

Хміль Ірина Олександрівна

Аспірант кафедри соціальних комунікацій та інформаційної діяльності
Національний університет «Львівська політехніка», Львів

**АНАЛІЗ ВЗАЄМОДІЇ СКЛАДОВИХ ІНФОРМАЦІЙНОГО ОБРАЗУ
ГЕОПРОСТОРОВИХ ОБ'ЄКТІВ У ВІРТУАЛЬНИХ СПІЛЬНОТАХ**

***Анотація.** Розглянуто взаємодію суб'єктів, що формують інформаційний образ геопросторових об'єктів у віртуальній спільноті у реальному та віртуальному середовищах. Також описано формальну модель корпусу релевантних текстів, екстрагованих з інформаційного наповнення віртуальних спільнот за допомогою словника синонімів назв геопросторових об'єктів.*

***Ключові слова:** геопросторовий об'єкт; віртуальна спільнота; інформаційний образ; модель; корпус текстів*

Вступ

Аналіз наповнення віртуальних спільнот у мережі Інтернет у сучасному суспільстві дає можливість отримати консолідований ресурс потрібної інформації про геопросторовий об'єкт (ГпрО) без безпосереднього контакту з великою кількістю суб'єктів досвіду – авторів. Активний і постійний аналіз наповнення віртуальних спільнот може надати нам ретроспективну картину попереднього досвіду як одного автора, так і загальну картину досвіду, що стосується обраного ГпрО – тобто загальний інформаційний образ.

Для отримання максимально близького до реальності інформаційного образу потрібно здійснити класифікацію авторів, що дасть нам змогу отримати певний рівень точності отриманого консолідованого досвіду. Такий рівень точності отримується за допомогою певних елементів, що знаходяться у структурі дописів віртуальних спільнот: власне тіло допису та інших додаткових даних, отримання яких забезпечують інструменти віртуальної спільноти. Саме тому автор, ГпрО і віртуальна спільнота (ВС) розглядаються як рівноправні складові інформаційного образу.

**Аналіз останніх досліджень
і публікацій**

Однією із важливих проблем аналізу інформаційного наповнення віртуальних спільнот на сьогодні є аналіз соціальних даних виділеного автора дописів. Проте, попри суттєве значення для подальшого розвитку цієї сфери досліджень є нерозробленими методи аналізу саме попереднього соціального досвіду учасників віртуальних спільнот з корпусу релевантних текстів.

Питаннями отримання та аналізу корпусу релевантних конкретному інформаційному запиту

текстів розглянуто у роботах [2 – 4]. Також в контексті релевантності важливим є питання формування словника синонімів назв геопросторових об'єктів, а також виділення переліку слів, що можуть їх замінити [10].

Питання отримання соціальних і персональних характеристик авторів у віртуальних спільнотах з власне текстового наповнення розглянуто у [7; 9], проте аналіз попередньо не задекларованих соціальних даних є малорозвиненим.

Це, своєю чергою, породжує актуальну проблему розроблення нових інструментів аналізу соціальних характеристик авторів віртуальних спільнот, які би мали належне наукове обґрунтування, формалізованість та універсальність.

Мета статті

Метою статті є аналіз взаємодії суб'єктів інформаційного образу у віртуальному середовищі, а також розглянути основні складові, які можна виділити з корпусу релевантних текстів у віртуальній спільноті, що реалізуються суб'єктами «геопросторовий об'єкт – автор – віртуальна спільнота».

Аналіз предметної області

У формуванні інформаційного образу геопросторових об'єктів у віртуальних спільнотах беруть участь три суб'єкти (рисунк):

- автор – особа, що взаємодіє з об'єктами реального світу і в результаті набуття досвіду генерують контент у віртуальній спільноті;
- геопросторовий об'єкт – об'єкт реального світу, який має свої однозначні координати, а також характеристики і результати діяльності, що є основою для формування досвіду авторів;
- віртуальна спільнота – інструмент для генерування, ідентифікації та аналізу контенту, що містить досвід автора.

Вони взаємодіють один з одним у двох середовищах:

- віртуальне середовище – середовище, в якому міститься некатегоризована та неаналізована, авторизована та неавторизована інформація про різні об’єкти реального і віртуального світу;

- реальне середовище – середовище існування реальних об’єктів, зокрема геопросторових, їх взаємодії з авторами та формування авторського попереднього досвіду.

