

ОСОБЛИВОСТІ АРХІТЕКТУРИ ІНТЕРНЕТ СИСТЕМИ УПРАВЛІННЯ КОМЕРЦІЙНИМ КОНТЕНТОМ НА ОСНОВІ МЕТОДІВ MACHINE LEARNING, WEB MINING ТА SEO-ТЕХНОЛОГІЙ

Висоцька В. А. – канд. техн. наук, доцент, доцент кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

Демчук А. Б. – канд. техн. наук, асистент кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

Литвин В. В. – д-р техн. наук, професор, завідувач кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

АНОТАЦІЯ

Актуальність. Сьогодні більшість корпорацій постійно переосмислює бізнес з точки зору можливостей Інтернет, а саме його доступність, широке охоплення і постійно мінливі потреби користувача. Web-ресурс е-комерції, який забезпечує зручний для користувача досвід, зокрема, можливість швидко знаходити необхідні згідно його портеб та смаку товари, більше підтримує конкурентні переваги.

Метою дослідження є розроблення загальної архітектури інтелектуальної системи поширення комерційного контенту в Інтернет-просторі на основі навчання нейронної мережі згідно історії постійної аудиторії для подачі унікального контенту з використанням підходу персоналізації та використання тегів.

Метод. Розроблено модель інформаційної системи персоналізації комерційного контенту згідно потреб користувача. Також розроблено метод поширення комерційного контенту на основі підходу персоналізації та використання тегів. При цьому використано навчання нейронної мережі для створення тег рекомендацій та доступні на ринку засоби персоналізації. Розроблений алгоритм персоналізації дозволяє пов'язати кожного користувача з списком продуктів, які найімовірніше його зацікавлять, а також може прогнозувати те, що клієнти можуть хотіти бачити, навіть якщо вони ще не знають про це. Розроблений метод можна використати для забезпечення більш релевантного набору контенту. Також розроблений метод дає можливість класифікувати відповідний контент або показати його раніше в процесі гортання сторінок для уникнення споживачами вибору неправильного контенту або витрати часу на прокручування при пошуку товару.

Результати. Розроблена система призначена для поширення продуктів інформаційних технологій (публікацій, книг, курсів, відео, файлів тощо) за допомогою Інтернет.

Висновки. Інтернет. Впровадження цієї системи дасть змогу отримувати доступ до певного роду контенту широкому загалу користувачів, адже сайт буде розміщено у всесвітній павутині, з другого боку інша частина мети створення цієї системи є комерційна складова, а саме отримання прибутків власником чи адміністратором інтелектуальної системи, через механізми е-комерції.

КЛЮЧОВІ СЛОВА: комерційний контент, персоналізація, Web mining, Machine Learning, SEO-технологія, метрики пошуку, електронна комерція, NLP, контент-моніторинг, контент-аналіз, статистичний лінгвістичний аналіз, квантитативна лінгвістика.

АБРЕВІАТУРА

БЗ – база знань;
ІА – інтелектуальний агент;
ІС – інформаційна система;
ІТ – інформаційна технологія;
ПО – предметна область.

НОМЕНКЛАТУРА

S_{ec} – модель системи персоналізації контенту згідно потреб та вподобань користувача;

$IdnU_{sr}$ – колекція правил ідентифікації користувача;

$CntRvw$ – колекція правил перегляду контенту;

$PchPcs$ – колекція правил оформлення замовлення;

$PdtPmt$ – колекція правил оплати замовлення;

α – оператор збору даних при перегляді контенту користувачем;

β – оператор пошукових тегів;

γ – оператор аналізу отриманих даних після збору даних при перегляді контенту;

$AdgSchTgs$ – колекція правил додавання пошукових тегів;

$RvwRzt$ – колекція правил формування результатів перегляду контенту користувачем;

$PdtSch$ – колекція правил пошуку продукту згідно персоналізації потреб користувача;

$PdtLstOr$ – колекція правил перегляду списку продуктів;

$AngPdt$ – колекція правил додавання в список персоналізованого продукту;

$RmdPdt$ – колекція правил перегляду персоналізованих продуктів;

$AngLst$ – колекція правил додавання продукту в список побажань;

$AngCrt$ – колекція правил додавання в кошик;

$PdtDsr$ – колекція правил опису продуктів;

δ – оператор додавання в список персоналізованого продукту;

χ – оператор роботи нейромережі аналізу рекомендованих тегів;

NrlNwRmd – нейромережа рекомендованих тегів;

SchRzt – колекція правил формування результатів пошуку на основі існуючих тегів та категорій;

ϕ – оператор формування результатів пошуку;

PsnCnt – множина персоналізованого контенту;

φ – оператор;

ChgCnt – колекція зміненого контенту на сайті;

μ – оператор аналізу збережених тегів та категорій в Sitecore;

SvdTgs – множина збережених тегів та категорій;

ν – оператор формування множини персоналізованого контенту;

UstHst – колекція параметрів історії перегляду користувача;

Recall – повнота пошуку;

Precision – точність пошуку;

Accuracy – акуратність пошуку;

AvgPrec – середня точність пошуку;

Error – помилка пошуку;

F-measure – F-міра пошуку;

Precision(n) – точність на рівні n документів пошуку;

R-precision – R-точність пошуку;

a – кількість документів, знайдених системою і релевантних з точки зору експертів;

b – кількість документів, знайдених системою, але нерелевантних з точки зору експертів;

c – кількість релевантних документів, не знайдених системою;

d – кількість нерелевантних документів, не знайдених системою;

P – список всіх можливих морфо-синтаксичних описів для даного слова;

n – довжина кодування морфо-синтаксичних описів (110 біт);

o – вихід нейронної мережі для поточного слова;

e – двійкове кодування для морфо-синтаксичних описів в P .

ВСТУП

Сьогодні ефективність функцій пошуку та навігація по Web-сайту е-комерції зростає експоненційно [1]. Це є важливою складовою успішної стратегії е-комерції [2]. Учасники Інтернет-ринку збільшують інвестиції в персоналізацію контенту на Web-ресурсі для отримання низки переваг в е-бізнесі [3]. Ключові слова від користувачів Web-ресурсу е-комерції при пошуці в реальному часі формують профіль поведінки постійних/потенційних клієнтів і надають неоціненні статистичні дані для подальших досліджень [4]. Вдало використана така інформація збільшує кількість продаж, покращує механізми

утримання клієнтів та збільшує обсяги потенційних і постійних користувачів Web-ресурсу е-комерції [5–8].

Метою ІС е-комерції є формування унікального контенту на основі підходу персоналізації та використання тегів [9–10]. Об'єктом дослідження є використання нейронних мереж для створення тегів рекомендацій на основі методів персоналізації даних. Предметом дослідження є метод формування рекомендацій кінцевому користувачу Web-ресурсу е-комерції на основі персоналізації даних.

Метою дослідження є розроблення загальної архітектури ІС поширення комерційного контенту в Інтернет на основі навчання нейронної мережі на основі методів персоналізації даних.

