всі ці метадані або прямо визначаються в контенті: в тегах, атрибутах, описах тощо, або непрямо - за допомогою посилань на інші джерела, стандарти та онтології. Але так чи інакше вони мають текстові значення, їх можна вважати текстовими або описовими. Більш структуровано загальна класифікація текстових метаданих на верхньому рівні представлена на рис.1 [9] нижче.

Великі дані за визначенням переважно мають не структурований контент значного розміру. Документ великих даних, окрім пов'язаних з ним описових елементів (семантичних анотацій), може мати характеристики, притаманні даному типу контенту, витягнення та аналіз яких дозволяє глибше дослідити відношення подібності між використовуваними поняттями, та отримати достовірніше уявлення про зміст контенту.

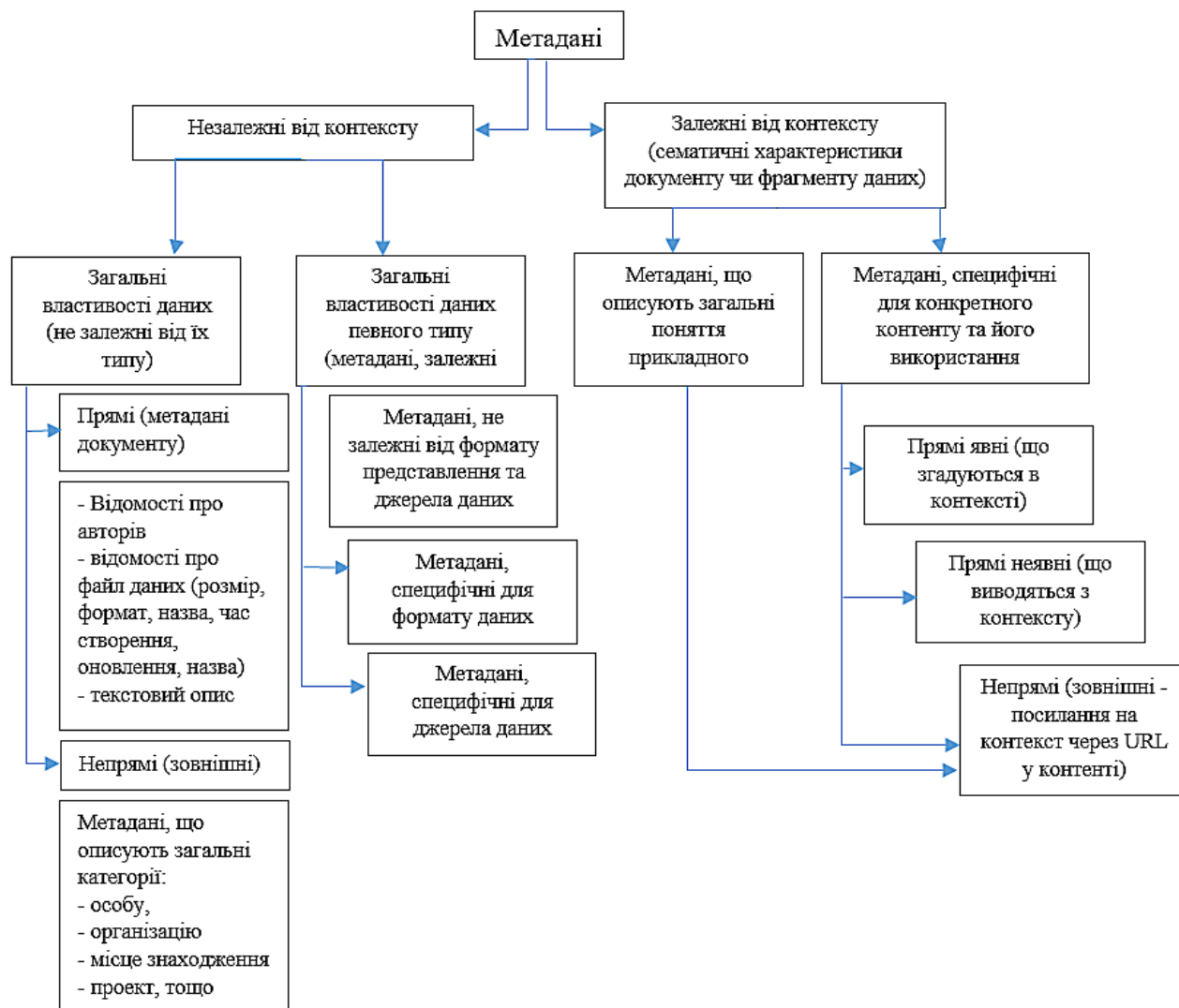


Рис.1. Загальна класифікація текстових метаданих верхнього рівня

Загалом, подібність понять є результатом суб'єктивного сприйняття людини. Окрім семантичної інформації, саме зміст контенту (зображення, відео, аудіо тощо) й забезпечує інформацію зорового/слухового сприйняття, що також відіграє важливу роль в доступі до відношень подібності понять.

Тому ці характеристики можуть бути використані для пошуку, порівняння й встановлення відповідності контентів, та навіть у методах побудови кількісних показників відповідності для підвищення їхньої точності. Для контентів типу зображення такими показниками можуть слугувати характеристики візуалізації.

Характеристики візуалізації

Характеристики візуалізації – це описові елементи зображення, що витягуються

саме з контенту, переважно, за допомогою стандартного BoVW (bag-of-visual-words) представлення зображень [10].

Цей підхід базується на поняттях ключових точок зображення, дескрипторів, дескрипторного простору та словника візуальних слів.

Ключовими точками [11, 12, 13] називають помітні фрагменти зображення, що містять багату локальну інформацію про зображення, яка може бути автоматично виявлена за допомогою різноманітних детекторів [13, 14] й представлена багатьма дескрипторами [15]. Потім ключові точки поєднуються у групи – у велику кількість кластерів, водночас ті точки, що мають подібні дескриптори, призначаються до одного й того самого кластера. Кожний кластер розглядається як «візуальне слово», що