Результатом такої взаємодії у будь-якому з двох середовищ є досвід, який формується у вигляді:

I – () (результат взаємодії автора *A* і геопросторового об’єкта *GsO* у реальному середовищі *RealEn*) – передбачає попередню взаємодію автора з геопросторовим об’єктом у реальному середовищі, тобто набуття ним первинного особистого досвіду (наприклад, відвідування закладу або проживання поряд з певним підприємством), набуття вторинного досвіду від людей, що мали власний безпосередній досвід взаємодії з геопросторовим об’єктом. При цьому автор вже є суб’єктивною складовою, оскільки як на первинний, так і на вторинний досвід впливає попередній досвід та інші фактори, такі як освіта, досвід взаємодії з геопросторовими об’єктами такої ж категорії тощо.

II – () (результат взаємодії автора *A* і геопросторового об’єкта *GsO* у віртуальному середовищі *VEn*, неавторизований контент про геопросторовий об’єкт) – передбачає генерування певним автором контенту про геопросторовий об’єкт, проте встановлення авторських характеристик цього контенту є неможливими, або такими, достовірність яких встановити важко чи неможливо.

III – () (результат мало ідентифікованої активності автора *A* у віртуальному середовищі *VEn*) – це будь-який контент генерований автором поза віртуальною спільнотою, що також не має згадок про геопросторовий об’єкт, який аналізується. В даному дослідженні контент не розглядається як складова.

IV – () (результат соціальної активності автора *A* у віртуальній спільноті *VC*, яка не стосується геопросторового об’єкта) – весь контент, генерований автором, який не містить жодних згадок про геопросторовий об’єкт, але який дозволяє отримати інформацію про автора як суб’єкта досвіду, який стосується його географічного перебування.

V – () (неавторизована інформація про геопросторовий об’єкт *GsO* у віртуальному середовищі *VEn*) – це інформація про геопросторовий об’єкт, яка не має чітко ідентифікованого автора, і яка здебільшого стосується опису його задекларованих у реальному середовищі характеристик (наприклад, об’єкт *X* розташований у місці *Y*, займає певну площу, є підприємством з виготовлення продукції *Z*).

VI – () (доступна у *BC*, неавторизована інформація про *ГпрО*) – це інформація, достовірність авторських характеристик якої встановити важко або неможливо.

VII – () (інформаційний образ *ГпрО* у *BC*) – передбачає контент, в якому можна визначити характеристики автора, однозначно визначити маркування контенту як такого, що містить первинний або вторинний досвід взаємодії автора з геопросторовим об’єктом, визначити за допомогою інструментів віртуальної спільноти інші характеристики контенту (час допису, однозначна ідентифікація автора тощо.)