1 ПОСТАНОВКА ПРОБЛЕМИ

Сучасні успішні проекти у вигляді ІС е-комерції зазвичай використовують персоналізацію на Web-сайті на відмінну від стандартної навігації [11]. Загальну модель ІС е-комерції подамо як кортеж:

$$S_{ec} = \langle IdnUsr, CntRvw, PchPcs, PdtPmt, \alpha, \beta, \gamma \rangle.$$

Коли люди здійснюють покупки з намірами (знають, чого хочуть), і мають можливість швидко та легко це знайти, то вони, швидше за все, здійснять цю покупку [12–14]. Якщо користувачу презентувати відразу шуканий товар по категоріях та розташувати його в блоках контенту на головній сторінці, шанси на привернення уваги різко збільшуються. Тому процес персоналізованого перегляду контенту, рекомендованого конкретному користувачу згідно його побажань, залежить від декількох факторів:

$$CntRvw = \alpha(IdnUsr, AdgSchTgs, RvwRzt).$$

Персоналізований підхід до користувача Web-ресурсу е-комерції призводить до більш високого коефіцієнта продаж. Клієнти, які не можуть знайти необхідну інформацію, зазвичай залишають Web-ресурс незадоволеними і шукають найкращі альтернативи [15]. Необхідно звертатися до Machine Learning для покращення формування рекомендацій кінцевому користувачу Web-ресурсу на основі результатів пошуку при кожному відвідуванні через періодичне поповнення бази пошукових тегів:

$$AdgSchTgs = \beta(PdtCrt, AgrRmdTgs).$$

Застосування технології Machine Learning дозволяє генерувати рейтинг пошуку, відсортований за релевантністю чи оціночною актуальністю:

$$RvwRzt = \gamma(PdtSch, PdtLstOrw, AngPdt, RmdPdt, AngLst, AngCrt).$$

Ця оцінка враховує специфічні пошукові терміни (теги), а також особливості конкретного профілю користувача (наприклад, віковий діапазон, попередні замовлення, попередні пошукові терміни) [16]:

$$AngPdt = \delta(PdtDsr, \chi(NrlNwkRmd)).$$

Алгоритми персоналізації дозволяють пов'язати кожного користувача зі списком найімовірніше необхідних товарів, поповнюючи альтернативними товарами, які не були об'єктом пошуку [17]:

$$SchRzt = \phi(PdtSch, PsnCnt, \phi(ChgCnt), \mu(SvdTgs)).$$

Розпізнавання контексту запиту користувачів за допомогою глибоких нейронних мереж забезпечує автоматичне додавання тегів в описи товару Web-ресурсу. Ці методи використовують для класифікації міміки і розпізнавати емоції користувача [18]:

$$PsnCnt = v(UstHst, SvdTgs, ChgCnt).$$

Кожний користувач має власні вимоги до формування запиту. Але типові пошукові системи повертають один і той самий результат для одного запиту, поданого різними користувачами. Для вирішення проблеми інформаційного перевантаження та надання користувачам необхідного контенту використовують персоналізацію. Вона підвищує точність пошуку, спрощує процес пошуку, зберігає час та надає необхідний контент користувачам. Персоналізація створює відчуття індивідуальності та унікальності. Шляхом сегментації та націлювання на різних покупців, персоналізація відповідає різним потребам кожного користувача Web-ресурсу.

2 АНАЛІЗ ЛІТЕРАТУРНИХ ДЖЕРЕЛ

Ефективність реалізації штучного інтелекту для персоналізації передбачення побажань користувача в результатах пошуку товару значно залежить від роботи вбудованої в Web-ресурс інформаційно-пошукової системи. Для її стандартного оцінення використовують компоненти [19–26] як:

- Колекція документів.
- Набір тестових інформаційних потреб, поданих у вигляді запитів користувачів.

– Набір оцінок релевантності, поданих через бінарні твердження релевантний-нерелевантний відносно кожної пари запит-контент.

Релевантність є мірою відповідності отриманого результату бажаному, тобто результатам пошуку відповідно до запиту [1–7, 12–18]. Колекція контенту і набір запитів повинні мати достатній обсяг: чим більша тестова вибірка, тим точніша оцінка якості роботи алгоритму [5–8, 27–32]. Інтерпретації тих або інших оцінок часто відмінні [33–35]. Більшість метрик в сучасній оцінці текстового пошуку ґрунтуються на матриці класифікації [1–5, 36–39]:

Recall характеризує здатність системи знаходити необхідний користувачу контент, але не враховує кількість виданого нерелевантного контенту [1–5, 40]. Наприклад, якщо *Recall* = 50%, то половина релевантного контенту системою не знайдена.

$$Recall = \frac{a}{a+c}.$$

Precision характеризує здатність системи видавати в списку результатів тільки релевантний контент [1–5, 41]. Наприклад, якщо *Precision* = 50%, то серед знайденого контенту половина релевантного.

$$Precision = \frac{a}{a+b}.$$

Для *Accuracy* передбачається, що система приймає рішення про приналежність до цієї категорії для кожного контенту колекції [1–5, 42].

$$Accuracy = \frac{a+d}{a+b+c+d}.$$

Error обчислюється як [43]:

$$Error = \frac{b+c}{a+b+c+d}.$$

F-measure використовують як єдину метрику, що об'єднує метрики *Recall* і *Precision* [1–5, 44]:

$$F = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}.$$

Відмітимо основні властивості метрики *F* [45]:

- $0 \leq F \leq 1$;
- якщо *Recall* = 0 або *Precision* = 0, то *F* = 0;
- *Recall* = *Precision*, то *F* = *Recall* = *Precision*;
- $\min(Recall, Precision) \leq F \leq \frac{Recall + Precision}{2}$.

Загальна формула для *F-measure* обчислюється як:

$$F_{\beta} = (1 + \beta^2) \cdot \frac{Precision \cdot Recall}{(\beta^2 \cdot Precision) + Recall}.$$

Для *AvgPrec* розглядають дві послідовності дій:

- вичислити метрики по кожному запиту окремо і потім їх усереднити (macroaverage);
- знайти загальну кількість документів, що відносяться до категорій табл. 1 і вже на їх основі вичислити шукану метрику (microaverage).

Таблиця 1 – Основні категорії документів

Кількість контенту	Релевантного	Не релевантного
виданого по запиту	<i>a</i> (правильно виданого контенту)	<i>c</i> (виданого нерелевантного даній рубриці контенту)
не виданого по запиту	<i>b</i> (неправильно виданого контенту)	<i>d</i> (не виданого і нерелевантного контенту)



Рисунок 1 – Категорії документів

Величина macroaverage характерна для оцінки завдань пошуку, в яких важливий результат в середньому за запитом, незалежно від потужності відповіді на цей запит. Величина microaverage знайшла більше застосування в оцінці класифікації і фільтрації, де необхідно враховувати обсяги запитів.

Метрики на послідовностях документів, відсортованого по релевантності, враховують не лише факт наявності документу в списку знайдених документів, але і його положення в цьому списку.

$Precision(n)$ визначається як кількість релевантних документів серед перших n виданих документів, ділене на n [1–5]. Якщо система видала більше за n документів, то $Precision(n)$ дорівнює точності системи на перших n документах результатів запити. Якщо система видала менш n документів, то $Precision(n)$ буде не вища за точність системи. $Precision(n)$ характеризує здатність системи видавати релевантні документи на початку списку результатів. Наприклад, якщо система видає не більше 10 документів на першій сторінці, то $Precision(n)$ відбиває якість результатів системи, що отримуються на першій сторінці. Ця метрика має ряд недоліків. Зокрема, для різних запитів метрики $Precision(n)$ можуть бути незрівняні. Наприклад, для ідеальної системи, яка видає тільки релевантні документи, $Precision(100)=0.2$ для запити, по якому існує 20 релевантних документів, і $Precision(100)=0,3$ для запити, по якому існує 30 релевантних документів. Але $Precision(n)$ є незамінною метрикою сучасних систем пошуку оскільки, зокрема, дозволяє оцінити корисність першої сторінки відповіді системи для користувача.