представляє конкретний локальний шаблон, який розділяється ключовими словами в цьому кластері. Це дозволяє отримати словник візуальних слів, який описує всі види таких локальних шаблонів зображень. У такому випадку зображення може бути представлено як «набір візуальних слів», або, зокрема, як вектор, який містить (зважену) кількість входжень кожного візуального слова в цьому зображенні, що використовується як вектор ознак в задачі класифікації. Такі вектори ознак називаються дескрипторами, а множина цих векторів ознак утворює дескрипторний простір.

Тобто, кожне візуальне слово є результатом аналізу певної області зображення. Первинне виявлення таких слів здійснюється шляхом розбору зображень, помічених текстовими тегами. Наприклад, маємо зображення, певна область якого помічена словом «кіт». Відповідний фрагмент зображення декомпонується на групи графічних елементів конкретних форм та кольорів. Підраховується кількість екземплярів в кожній групі та з отриманих значень формується вектор ознак, що відповідає візуальному слову «кіт» (рис.2).



Рис.2. Формування вектору ознак

Вектор ознак наведеного на рисунку прикладу базується на графічних складових зображення. Це відповідає SIFT (Scale-Invariant Feature Transform) системі характеристик візуалізації, але можуть використовуватися й інші системи характеристик. Наприклад, таких, що базуються лише на інформації з кольорових складових зображення [19], або ребрах [17], текстурах [18], формах [19] тощо.

Теги, що помічають зображення, також представляються векторами. Обидва типи векторів потребують загальної нормалізації, після чого їх можна конкатенувати.

Вимірювання простору в даному випадку можна сформувати наступним чином: значення кожного вимірювання обчислюється як сума входжень відповідного візуального слова у всіх зображеннях колекції, помічених цим словом.