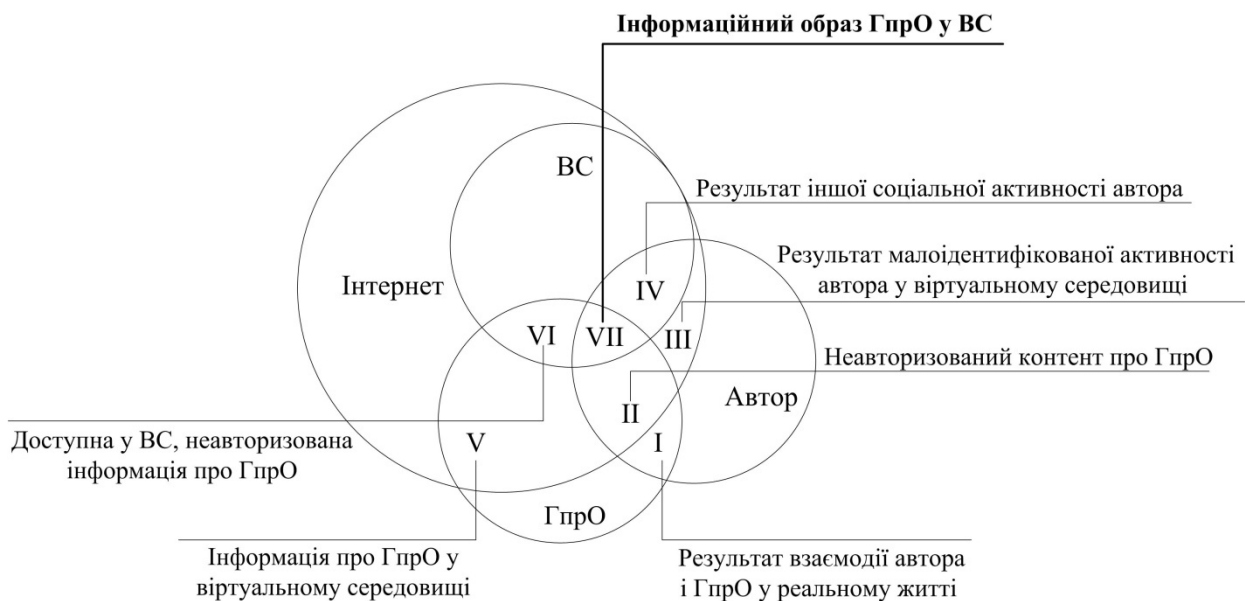


Рисунок – Загальна схема взаємодії суб’єктів інформаційного образу у реальному і віртуальному середовищах

Автори на першому етапі формують особистий образ об'єкта на основі свого попереднього досвіду: безпосередньо, на основі власних спостережень а також на основі досвіду інших людей у реальному житті або на основі отриманих від інших осіб чи колективів (група зацікавлених авторів, метою яких є надати оцінку чи викласти певну думку з метою отримання вигоди).

На другому етапі автори формують власний цілісний образ, на основі якого вони генерують свої первинні повідомлення, створюючи одиничний авторський інформаційний образ.

Наступним етапом є модифікація одиничного авторського інформаційного образу – це відповіді і дискусія, з якої екстрагується всі думки і оцінки одного визначеного автора, що є комплексним авторським інформаційним образом.

Під поняттям модифікації розглядається зміна попереднього (первинного чи вторинного) досвіду автора, в результаті якої відбувається генерування нового контенту з новим досвідом. Тобто, це трактується як зміна образу в часі.

Формальна модель корпусу текстів у віртуальних спільнотах

Віртуальна спільнота є основним джерело формування корпусу релевантних текстів – дописів авторів у спільноті, з яких ми екстрагуємо набір релевантних текстів будь-якій попередньо задекларованій назві (переліку синонімів) геопросторового об'єкта.

$$Post = \{Post_i\}_{i=1}^{N^{Post}} \in NameGsO,$$

де N^{Post} – кількість дописів у корпусі, $NameGsO$ – назва ГпрО, яка повинна бути у текстах дописів.

У такому випадку назва ГпрО може бути не лише констатованою і єдиною, а бути списком найбільш вживаних або вживаних коли-небудь назв – назви рідною мовою, транслітерація, такі, що склались історично, але і досі активно використовуються. Також під час аналізу потрібно звертати увагу до дублетних назв, які виражають непряму назву, а якусь характеристику ГпрО, але всі розуміють, про який об'єкт іде мова (Венеція – місто на воді; Рим – місто на семи пагорбах тощо). Проте часто такі назви є вже компонентами, які відображають сам інформаційний образ, тому при попередньому аналізі потрібно чітко розділяти імена і дескриптори [10; 11].

Кожний конкретний допис $Post_i$ є набором таких складових:

$$Post_i = \langle Text_body_i, Author_i, Com_char_i \rangle,$$

де $Text_body_i$ – власне текст одного допису чи комунікативної ситуації [6]; $Author_i$ – відомості про автора, які отримуємо з додаткових відомостей, що є обов'язковими елементами реєстрації дописувачів [1; 8]; Com_charac_i – інші відомості, що є важливими для подальшого аналізу ІО.

Зокрема, важливим для комплексного аналізу є питання часових характеристик. Йдеться про чітке маркування допису чи рамок комунікативної ситуації конкретним часовим моментом чи періодом публікації. Таким чином, ми отримуємо дані, в яких прослідковується ретроспекція активності одного автора про конкретний ГпрО, які ми формуємо у базу часово-залежних даних [5].