R -precision дорівнює точності на рівні n документів для n рівної кількості релевантних документів для цього запити [1–5]. Ця метрика покликана замінити $Precision(n)$ в тих випадках, коли необхідно врахувати велику різницю у кількості релевантних документів різних запитів. $AvgPrec$ для цього запити визначається таким чином [1–5]: нехай для цього запити є k релевантних документів. Точність на рівні i -го релевантного документу $prec_rel(i) = Precision(pos(i))$, якщо i -й релевантний документ знаходиться в результатах запити на позиції $pos(i)$. Якщо i -й релевантний документ не знайдений, то $prec_rel(i)=0$. $AvgPrec$ для цього запити дорівнює середньому значенню величини $prec_rel(i)$ по усім k релевантним документам [1–5, 46]:

$$AvgPrec = \frac{1}{k} \sum_{i=1}^k prec_rel(i).$$

Основні властивості метрики $AvgPrec$:

- $AvgPrec \leq Recall$;
- якщо релевантні документи знаходяться тільки на початку списку, то $AvgPrec \approx Recall$;
- якщо релевантні документи рівномірно розподілені в списку, то $AvgPrec \approx Precision \cdot Recall$;
- кількість документів, які ранговані нижче останнього релевантного, не впливає на значення.

$AvgPrec$ дозволяє оцінювати якість роботи системи, враховуючи пріоритет високо ранжированих документів перед документами, що знаходяться у кінці списку. На відміну від метрик $Precision(n)$ і R -precision, $AvgPrec$ враховує усі знайдені документи.

3 МАТЕРІАЛИ ТА МЕТОДИ

Покращити персоналізацію можна в два етапи.

1. Сгенерувати найбільш релевантну персоналізовану колекцію контенту.

2. Класифікувати відповідний персоналізований контент за потребами користувача та демонструвати його в зручний спосіб в процесі гортання сторінок для уникнення вибору нерелевантного контенту або втрати часу на пошук релевантного контенту.

Користувач Web-ресурсу користується послугами е-комерції для здійснення покупок для зручності, економії часу та зусиль. Покращення умов співпраці з кожним кінцевим користувачем значно спростить процес ведення е-бізнесу та зменшить зусилля користувачу на пошук необхідного товару (рис. 2).

Блок реєстрації клієнтів дозволяє отримати знижки або запрошення на закриті продажі. Блок перегляду товару включає додавання пошукових тегів. Також клієнт може шукати товари, переглядати список продуктів, переглядати рекомендації, додавати продукти до кошику або списку бажань.

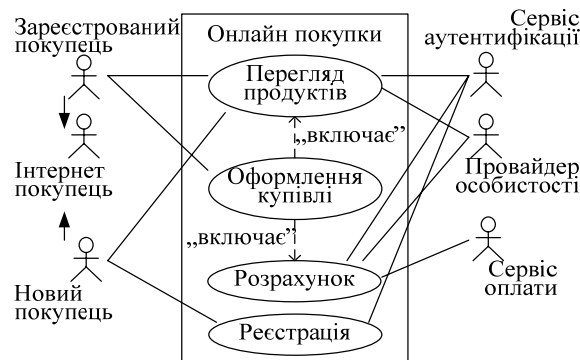


Рисунок 2 – Use case діаграма для Web-ресурсу

Блок додавання пошукових тегів розширюється за допомогою автогенерування рекомендованих тегів, та включає створення товару (рис. 3). Блок автентифікації клієнтів включає блок перегляду рекомендованих продуктів та доданих до списку бажань, оскільки обидва вимагають перевірки автентичності клієнта. У той же час товар може бути доданий до кошика для покупок без перевірки автентичності користувача. Система прийняття рішень для Web-сайту е-комерції при додаванні контент-модератором кожного нового товару забезпечує надання розширеного опису продуктів. Також надає відповідні рекомендовані категорії та теги з використанням синонімічних варіантів. Синонімічний ряд визначається через нейронну мережу та алгоритми Machine Learning, наприклад, на основі .Net CMS Sitecore (рис. 4), яка володіє засобами персоналізації, доступними для розширення і додавання власних правил, базуючись на доступному core-функціоналі (рис. 5). Кожний клієнт має унікальний ідентифікатор, пов'язаний з точно одним обліковим записом (рис. 6). Web-користувач

може перебувати в кількох станах: нові, активні, тимчасово заблоковані або заборонені, і бути пов'язаними з кошиком для покупок. Клієнт може не мати замовлень чи історії попередніх пошуків, чи переглядів сторінок. Замовлення клієнтів сортуються та є унікальними. Продукти мають теги, які їх позначають, а також категорії, до яких відносяться. Кожен продукт може відноситись до одної категорії, та мати багато тегів.

Персоналізація є методом відображення цільового, релевантного контенту для користувачів з урахуванням їх характеристик і поведінки, наприклад, місцезнаходження, статі, контенту і/або попередніх візитів. Через персоналізацію необхідний контент досягає відповідних користувачів, наприклад, можна:

– показати інший контент для різних користувачів на основі даних геолокації;

– сховати реєстраційну форму користувачів, які раніше заповнили форму;

– змінити текст на Web-сайті банера через посилання на Web-сайт користувача.

Для умовної візуалізації контенту використовують умовні відтінки, щоб контролювати, як відвідувачі переглядають та взаємодіють з Web-ресурсом.

Персоналізація на відміну від умовної візуалізації відноситься до широкого процесу доставки цілеспрямованого необхідного контенту для відповідних користувачів. Персоналізація включає:

– адаптивну/динамічну зміну контенту Web-ресурсу на основі поведінки користувача;

– на основі створення та впровадження правил умовного відтворення контенту Web-ресурсу.