Поєднання текстового та візуального представлень створюють єдиний мультимодальний простір. Для конкатенації векторів представлена можуть використовуватися різні підходи. У [20] запропоновано лінійну зважену функцію, що нормалізує та комбінує вектори текстового (F_t) та візуального (F_v) представлення:

$$F = \alpha F_t \otimes (1 - \alpha) F_v \quad F = \alpha F_t \otimes (1 - \alpha) F_v, \quad (1)$$

де \otimes позначає операцію конкатенації, а α – зважений параметр, що встановлює співвідношення між текстовим та візуальним представленнями контенту.

У літературі описано чимало методів, в тому числі експериментальних, до визначення візуальних, текстових та гібридних моделей [21] з аналізом ефективності їх застосування. Очевидно, що від якості представлення вектора візуалізації суттєво залежить адекватність вирішення подальших задач. Підвищити ефективність вектора візуалізації можливо шляхом:

- скорочення словника візуальних слів,
- видалення з вектора візуального представлення «стоп-слів»
- зважування візуальних слів у представленні,
- становлення специфічних критеріїв вимірювання інформативності візуальних слів та видалення «не інформативних» слів з представлення,
- збагачення представлення просторовою інформацією, а саме врахування просторового розташування ключових точок.

Але дане дослідження не має за мету розробку або розгляд тих чи інших підходів до побудови векторів візуалізації, а швидше присвячене інтеграції семантичної та візуальної інформації зображення та визначенню оцінок подібності інформації на основі

створеного мультимодального простору, й використанню їх для встановлення ступеня подібності даного типу контентів та для їх класифікації (чи категоризації). Водночас в основу побудови оцінок подібності покладені метрики, засновані на онтологіях, та, фактично, отримані показники повинні являти собою оптимізацію вказаних метрик.

Використання характеристик візуалізації в інтегрованих показниках семантичної подібності

Перевагою визначення подібності понять онтологій є те, що ми одразу маємо справу з готовими концептами, тобто інформація є очищеною, без «шумів». У загальному випадку, у роботі зі складними контентами існує велика кількість «зашумлених» даних, що призводить до неточних мір подібності. Використання візуальних особливостей зображення разом із пов'язаними з ними семантичними анотаціями дозволяють оптимізувати відношення семантичної подібності. Поєднання в онтології різних семантичних знань (що стосуються даного типу контенту), таких як ієрархічна структура та семантичні відношення, дозволяє сформуванню багатшу базу семантичних знань. Включення візуальної інформації та інтеграція декількох баз знань разом дозволяє різнобічно виразити відношення семантичної подібності понять.

Метрики подібності концептів, що базуються на онтологіях, можна розділити на чотири категорії: методи на основі довжини шляху [22], методи на основі властивостей [25], методи на основі інформаційного контенту [23, 24] й так звані «gloss-based» (на основі глосів) методи [31]. Аналіз та метрики перших трьох груп методів було наведено у [26].

Метод на основі глосів (тлумачень) вперше був запропонований для усунення семантичної неоднозначності. У [32] порівнюються глоси слова у фразі з глосами інших, знаходячи найбільш подібні сенси як сенс слова в обраній фразі. Глоси описувалися у словнику. Пізніше, у [27] цей словник був заміщений загальним словником Wordnet, та на його основі було створено відповідну онтологію. Передумовою даного методу було існування ідеальної структури,

яка включає деталізований глосарій слів. Wordnet взмозі задовільнити вимоги щодо забезпечення ієрархічної структури понять і великої кількості семантичних тлумачень.