Провести відбір релевантних дописів у ВС – що дасть нам множину Post:

$$Text_body_i = \left\langle \begin{array}{l} GsO_i, DesW_i, \\ resident_mark_i, Consumer_mark_i, \\ correct_index_i, servise_w_i \end{array} \right\rangle,$$

де GsO_i – слово (талій вираз), що позначає ГпрО – попередньо задекларована назва, займенник чи інший заміник («це місто», «воно» тощо); $DesW_i$ – дескриптори – слова, що власне несуть інформаційний образ; $resident_mark_i$ – маркер, за допомогою якого ми ідентифікуємо міру причетності, проживання; $Consumer_mark_i$ – маркер, за допомогою якого ми ідентифікуємо міру споживання; $correct_index_i$ – індекс точності – виражає, наскільки інформація подана автором у дописі відповідає дійсності (розрізнення моменту публікації і моменту отримання досвіду); $servise_w_i$ – слухові, допоміжні слова, які не використовують при аналізі ГпрО.

З релевантних дописів ми екстрагуємо всі дескриптивні слова (дескриптори), що описують ГпрО. У такому випадку всі інші слова і допоміжні конструкції вилучаються з аналізу і у формуванні ІО участі не беруть.

Наступним кроком є надання дескрипторам абсолютного і відносного числа згадок у вибраному масиві дописів:

$$DesW = \{DesW_i\}_{i=1}^{N^{DesW}},$$

де N^{DesW} – число дескрипторів.

У загальному випадку $DesW_i$ є набором:

$$DesW_i = \left\langle \begin{array}{l} Type_i, AbsoluteVal_i, \\ relativeVal_i, relation_i \end{array} \right\rangle,$$

де $Type_i$ – тип дескриптора; $AbsoluteVal_i$ – абсолютне значення кількості зв'язків з ГпрО (концептом) дескриптора в масиві релевантних дописів; $relativeVal_i$ – відносне значення (процентне відношення кількості згадок до найбільш згадуваних дескрипторів чи самого ГпрО; $relation_i$ – співвідношення двох дескрипторів.

Наступним елементом структури є відношення між дескрипторами і між дескрипторами і назвами ГпрО. Ці відношення мають три типи за розташуванням:

- близьке – слова розташовані поряд або у одному реченні;

- середнє – слова не розташовані поруч, але належать до одного допису;
 - далеке – слова не розташовані поряд, але згадуються в межах дискусії про аналізований ГпрО.

Таким чином, ми отримуємо набір зв'язків R:

$$R = \{R\}_{i=1}^{N^R},$$

де N^R – кількість зв'язків.

У загальному випадку зв'язок R_i описується так:

$$R_i = \left\langle \begin{matrix} Type_i, AbsoluteVal_i, \\ relativeVal_i, DW_k, DW_m \end{matrix} \right\rangle,$$

де $Type_i$ – тип зв'язку (описано вище); $AbsoluteVal_i$ – абсолютне числове значення (вага) зв'язку між двома обраними дескрипторами; $relativeVal_i$ – відносне числове значення (вага) зв'язку між двома обраними дескрипторами (відсоткове співвідношення,

максимальне число у відсотках, що виражає відношення між двома дескрипторами – два дескриптори при високому відносному значенні можемо розглядати як сталий вираз в межах аналізу IO); DW_k і DW_m – два дескриптори, між якими розглядається зв'язок.

Висновки

Створення формальної моделі корпусу релевантних запиту про обраний геопросторовий об'єкт текстів допомагає подальшому аналізу інформаційного образу цих об'єктів, зокрема при створення баз даних, що міститимуть ретроспективну інформацію.

Групування дескрипторів і надання числових значень дає змогу проводити обрахунок їх важливості для формування інформаційного образу.