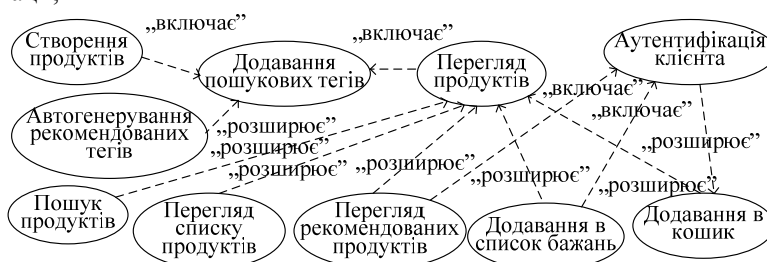


Рисунок 3 – Use case діаграма для блоку перегляд продуктів

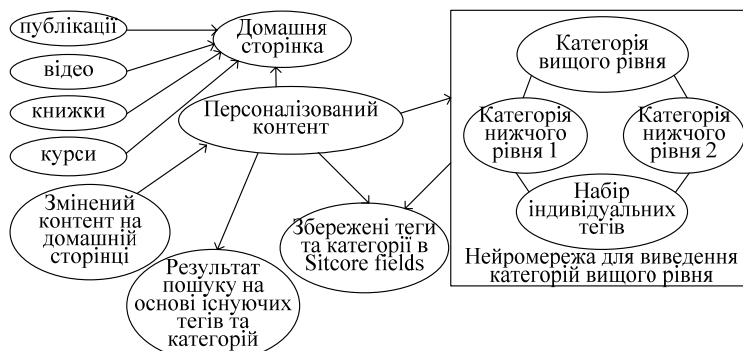


Рисунок 4 – Use case діаграма для процесу персоналізації контенту

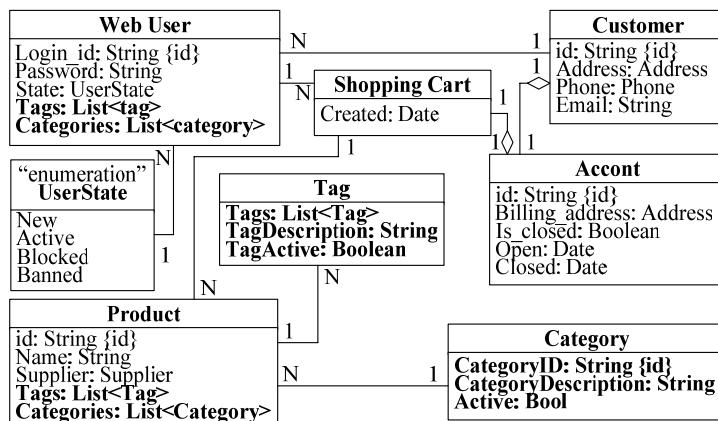


Рисунок 6 – ER-діаграма для для Web-ресурсу

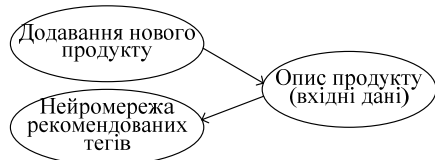


Рисунок 5 – Use case для процесу додавання продукту

В Sitecore встановлюють умови для умовної візуалізації в редакторі правил набору на основі логіки для визначення істинності умови. Визначають також дії як умовний рендерінг, який набуває чинності, якщо умова дійсна. Умова, пов'язана з дією, є правилом. Доступ до редактора коду правил отримують з редактора подій, редактора контенту та панелі керування маркетингом, але зазвичай створюють умовні виправлення в редакторі досвіду. Профілі контенту є категоріями для відстеження поведінки користувача під час навігації по Web-ресурсу е-комерції. Він допомагає краще зрозуміти поведінку, дії та інтереси користувачів. Вміст профілів складаються з ключів профілю; значення профілю; профільної картки. Також створюють в Sitecore власні властивості для відображення типових користувачів, яким також призначають картки профілю. Користувачів використовують для здійснення персоналізації правил. Створюють профілі контенту, ключі профілю, профілі та особи в центрі маркетингу. Значення профілю для елементів призначають у редакторі досвіду. Ключі профілю описують різні аспекти профілів. Призначають цифрові значення профілю для ключів профілю, а потім використовують значення профілю для відстеження взаємодії користувачів з Web-ресурсом е-комерції. Sitecore має деякі попередньо визначені профілі, які вже мають призначені профілі. Також можна створити власні ключі профілю. Коли користувачі здійснюють навігацію через Web-ресурс, їм призначаються значення профілю контенту, які визначаються для кожного відвіданого елемента. Ці значення накопичуються при повторних відвідуваннях ресурсу користувачем. Це допомагає створити контактний профіль постійних користувачів. Інформація про дії користувачів (переглянуті сторінки, виконані цілі, і шлях навігації по Web-ресурсу, тощо) визначає ділянки Web-ресурсу для вдосконалення. Також це використовують для сегментування користувачів та створення правил персоналізації для продажу. Наприклад, якщо користувач часто здійснює навігацію у певних категоріях або досяг високого профілю, то він є потенційною клієнтом цього напрямку. Тоді вносять у систему CRM як потенційного покупця або надсилають йому е-повідомлення.

Картки профілю містять збережені ключі профілю та значення профілю. Профільні картки використовують для призначення стандартних значень профілів для елементів Web-ресурсу. Під час налаштування профілів для контенту створюють осіб.

Це вигадані персонажі як певні типи користувачів у межах цільової демографічної групи. Особи описують життя, вік, звички, передумови, інтереси та професію вигаданого персонажа, який може використовувати Web-ресурс е-комерції певним чином. Створюють профільні карти, які описують спосіб, яким особа споживає контент Web-ресурсу. Вихідними даними профілів користувачів Web-ресурсу в Sitecore є вхідними даними для нейронної мережі в Machine Learning для персоналізації при формуванні рекомендацій кінцевому клієнту е-комерції.

Введення тексту в нейронній мережі особливо складне. Є певні проблеми, з якими необхідно боротися. Окремі слова мають різну довжину. Нейронні мережі вимагають фіксованого вхідного та вихідного розмірів. Використання рекурентної нейронної мережі, наприклад, нейронної мережі Елмана, вирішує частину проблем при розпізнаванні літер. Вона використовує контекстний шар для запам'ятовування замовлення. Опрацювання тексту використовує один довгий потік повідомлень. Кожен вхід є кількістю одного конкретного слова. Весь вхідний вектор міститиме одне значення для кожного унікального слова, наприклад, «з брюками і без», «двоє спортивних брюк», «темний спортивний костюм» та «брюки та чорні брюки». Є такі унікальні слова. Це словник.

- Input 0: та
- Input 1: темний
- Input 2: костюм
- Input 3: спортивний
- Input 4: без
- Input 5: брюки
- Input 6: чорний
- Input 7: з
- Input 8: два

Рядки кодуєть наступним чином, заповнюючи пропущені слова з нулем:

- «з брюками і без» [0 4 5 7] → [1, 0, 0, 0, 1, 1, 0, 1, 0];
- «двоє спортивних брюк» [1 5 8] → [0, 1, 0, 0, 0, 1, 0, 0, 1];
- «темний спортивний костюм» [1 2 3] → [0, 1, 1, 1, 0, 0, 0, 0, 0];
- «брюки та чорні брюки» [0 5 6] → [1, 0, 0, 0, 0, 2, 1, 0, 0].

Тепер є постійна векторна довжина. Дев'ять – загальна кількість слів у словнику. Кожен номер компонента у векторі – індекс у словнику доступних слів. На кожному векторному компоненті зберігається підрахунок кількості слів для цього словника. Кожен рядок зазвичай містить лише невелику підмножину словника. Як результат, більшість векторних значень є нульовим. Один з найважливіх аспектів програмування Machine Learning перекладає проблему на масив фіксованої кількості чисел з плаваючою точкою.

4 ЕКСПЕРИМЕНТИ

Після збору персоналізованих даних користувачів будемо нейронну мережу на прикладі хеш-таблиці. Її використовують для відображення ключів до значень, наприклад, «чуєш» → «сприймається вухом»; «бігти» → «йти швидше, ніж прогулянка»; «написати» → «для формування інструментом зображення на поверхні».

Це відображення між словами та визначенням кожного слова. Використовують ключ рядка для іншого значення рядка. Ключ повертає значення. Нейронна мережа є двонаправленою асоціативною пам'яттю, тобто фактично дозволяє також передавати значення і отримувати ключі. Наприклад, шаблон, який надсилається на вхідний шар нейронної мережі, подібний на процес введення ключа до хеш-таблиці. Але нейронні мережі не повертають порожній результат, а знаходять найближче значення.