Онтологія надає різноманітні джерела знань для вимірювання семантичної подібності, містить багато інформації про тлумачення поняття та інші семантичні зв'язки з ієрархічною структурою, що базується на відносинах гіпонімії. Проте, це все семантична текстова інформація. Останнім часом, разом зі швидким зростанням кількості зображень спостерігається суттєве збагачення доступної візуальної інформації. З візуальної точки зору легше зрозуміти, чи є подібними два поняття. Тому інтеграція семантичної та візуальної інформації зображення здатна забезпечити оптимізацію методу оцінювання подібності, заснованого на онтологіях, та дозволяє отримати більш узгоджені з уявленням людини метрики подібності.

Припустимо, існує N зображень, кожне з яких анотоване однією або декількома мітками з M концептів $c_i, i = 1, \dots, M$. Нехай,

$$s_{ij} = \begin{cases} 1, & \text{якщо зображення } j \text{ анотовано} \\ & \text{концептом } c_i \\ 0, & \text{навіпаки} \end{cases}$$

Отримуємо так звану матрицю семантичних відношень S розміром $M \times N$.

Візуальні оцінки можуть бути сформовані за допомогою матриці кореляцій V , елементи якої виражають подібність векторів візуальних ознак окремих концептів та зображень. Кожному концепту c_i відповідає підмножина зображень, що ним анотовані. Тоді візуальні властивості концепта c_i можуть бути описані цією підмножиною зображень. Середнє візуальних властивостей всіх цих зображень підмножини саме і є візуальною оцінкою концепта c_i . У випадку, якщо позначити вектор візуальних ознак концепта c_i як f_i , а візуальних ознак зображення j як f_j , то елементи матриці V - v_{ij} саме виражають кореляцію між цими двома векторами. У [26] для визначення величини кореляції пропонується використовувати косинус кута між векторами візуалізації:

$$v_{ij} = \cos(f_i, f_j) = \frac{\langle f_i, f_j \rangle}{\|f_i\| \|f_j\|}$$

$$v_{ij} = \cos(f_i, f_j) = \frac{\langle f_i, f_j \rangle}{\|f_i\| \|f_j\|}$$

Можна сказати, що даний показник визначає косинусну подібність векторів візуальних ознак, а саме: схожі вектори будуть мати величину кореляції, значення якої наближається до нуля.

Далі ці оцінки можуть бути використані у формуванні комплексних векторів ознак:

$$VS = \mu * V \oplus (1 - \mu) * S$$

$$VS = \mu * V \oplus (1 - \mu) * S,$$

Де \oplus - операція конкатенації, а $\mu \in [0,1]$ – ваговий коефіцієнт, що встановлює баланс семантичної та візуальної інформації в матриці подібності концептів. μ досягає 0, коли матриця семантичної подібності обчислюється лише на основі семантичних анотацій та навпаки, $\mu = 1$, якщо враховуються тільки візуальні ознаки.

Отож, отримана матриця є об'єднаною матрицею ознак $VS = [y_1, \dots, y_i, \dots, y_m]^T$ $VS = [y_1, \dots, y_i, \dots, y_m]^T$, де кожний рядок y_i y_i є комплексним вектором ознак концепту $c_i c_i$. Слід зазначити, якщо два концепти є семантично близькими, то їхні комплексні вектори також мають бути дуже схожими. Це обумовлює доцільність використання оцінок косинусної подібності також і на етапі визначення комплексної семантичної подібності концептів.

У такому випадку, значення комплексної семантичної подібності двох концептів $c_i c_i$ та $c_j c_j$ на основі семантичних анотацій та візуальних ознак зображення можна визначити як:

$$r_{ij}^{VS} = \cos(y_i, y_j) = \frac{\langle y_i, y_j \rangle}{\|y_i\| \|y_j\|}$$

$$r_{ij}^{VS} = \cos(y_i, y_j) = \frac{\langle y_i, y_j \rangle}{\|y_i\| \|y_j\|},$$

де y_i, y_j y_i, y_j –

комплексні вектори ознак концептів $c_i c_i$ та $c_j c_j$, відповідно. Елементи $r_{ij}^{VS} r_{ij}^{VS}$ формують матрицю комплексної семантичної подібності R.