Список літератури

1. Johnson T. J. *Webelievability: A Path Model Examining How Convenience and Reliance Predict Online Credibility.* / T. J. Johnson and B. Kaye // *Journalism and Mass Communication Quarterly* (79). – 2002. – 619-642 p.
2. Oudshoff A.M., Bosloper I.E., Klos T.B., Spaanenburg L. *Knowledge discovery in virtual community texts: Clustering virtual communities* // *Journal of Intelligent & Fuzzy Systems.* – IOS Press, 2003. - Vol.14.- №1.- P.13-24.
3. Pekar V., Ou Sh. *Discovery of Subjective Evaluationsof Product Features in Hotel Reviews* // *Journal of Vacation Marketing.* – 2008. – Vol. 14. – № 2. – P. 145 – 155.
4. Sauri R, Pustejovsky J. *FactBank: a corpus annotated with event factuality* // *Language Resources and Evaluation.* - Springer Netherlands, 2009.-Vol.43. - №3. – P.227-268.
5. Жежнич П.І. *Часові бази даних: моделі та методи реалізації* : [Монографія] / П.І. Жежнич – Львів: Вид-во Нац. ун-ту «Львівська політехніка», 2007. – 260 с.
6. Тимовчак-Максимець О.Ю. *Аналіз комунікативної взаємодії на веб-форумах: інформаційна поведінка та учасники* / О.Ю. Тимовчак-Максимець, А.М. Пелецишин, К.О. Слобода // *Інформаційні системи та мережі: [збірник наукових праць] / відповідальний редактор В. В. Пасічник.* - Львів : Вид-во Національного університету "Львівська політехніка", 2011. - 387 с. : іл. - (Вісник / Національний університет "Львівська політехніка" ; № 699). – С. 352-361.
7. Федущко С. *Методи визначення соціально-демографічних характеристик користувачів соціальних комунікацій* / С. Федущко, А. Пелецишин, Ю. Серов // *Комп'ютерні науки та інженерія (CSE-2011): тези V міжнародної науково-практичної конференції молодих вчених, 24-26 листопада 2011 р., Львів.* – Львів: Вид-во НУ „Львівська політехніка”, 2011.– С.358-361.
8. Федущко С.С. *Алгоритм валідації інтернет-імен користувачів віртуальних спільнот* / С.С. Федущко, Ю.О. Серов // *Вісник НУ "Львівська політехніка": Комп'ютерні науки та інформаційні технології.* – Львів: Вид-во НУ "ЛП", 2011. – №719. – С.215-219.
9. Федущко С.С. *Особливості визначення та опису соціально-демографічних характеристик в соціальних комунікаціях* / С.С. Федущко // *Вісник Національного університету "Львівська політехніка": Комп'ютерні науки та інформаційні технології.* – Львів: Вид-во НУ "ЛП", 2011. – №694. – С.75-85.
10. Хміль І. *Побудова концептуальних сіток для аналізу тематичної близькості геопросторових об'єктів* // *Інформація, комунікація, суспільство: матеріали III Міжнародної наукової конференції ІКС-2014, 21-24 травня, 2014 року, Львів-Славське.* – Львів: Вид. Львівської політехніки, 2014. – С. 200-201.
11. Хміль І. *Формування пошукових запитів для опрацювання інформації у геоінформаційних системах* // *Інформація, комунікація, суспільство: матеріали I Міжнародної наукової конференції ІКС-2012, 25-28 квітня, 2012 року, Львів.* – Львів: Вид. Львівської політехніки, 2012. – С. 122-123.

Стаття надійшла до редколегії 24.12.2014

Рецензент: д-р техн. наук, проф. А.М. Пелецишин, Національний університет «Львівська політехніка», Львів.

Хміль Ирина Александровна

Аспирант кафедри соціальних комунікацій и інформаційної діяльності
 Национальний університет «Львовская політехніка», Львов

**АНАЛИЗ ВЗАИМОДЕЙСТВИЯ СОСТАВЛЯЮЩИХ ИНФОРМАЦИОННОГО ОБРАЗА
ГЕОПРОСТРАНСТВЕННЫХ ОБЪЕКТОВ В ВИРТУАЛЬНЫХ СООБЩЕСТВАХ**

Аннотация. Рассмотрено взаимодействие субъектов, формирующих информационный образ геопространственных объектов в виртуальном сообществе в реальной и виртуальной средах. Также описана формальная модель корпуса релевантных текстов, экстрагированных из информационного наполнения виртуальных сообществ с помощью словаря синонимов названий геопространственных объектов.