Якщо задати ідеальний результат, то це контрольоване навчання і навпаки. Наглядова підготовка вчить нейронну мережу виробляти ідеальний вихід. Непідконтрольний тренінг зазвичай вчить нейронну мережу групувати вхідні дані в декілька груп, визначених вихідним числом нейронів. Наглядова та безконтрольна підготовка є ітераційним процесом. Для керованого тренування кожна ітерація тренування обчислює наскільки близький фактичний результат до ідеального виходу. Ця близькість виражається як відсоток помилки. Кожна ітерація змінює внутрішні матриці ваги нейронної мережі, щоб отримати рівень помилки на досить низькому рівні. Непідконтрольний тренінг також є ітераційним процесом. Проте обчислення помилки не таке просте. Немає очікуваного виходу, тому не можна виміряти, наскільки непідконтрольна нейронна мережа від ідеального виходу, бо немає ідеального виходу. Часто повторюють кілька ітерацій, а потім використовують мережу. Якщо потрібні додаткові тренування, то це навчання. Інший дуже важливий аспект вищезазначених тренувальних даних полягає у можливості використання в будь-якому порядку.

Тегер при Machine Learning автоматично призначає чотири фіктивних теги (два на початку і два наприкінці) для цільового висловлювання. Тоді нейронна мережа навчається автоматично призначати морфо-синтаксичні описи, що враховує контекст, тобто два раніше призначені теги та можливі теги для поточних та наступних двох слів. Навчальний приклад складається з функцій, витягнутих для одного слова всередині висловлювання як введення. Це морфо-синтаксичні описи в межах цього виразу як виведення. Функції витягуються з 5 слів, орієнтованих на поточне слово. Одне слово характеризується вектором, який кодує його морфо-синтаксичний опис. Для кодування можливих морфо-синтаксичних описів використовуємо $P(a|w)$, де кожен можливий атрибут має одну відповідну позицію усередині закодованого вектора. Вектори використовують для кодування можливих морфо-синтаксичних описів для поточного слова та наступних двох слів.

$$P(a|w) = \frac{C(w,a)}{C(w)}$$

Під час навчання обчислюємо список суфіксів з відповідними морфо-синтаксичними описами, які використовують під час виконання, щоб створити

можливий морфо-синтаксичний вектор для невідомих слів. Коли такі слова зустрічаються в даних тесту, наближаємо їх можливий морфо-синтаксичний вектор, використовуючи метод Брантса. Коли тегер застосовують до нового висловлювання, система ітеративно обчислює вихідний морфо-синтаксичний опис для кожного окремого слова. Після того, як мітку присвоєно одному слову, пов'язаний із цим словом вектор змінюється таким чином, що буде мати значення 1 для кожного атрибута, присутнього у його знову призначеному морфо-синтаксичному описі. Як наслідок кодування кожного окремого атрибута окремо для морфо-синтаксичних описів, тегер призначає нові теги, які ніколи не були пов'язані з поточним словом у навчальному процесі. Хоча це є недоліком для роботи з невідомими словами. Тоді використовуємо додатковий список слів із їх дозволеними морфо-синтаксичними описами. Для слова список розраховується як об'єднання з усіх морфо-синтаксичних описів, що з'являються з суфіксами, які застосовують до цього слова. Коли тегер призначає морфо-синтаксичний опис до певного слова, обирає один з морфо-синтаксичних описів можливих словоформ у списку зі словом, використовуючи функцію відстані:

$$\min_{e \in P} \sum_{k=0}^n |o_k - e_k|$$

5 РЕЗУЛЬТАТИ

Розглянемо аналіз імен людей та місць у тексті для автоматичного додавання тегів про новини з людьми та назвами місць, що містяться в статтях. Особливість для ідентифікації назв і місць у тексті – це дані у файлі test data / propername.ser – спеціальний файл даних Java, що містить хеш-таблиці для людей і назв місць. Ці дані читаються в конструкторі класу Names.

Значення хеш-таблиць використовується так:

```
while (keysE.hasMoreElements()) {  
    Object key = keysE.nextElement();  
    System.out.println(key+": "+placeNameHash.get(key)); }
```

Які виведуть наступне:

```
Mauritius : country  
Port-Vila : country_capital  
Hutchinson : us_city  
Mississippi : us_state  
Lithuania : country
```

Наступний приклад використовує методи

```
isPlaceName, isHumanName, та getProperNames:  
System.out.println("Los Angeles: " +  
names.isPlaceName("Los Angeles"));  
System.out.println("President Bush: " +  
names.isHumanName("President Bush"));  
System.out.println("President George Bush: " +  
names.isHumanName("President George Bush"));  
System.out.println("President George W. Bush: " +  
names.isHumanName("President George W. Bush"));  
ScoredList[] ret = names.getProperNames("George Bush  
played golf. President \George W. Bush went to London  
England, \and Mexico to see Mary \Smith in Moscow.  
President Bush will \return home Monday.");  
System.out.println("Human names: " +  
ret[0].getValuesAsString());  
System.out.println("Place names: " +  
ret[1].getValuesAsString());
```

Вихідні значення цього прикладу є таким:

```
Los Angeles: true
President Bush: true
President George Bush: true
President George W. Bush: true
* place name: London,
  placeNameHash.get(name): country_capital
* place name: Mexico,
  placeNameHash.get(name): country_capital
* place name: Moscow,
  placeNameHash.get(name): country_capital
Human names: George Bush:1,
  President George W . Bush:1,
  Mary Smith:1,
  President Bush:1
Place names: London:1, Mexico:1, Moscow:1
```

Методи `HumanName` і `isPlaceName` шукають рядок в хеш-таблиці з назвою людини або місця:

```
public boolean isPlaceName(String name){
    return placeNameHash.get(name) != null;}
```

Версії API, які опрацьовують імена, що містять кілька слів, є трохи складнішими. Необхідно побудувати рядок із слів між початковими та кінцевими індексами та перевірити, чи це нове значення рядка є дійсним ключем у хеш-таблицях імен людей або назв місць:

```
public boolean isPlaceName(List<String> words, int
startIndex, int numWords){
    if ((startIndex + numWords) > words.size()) {
        return false; }
    if (numWords == 1) {
        return isPlaceName(words.get(startIndex)); }
    String s = "";
    for (int i = startIndex;
        i < (startIndex + numWords); i++){
        if (i < (startIndex + numWords - 1)){
            s = s + words.get(startIndex) + " "; }
        else{ s = s + words.get(startIndex); } }
    return isPlaceName(s);}
```

Токенінг тексту є процесом розщеплення рядка до окремих токенів. В результаті цього відбувається зменшення кількості слів до скороченого кореня слова, що дозволяє легко порівнювати рівність подібних слів. Тегування є визначенням того, яка частина кожного слова знаходиться у вхідному тексті. Позначення міток ускладнюється багатьма словами, що мають різні частини мовлення залежно від контексту (наприклад, «bank the airplane», «the river bank», тощо). Перш ніж опрацьовувати будь-який текст, необхідно розбити на окремі токени (слова, цифри та пунктуаційні символи). Клас `Tokenizer` має два статичні способи, обидва беруть вхідний рядок для токенування та повертають список токенів. Другий спосіб має додатковий аргумент для визначення максимальної кількості токенів:

```
static public List<String> wordsToList(String s)
static public List<String> wordsToList(String s, int
maxR)
String text = "The ball, rolling quickly, went down the
hill.";
List<String> tokens = Tokenizer.wordsToList(text);
System.out.println(text);
for (String token : tokens)
System.out.print("\n"+token+"\n ");
System.out.println();
```