Врахування візуальної інформації при обчисленні ступеня подібності концеп-

тів дозволяє компенсувати брак семантичної інформації (якщо має місце неповнота семантичних анотацій) та уникнути проблеми «розрідженості» матриці подібності. А запобігти зміщенню центру візуальних ознак дозволяє встановлення ефективного значення коефіцієнту μ , який би дійсно забезпечував баланс між цими видами семантичної інформації. Його визначення є окремою математичною задачею. На сьогодні відомі різні методи та підходи до її реалізації [23, 29, 30].

Як було зазначено вище, наведений показник побудований на основі семантичних анотацій зображення та його візуальних ознак. Далі він може бути використаний для оптимізації онтологічних показників подібності [26]. Результуюча оцінка повинна бути композитною функцією від онтологічного показника подібності концептів та їх комплексної оцінки подібності, на основі семантичних та візуальних ознак. Тобто, якщо $r_{ij}^{OO} r_{ij}^{OO}$ визначає онтологічну подібність концептів $c_i c_i$ та $c_j c_j$ (наприклад, через довжину шляху в онтології, яка їх містить), а $r_{ij}^{VS, VS} r_{ij}^{VS, VS}$ – оцінка, що побудована на основі комплексних векторів ознак цих концептів, то результуючий показник подібності концептів $c_i c_i$ та $c_j c_j$:

$$r_{ij}^{OVS} = F_1(r_{ij}^{OO}) * F_2(r_{ij}^{VS})$$

$$r_{ij}^{OVS} = F_1(r_{ij}^{OO}) * F_2(r_{ij}^{VS})$$

Поєднання семантичних та візуальних ознак у традиційну онтологічну метрику дозволяє компенсувати недостатність різнобічних та складних відношень в онтології.

Висновки

Фокусом даного дослідження є питання вдосконалення підходів до визначення подібності інформаційних контентів, або контентів великих даних, що базуються на онтологіях, на основі всебічного інтелектуального аналізу відповідної інформації. Зокрема, в даній статті розглянуто такі нетривіальні типи великих даних як зображення, або відеосцени, тобто візуальні контенти. Явні або неявні метадані, пов'язані з контентом, надають його текстовий семантичний опис. Доцільність використання он-

тологій для визначення таких метаданих обумовлена їх високим рівнем формалізації та наявністю ефективних механізмів міркувань. Це, в свою чергу, уможливорює автоматизований аналіз цих контентів. Онтологічні міри подібності забезпечують швидке та достовірне визначення ступеня відповідності контентів на основі схожості концептів метаданих, що їх описують. Однак, ці оцінки не враховують зорового сприйняття зображення людиною. Підхід, що розглядається у даній статті, спрямований саме на оптимізацію онтологічних метрик шляхом розширення їх показниками, які враховують саме характеристики візуалізації зображень. Такі оцінки комплексної семантичної подібності концептів визначаються шляхом композиції двох функцій. Перша з них фактично є онтологічною мірою подібності, а друга будується на основі його комплексного вектора ознак, що є конкатенацією семантичних та візуальних характеристик із встановленим ваговим балансом між цими двома різновидами інформації.