Ключевые слова: геопространственный объект; виртуальное сообщество; информационный образ; модель; корпус текстов

Khmil Irina

Aspirant kafedry sotsial'nykh kommunikatsiy i informatsionnoy deyatel'nosti
Natsional'nyu universitet «L'vovskaya politekhnika», L'vov

**ANALYSIS OF THE INTERACTION OF GEOSPATIAL OBJECTS' IMAGE INFORMATION
COMPONENTS IN VIRTUAL COMMUNITIES**

Abstract. The article considers the interaction of components of an information image of the author as the subject of previous experience, geospatial object, which is the main object of the formation of the previous experience and virtual communities as a tool for generating and obtaining relevant to geospatial objects text corpus. We analyzed the result of the interaction of two media: a real-world environment, which is formed by the primary and secondary experience of the author in relation to geospatial object and the virtual, the Internet, where a corresponding array of information displays any previous experience of the author, is structured according to the content of the service in which it is located, but does not have the clear classification, the author and his characteristics are not identified. Also we described a formal model of the relevant body of texts that is extravagant of content virtual communities through the use of pre-formed dictionary of synonyms of the names of geospatial objects. The model highlights the main (body text, consisting of descriptors, markers and other auxiliary parts) and additional (previously declared info about the author, which is available from the online community) components that are involved in further shaping the image information and its analysis.

Keywords: geospatial object; virtual community; information image; model; texts` corpus

References

1. Johnson, T. J. (2002). *Webelievability: A Path Model Examining How Convenience and Reliance Predict Online Credibility* / T. J. Johnson and B. Kaye // *Journalism and Mass Communication Quarterly* (79). – 619-642.
2. Oudshoff, A.M. (2003). *Knowledge discovery in virtual community texts: Clustering virtual communities* / Oudshoff A.M., Bosloper I.E., Klos T.B., Spaanenburg L. // *Journal of Intelligent & Fuzzy Systems*. - Vol.14.- №1.- 13-24.
3. Pekar, V. (2008). *Discovery of Subjective Evaluation of Product Features in Hotel Reviews* / Pekar V., Ou Sh. // *Journal of Vacation Marketing*. – Vol. 14. – № 2. – 145 – 155.
4. Saur, i R. (2009). *FactBank: a corpus annotated with event factuality*/ Sauri R, Pustejovsky J. // *Language Resources and Evaluation*. - Vol.43. - №3. –227-268.
5. Zhezhnych, P. I. (2007). *Temporal databases: models and methods of implementation: Lviv Polytechnic National University Publishing house*, 260.
6. Tymovchak-Maksymets, O.Yu. (2011). *Analysis of communicative interaction on web forum: information behavior and participants* / Tymovchak-Maksymets, O.Yu., Peleshchyshyn, A. M., Sloboda, K.O. // *Information Systems and Networks*, 699. – 352-361.
7. Fedushko, S. S. (2011). *Methods for determining of the social communications users socio-demographic characteristics*/ Fedushko, S. S., Peleshchyshyn, A. M., Syerov, Yu. O. // *Computer Science and Engineering (CSE-2011)*. – 358-361.
8. Fedushko, S. S. (2011). *Algorithm Validation online usernames virtual communities*/ Fedushko, S. S., Syerov, Yu. O. // *Computer Science and Information Technologies*. – №719. – C.215-219.
9. Fedushko, S. S. (2011). *Features of definition and description of the socio-demographic characteristics in social communications*. *Computer Science and Information Technologies*, 694. – 75-85.
10. Khmil, I. (2014). *Building a conceptual grids for analysis of geospatial objects thematic proximity*. *Information, communication, society: materials III International conference ICS-2014*. – 200-201.
11. Khmil, I. (2012). *Search queries forming for information processing in GIS*. *Information, communication, society: materials I International conference ICS-2012*. – 122-123.

Посилання на публікацію

- APA Khmil, I. O. (2015). *Analysis of the interaction of geospatial objects` image information components in virtual communities*. *Management of Development of Complex Systems*, Issue 21, P. 87 – 91 [in Ukrainian].
- ГОСТ Хміль І.О. *Аналіз взаємодії складових інформаційного образу геопросторових об'єктів у віртуальних спільнотах [Текст] / І.О. Хміль // Управління розвитком складних систем. – 2015. - № 21. – С. 87 - 91.*