Цей фрагмент коду виводить наступне:

```
The ball, rolling quickly, went down the hill.
"The" "ball" ",," "rolling" "quickly" ",," "went" "down"
"the" "hill" ".,"
```

Краще витягнути токени слова для спрощення порівняння подібних слів. В класі є два зручні

інтерфейси API для створення рядка з декількох слів та для одого токена слова:

```
public List<String> stemString(String str)
public String stemOneWord(String word)
```

Приклади опису тегів:

Tag	Description	Examples
NN	singular noun	dog
NNS	plural noun	dogs
NNP	singular proper noun	California
NNPS	plural proper noun	Watsons
CC	conjunction	and, but, or
CD	cardinal number	one, two
DT	determiner	the, some
IN	preposition	of, in, by
JJ	adjective	large, small, green
JJR	comparative adjective	bigger
JJS	superlative adjective	biggest
PP	proper pronoun	I, he, you
RB	adverb	slowly
RBR	comparative adverb	slowest
RP	particle	up, off
VB	verb	eat
VBN	past participle verb	eaten
VBG	gerund verb	eating
VBZ	present verb	eats
WP	wh* pronoun	who, what
WDT	wh* determiner	which, that

`FastTag` використовує методи `Machine Learning` для визначення правила переходу для тегів тексту, використовуючи ручне позначення тексту як приклад навчання. Клас `Tagger` читає файлоу лексику як потік ресурсів або як локальний файл. Кожен рядок у файлі `lexicon.txt` проходить через метод `utility parseLine`, який опрацьовує вхідний рядок, використовуючи перший токен у рядку як хеш-ключ, і розміщає решту токенів у масиві, що є хеш-значенням. Отже, будемо опрацьовувати рядок «fair JJ NN RB» як хеш-ключ «fair», а хеш-значення – масив рядків: [«JJ», «NN», «RB»]. Коли `Tagger` опрацьовує список токенів слова, кожний токен звертається до хеш-таблиці та зберігає перший можливий тип тегів для цього слова. У прикладі слово «fair» буде призначено (можливо, тимчасово) тегом «JJ». Тепер є список токенів слова та пов'язаний список можливих типів тегів. Детально розглянемо rule 1: i – це змінна циклу в діапазоні $[0, \text{кількість токенів слова} - 1]$, а слово – поточне слово в індексі i :

```
// rule 1: DT, {VBD | VBP} --> DT, NN
if (i > 0 && ret.get(i - 1).equals("DT")) {
    if (word.equals("VBD") || word.equals("VBP") ||
word.equals("VB")) {ret.set(i, "NN"); } }
```

Англійською це правило позначає визначник (DT) в токени слова (індекс $i-1$ супроводжується минулим часом дієсловом (VBD)), або дієслово теперішнього часу (VBP), після чого змінюється тип тегу індексу i на «NN». Інші правила у короткому синтаксисі:

– rule 2: перетворити іменник на номер (CD), якщо «.» з'являється у слові;

– rule 3: перетворити іменник на дієприкметник в минулому часі, якщо `words.get(i)` закінчується на «ed»;

Можливості створення профілів теоретично необмежені. Після використання Chrome браузера в якості одного унікального клієнта, та здійснення купівельної активності на сайті розглянемо отримані результати профілю користувача. Підсумуємо отримані деталі профілю користувача:

- surfing behaviour: new visitors vs. returning visitors, websites visited, etc.;
- purchasing behaviour: popular brands, price ranges, related products, etc.;
- geographical data;
- demographics: age and gender;
- sociological data: family situation, profession, interests, etc.

Запустимо консольну програму для навчання нейромережі для отримання вихідних даних у вигляді таблиці наближеності до списку заданих в системі тегів для контенту (рис. 8).

```
C:\Windows\System32\cmd.exe
* ===== Tags recommendations report =====
* model = [W52/WSJTP.TRAIN.W5]
* testset = [/home/me/WSJTP/WSJTP.TEST]
* ===== TAGGING SUMMARY =====
#WORDS = 129654
#KNOWN = 126005 / 129654 --> (97.1856 %)
#UNKNOWN = 3649 / 129654 --> (2.8144 %)
#AMBIGUOUS = 45779 / 129654 --> (35.3086 %)
* ===== ACCURACY PER LEVEL OF AMBIGUITY =====
POS HITS TRIALS ACCURACY MFT-ACCURACY
-----
# 15 / 15 = 100 % 100 %
$ 943 / 943 = 100 % 100 %
€ 1044 / 1045 = 99.9043 % 99.0431 %
( 186 / 186 = 100 % 100 %
) 187 / 187 = 100 % 100 %
* 6876 / 6876 = 100 % 100 %
. 5381 / 5381 = 100 % 100 %
: 752 / 752 = 100 % 100 %
CC 3237 / 3250 = 99.6 % 99.5692 %
CD 4789 / 4823 = 99.295 % 90.6075 %
DT 11117 / 11183 = 99.4098 % 98.453 %
EX 126 / 126 = 100 % 100 %
FW 7 / 30 = 23.3333 % 20 %
IN 13322 / 13492 = 98.74 % 98.3398 %
JJ 7617 / 8215 = 92.7206 % 85.2708 %
JJR 388 / 423 = 91.7258 % 95.9811 %
JJS 262 / 267 = 98.1273 % 95.5056 %
LS 10 / 15 = 66.6667 % 0 %
MD 1264 / 1267 = 99.7632 % 99.8421 %
NN 17257 / 17834 = 96.7646 % 91.9143 %
NNP 12717 / 13177 = 96.5091 % 85.0118 %
NNPS 98 / 170 = 57.6471 % 49.4118 %
* ===== OVERALL ACCURACY =====
*
125966 / 129654 = 97.1555 % 91.3015 %
*
C:\Storage\Library\projects\Naz_mag\new>
```

Рисунок 8 – Проміжні результати навчання нейромережі

Оцінимо вихідні дані:

```
# single output {
'governance': 0.00004324968926091062,
'risk': 0.007702528578033991,
'compliance': 0.0002575132225946431,
'risk management',
'data management': 0.2071775132225946431,
'big data': 0.008160047807935744,
'administration': 0.00015069427192724994 }
```

ВИСНОВКИ

В роботі вирішено задачу розроблення загальної архітектури ІС поширення комерційного контенту в Інтернет-просторі на основі навчання нейронної

мережі згідно історії постійної аудиторії для подачі унікального контенту з використанням підходу персоналізації та використання тегів. В роботі також сформульовані загальні вимоги до типової архітектури ІС поширення комерційного контенту на основі підходу персоналізації та використання тегів.

Наукова новизна полягає у розробленні моделі та методу поширення комерційного контенту на основі підходу персоналізації та використання тегів. При цьому використано навчання нейронної мережі для створення тег рекомендацій та доступні на ринку засоби персоналізації. А алгоритм персоналізації дозволяє пов'язати кожного користувача з списком продуктів які найімовірніше його зацікавлять, а також можуть прогнозувати те, що клієнти можуть хотіти бачити, навіть якщо вони ще не знають про це.

Як правило, кожен користувач має різні вимоги до інформації для запиту. Але типові пошукові системи повертають один і той самий результат для одного запиту, поданого різними користувачами. Для вирішення проблеми інформаційного перевантаження та надання користувачам відповідної інформації використовується Web-персоналізація. Персоналізація в Інтернет підвищує точність пошукової системи, спрощує процес пошуку, зберігає час та надає відповідну інформацію користувачам. Персоналізація створює почуття індивідуальності та унікальності. Клієнти почувають себе особливими та важливими, наче компанія звертає на них особливу увагу. Більш того, шляхом сегментації та націлювання на різних покупців, персоналізація відповідає різним потребам кожного клієнта, тим самим оптимізуючи клієнтський досвід, а також той же середній досвід для всіх.