Література

1. О. Захарова (2019). Використання метаданих для вирішення задач великих даних. Проблеми програмування №4, 2019, С.81 – 91.
2. <https://www.dublincore.org/specifications/dublin-core/dces/>
3. <http://www.foaf-project.org/>
4. <http://www.ontoweb.org/>
5. <http://knowledgeweb.semanticweb.org/>
6. <https://www.exif.org/category/specifications>
7. <http://exif.org/dcf.PDF>
8. <http://www.esperanto.net>
9. О. Захарова (2020). Основні аспекти семантичного анотування великих даних. Проблеми програмування №4, 2020, С.22 – 33.
10. J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In Proc. of 9th IEEE Int'l Conf. on Computer Vision, Vol. 2, 2003.
11. S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Proc. of 2006 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, volume 2, pages 2169{2178, 2006.
12. F.-F. Li and P. Perona. A bayesian hierarchical model for learning natural scene categories. In Proc. of the 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pages 524{531, 2005.
13. J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: An in-depth study. In Technical report, INRIA, 2005.
14. K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. Int. J. Comput. Vision, 60(1):63-86, 2004.
15. K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(10):1615-1630, 2005.
16. Mark D. Fairchild. 2005. Status of cie color appearance models.
17. John Canny. 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell, 36(4):679–698.
18. Song Chun Zhu, Cheng en Guo, Ying Nian Wu, and Yizhou Wang. 2002. What are textons? In Computer Vision - ECCV 2002, 7th European Conference on Computer Vision, Copenhagen, Denmark, May 28-31, 2002, Proceedings, Part IV, pages 793–807. Springer.
19. Aude Oliva and Antonio Torralba. 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. Int. J. Comput. Vision, 42:145–175.
20. Elia Bruni, Giang Binh Tran, and Marco Baroni. 2011. Distributional semantics from text and images. In Proceedings of the EMNLP GEMS Workshop, pages 22–32, Edinburgh.
21. Marco Baroni and Alessandro Lenci. 2010. Distributional Memory: A general framework for corpus-based semantics. Computational Linguistics, 36(4):673–721.
22. Collet, C., Huhns, M.N., Shen, W.M.: Resource integration using a large knowledge base in carnot. IEEE Computer 24 (1991) 55–62
23. Tversky, A.: Features of similarity. Psychological Review 84 (1997) 327–352
24. Jang, J., Conrath, D.: Semantic similarity based on corpus statistic and lexical taxonomy. In: Proceedings of the International Conference on Computational Linguistics. (1997)
25. 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 11 (1999) 95–130.
26. O. Zakharova. Defining degree of semantic similarity using description logic tools. Проблеми програмування. — 2021. — № 2. — С. 24-33.
 27. Banerjee S, Pedersen T. An adapted Lesk algorithm for word sense disambiguation using WordNet[M]. Computational linguistics and intelligent text processing. Springer Berlin Heidelberg, 2002: 136-145. DOI: http://dx.doi.org/10.1007/3-540-45715-1_11
 28. Mengyun Wang, Xianglong Liu, Lei Huang, Bo Lang, Hailiang Yu. 2014. Ontology-based Concept Similarity Integrating Image Semantic and Visual Information. Proceedings of the 2014 Federated Conference on Computer Science and Information Systems pp. 289–296.
 29. Rodríguez M A, Egenhofer M J. Determining semantic similarity among entity classes from different ontologies. Knowledge and Data Engineering, IEEE Transactions on, 2003, 15(2): 442-456.
 30. Zhou Z, Wang Y, Gu J. A new model of information content for semantic similarity in WordNet. Future Generation Communication and Networking Symposia, 2008. FGCNS'08. Second International Conference on. IEEE, 2008.
 31. Patwardhan S, Pedersen T. Using WordNet-based context vectors to estimate the semantic relatedness of concepts. Proceedings of the EACL 2006 Workshop Making Sense of Sense-Bringing Computational Linguistics and Psycholinguistics Together. 2006, 1501:1-8.
 32. Lesk M. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. Proceedings of the 5th annual international conference on Systems documentation. ACM, 1986: 24-26. DOI:<http://dx.doi.org/10.1145/318723.318728>

Одержано: 16.01.2023

Про автора:

Захарова Ольга Вікторівна,
кандидат технічних наук,
старший науковий співробітник.
Кількість наукових публікацій в українських виданнях – 33.
<http://orcid.org/0000-0002-9579-2973>.

Місце роботи автора:

Інститут програмних систем НАН України,
проспект Академіка Глушкова, 40.
Тел.: 526 5139.
E-mail: ozakharova68@gmail.com.
Моб.тел.: +38(068)594756