Робота має практичну цінність у розробленні правил перетворення ключових слів інформаційного пошуку конкретного користувача на персоналізований список тегів для навчання нейронної мережі формування рекомендацій кінцевому користувачу згідно його вподобань.

Перспективи подальших досліджень полягають у аналізі статистики функціонування впроваджених аналогічних систем для вдосконалення методів персоналізації у формуванні рекомендацій кінцевому користувачу. Необхідно адаптувати розроблені алгоритми створення списків персоналізованих тегів для україномовного пошуку та відповідно формування рекомендацій кінцевому користувачу. Впровадження цієї системи дасть змогу отримувати доступ до необхідного контенту широкому загалу користувачів, адже сайт буде розміщено у всесвітній павутині, з другого боку інша частина мети створення цієї системи є комерційна складова, а саме отримання прибутків власником чи адміністратором ІС, через механізми е-комерції. Проект у закінченому вигляді має практичне своє застосування, а саме може бути використаний, як один з Інтернет-магазинів у всесвітній мережі, за умови заповнення його реальними продуктами та послугами. Якщо надати оцінку ступеня готовності до впровадження то

потрібно спочатку провести відповідні дослідження попиту та пропозиції представлених послуг на ринку, перш ніж запускати весь цикл робіт, пов'язаних з електронною комерцією. Успіх залежить не тільки від грамотності та якості побудови сайту, але й від зацікавленості в запропонованих послугах, реклами та відомості ресурсу.

ПОДЯКИ

Роботу виконано в рамках держбюджетної теми «Методи та засоби функціонування систем підтримки прийняття рішень на основі онтологій» (ID:839 2017-05-15 09:20:01 (2459–315)). Дослідження провадились в межах спільних наукових досліджень кафедри інформаційних систем та мереж НУ «Львівська політехніка» на тему «Дослідження, розроблення і впровадження інтелектуальних розподілених інформаційних технологій та систем на основі ресурсів баз даних, сховищ даних, просторів даних та знань з метою прискорення процесів формування сучасного інформаційного суспільства». Наукові дослідження провадилися також в рамках ініціативної тематики досліджень кафедри ІСМ НУ «Львівська політехніка» на тему «Розроблення інтелектуальних розподілених систем на основі онтологічного підходу з метою інтеграції інформаційних ресурсів».

СПИСОК ЛІТЕРАТУРИ

1. Mobasher B. Data mining for web personalization / B. Mobasher // *The adaptive web*. – 2007. – Vol. 4321. – P. 90–135.
2. Dinucă C. Web Content Mining. In: University of Petroșani / C. Dinucă, D. Ciobanu // *Economics*. – 2012. – Vol. 12. – P. 85–92.
3. Xu G. Web content mining / G. Xu, Y. Zhang, L. Li // *Web Mining and Social Networking*. – 2011. – Vol. 6. – P. 71–87.
4. Khribi M. K. Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval / M. K. Khribi, M. Jemni, O. Nasraoui // *Advanced Learning Technologies : International Conference, 1–5 July 2008 : proceedings*. – Santander, Cantabria, Spain : IEEE, 2008. – P. 241–245.
5. Automatic web content personalization through reinforcement learning / [S. Ferretti, S. Mirri, C. Prandi, P. Salomoni] // *Journal of Systems and Software*. – 2016. – Vol. 121. – P. 157–169.
6. User attitudes towards news content personalization / [T. Lavie, M. Sela, I. Oppenheim et al] // *International journal of human-computer studies*. – 2010. – Vol. 68(8). – P. 483–495.
7. Fredrikson M. Repriv: Re-imagining content personalization and in-browser privacy / M. Fredrikson, B. Livshits // *Symposium on Security and Privacy: Conference, 22–25 May 2011 : proceedings*. – Berkeley, CA, USA : IEEE, 2011. – P. 131–146.
8. Application of neural networks and Kano's method to content recommendation in web personalization / [C. C. Chang, P. L. Chen, F. R. Chiu, Y. K. Chen] // *Expert Systems with Applications*. – 2009. – Vol. 36(3). – P. 5310–5316.
9. Pat. US7,571,226B1 US Content personalization over an interface with adaptive voice character / H. Partovi, R. Brathwaite, A. Davis, M. McCue, B. Porter, J. Giannandrea, Z. Li (US) ; TellMe Networks, Inc., Mountain View, CA (US). – No.: 09/523,853 ; Marz 14, 2009; August 4, 2009, Patent and Trademark Office. – 20 p.
10. Pat. US2009/0171968A1 US Widget-assisted content personalization based on user behaviors tracked across multiple web sites / F. J. Kane, C. Hicks (US) ; Amazon Technologies Inc (US). – No.: 11/966,817 ; December 28, 2007; July 2, 2009, Google Patents. – 24 p.
11. Mirri S. Experiential adaptation to provide user-centered web content personalization / S. Mirri, C. Prandi, P. Salomoni // *Advances in Human oriented and Personalized Mechanisms, Technologies, and Services : The Sixth International Conference, October 27 – November 1, 2013: proceedings*. – Venice, Italy : IARIA, 2013. – P. 31–36.
12. Fernandez-Luque L. Review of extracting information from the Social Web for health personalization / L. Fernandez-Luque, R. Karlsen, J. Bonander // *Journal of medical Internet research*. – 2011. – Vol. 13(1). – P. 15.
13. Pat. US8,019,777B2 US Digital content personalization method and system / E. Hauser (US) ; CRICKET MEDIA Inc (US). –No.: 12/795,419 ; June 7, 2010; September 13, 2011, Patent and Trademark Office. – 15 p.
14. Ho S. Y. Timing of adaptive web personalization and its effects on online consumer behavior / S. Y. Ho, D. Bodoff, K. Y. Tam // *Information Systems Research*. – 2011. – Vol. 22(3). – P. 660–679.
15. Uchyigit G. Personalization techniques and recommender systems / G. Uchyigit, M. Y. Ma. – Singapore: World Scientific, – 2008. – 322 p.
16. Pat. US2006/0020883A1 Web page personalization / N. Kothari, M. Harder, R. Howard, A. Sanabria, S. Schackow (US) ; Microsoft Technology Licensing LLC (US). – No.: 10/857,724 ; May 28, 2004; Januar 26, 2006, Patent and Trademark Office. – 18 p.
17. Zhang H. Construction of ontology-based user model for web personalization / H. Zhang, Y. Song., H. T. Song // *Lecture Notes in Computer Science*. – 2007. – Vol. 4511. – P. 67–76.
18. Pat. US 8,254,892 B2 US Methods and apparatus for anonymous user identification and content personalization in wireless communication / H. Chien (US) ; AT&T Mobility II LLC (US). – No.: 12/468,708 ; September 10, 2009; August 28, 2012, Patent and Trademark Office. – 9 p.
19. Pat. US7,970,664B2 US Content personalization based on actions performed during browsing sessions / G. D. Linden, B. R. Smith, N. K. Zada (US) ; Amazon Technologies Inc (US). – No.: 11/009,732 ; December 10, 2004; June 28, 2011, Patent and Trademark Office. – 36 p.
20. Web personalization using web mining: concept and research issue / [P. Mehtaa, B. Parekh, K. Modi, P. Solanki] // *International Journal of Information and Education Technology*. – 2012. – Vol. 2(5). – P. 510–512.
21. Zhezhnych P. Linguistic Comparison Quality Evaluation of Web-Site Content with Tourism Documentation Objects / P. Zhezhnych, O. Markiv // *Advances in Intelligent Systems and Computing*. – 2018. – Vol. 689. – P. 656–667.
22. Basyuk T. The main reasons of attendance falling of internet resource / T. Basyuk // *Computer Sciences and Information Technologies : Xth International Scientific and Technical Conference, 14–17 September 2015 : proceedings*. – Lviv : IEEE, 2015. – P. 91–93.
23. Uniform Method of Operative Content Management in Web Systems / [A. Gozhyj, L. Chyrun, A. Kowalska-Styczen,

- O. Lozynska] // CEUR Workshop Proceedings. – 2018. – Vol. 2136. – P. 62–77.
24. Kravets P. The control agent with fuzzy logic / P. Kravets // *Perspective Technologies and Methods in MEMS Design : VIth International Conference, 20–23 April 2010* 2015 : proceedings. – Lviv : IEEE, 2015. – P. 40–41.
25. Davydov M. Linguistic Models of Assistive Computer Technologies for Cognition and Communication / M. Davydov, O. Lozynska // *Computer Science and Information Technologies : XIth International Scientific and Technical Conference, 6–10 September 2016* : proceedings. – Lviv : IEEE, 2016. – P. 171–175.
26. Design and implementation of visitors queue density analysis and registration method for retail videosurveillance purposes / [Peleshko D., Ivanov Y., Sharov B., Izonin I., Borzov Y.] // *Data Stream Mining & Processing : First International Conference, 23–27 August 2016* : proceedings. – Lviv : IEEE, 2016. – P. 159–162.
27. Adaptive moving object segmentation algorithms in cluttered environments / [Y. Ivanov, D. Peleshko, O. Makoveychuk et al] // *The Experience of Designing and Application of CAD Systems in Microelectronics : Conference, 24–27 February 2015* : proceedings. – Lviv : IEEE, 2015. – P. 97–99.
28. Vitynskyi P. Hybridization of the SGTm Neural-like Structure through Inputs Polynomial Extension / [P. Vitynskyi, R. Tkachenko, I. Izonin, H. Kutucu] // *Data Stream Mining & Processing : Second International Conference, 21–25 August 2018* : proceedings. – Lviv : IEEE, 2018. – P. 386–391.
29. Development of the Non-Iterative Supervised Learning Predictor Based on the Ito Decomposition and SGTm Neural-Like Structure for Managing Medical Insurance Costs / [R. Tkachenko, I. Izonin, P. Vitynskyi et al] // *Data*. – 2018. – Vol. 3(4). – P. 1–14.
30. Mykich K. Algebraic model for knowledge representation in situational awareness systems / K. Mykich, Y. Burov // *Computer Sciences and Information Technologies : International Scientific and Technical Conference, 6–10 September 2016* : proceedings. – Lviv : IEEE, 2016. – P. 165–167.
31. Mykich K. Uncertainty in situational awareness systems / K. Mykich, Y. Burov // *Modern Problems of Radio Engineering, Telecommunications and Computer Science : 13th International Conference, 623–26 Februar 2016* : proceedings. – Lviv : IEEE, 2016. – P. 729–732.
32. Mykich K. Algebraic Framework for Knowledge Processing in Systems with Situational Awareness / K. Mykich, Y. Burov // *Advances in Intelligent Systems and Computing*. – 2017. – Vol. 512. – P. 217–227.
33. Mykich K. Research of uncertainties in situational awareness systems and methods of their processing / K. Mykich, Y. Burov // *EasternEuropean Journal of Enterprise Technologies*. – 2016. – Vol. 1(79). – P. 19–26.
34. The Risk Management Modelling in Multi Project Environment / [V. Lytvyn, V. Vysotska, O. Veres et al] // *Computer Science and Information Technologies : 12th International Scientific and Technical Conference, 5–8 September 2017* : proceedings. – Lviv : IEEE, 2017. – P. 32–35.
35. Vysotska V. Linguistic Analysis of Textual Commercial Content for Information Resources Processing / V. Vysotska // *Modern Problems of Radio Engineering, Telecommunications and Computer Science : 13th International Scientific and Technical Conference, 23–26 February 2016* : proceedings. – Lviv : IEEE, 2016. – P. 709–713.
36. Information resources processing using linguistic analysis of textual content / [J. Su, V. Vysotska, A. Sachenko et al] // *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications : 9th International Conference, 21–23 September 2017* : proceedings. – Bucharest, Romania: IEEE, 2017. – P. 573–578.
37. Content Linguistic Analysis Methods for Textual Documents Classification / [V. Lytvyn, V. Vysotska, O. Veres et al] // *Computer Science and Information Technologies : 11th International Scientific and Technical Conference, 6–10 September 2016* : proceedings. – Lviv : IEEE, 2016. – P. 190–192.
38. Bisikalo O.V. Identifying keywords on the basis of content monitoring method in ukrainian texts / O. V. Bisikalo, V. A. Vysotska // *Radio Electronics, Computer Science, Control*. – 2016. – Vol. 1(36). – P. 74–83.
39. Bisikalo O. V. Sentence syntactic analysis application to keywords identification Ukrainian texts / O. V. Bisikalo, V. A. Vysotska // *Radio Electronics, Computer Science, Control*. – Vol. 3(38). – 2016. – P. 54–65.
40. Lytvyn V. Application of algorithmic algebra system for grammatical analysis of symbolic computation expressions of propositional logic / V. Lytvyn, I. Bobyk, V. Vysotska // *Radio Electronics, Computer Science, Control*. – 2016. – Vol. 4(39). – P. 54–67.
41. Aliksieieva K. Technology of commercial web-resource management based on fuzzy logic / K. Aliksieieva, A. Berko, V. Vysotska // *Radio Electronics, Computer Science, Control*. – Vol. 3(34). – 2015. – P. 71–79.
42. Peculiarities of Content Forming and Analysis in Internet Newspaper Covering Music News / [M. Korobchinsky, V. Vysotska, L. Chyrun, L. Chyrun] // *Computer Science and Information Technologies : 12th International Scientific and Technical Conference, 5–8 September 2017* : proceedings. – Lviv : IEEE, 2017. – P. 52–57.
43. Intellectual System Design for Content Formation / [O. Naum, L. Chyrun, O. Kanishcheva, V. Vysotska] // *Computer Science and Information Technologies : 12th International Scientific and Technical Conference, 5–8 September 2017* : proceedings. – Lviv : IEEE, 2017. – P. 131–138.
44. Application of Sentence Parsing for Determining Keywords In Ukrainian Texts / [Vasyl Lytvyn, Victoria Vysotska, Dmytro Dosyn, Roman Holoschuk, Zoriana Rybchak] // *Computer Science and Information Technologies : 12th International Scientific and Technical Conference, 5–8 September 2017* : proceedings. – Lviv : IEEE, 2017. – P. 326–331.
45. Vysotska V. Process analysis in electronic content commerce system / V. Vysotska, R. Hasko, V. Kuchkovskiy // *Computer Science and Information Technologies : Xth International Scientific and Technical Conference, 14–17 September 2015* : proceedings. – Lviv : IEEE, 2015. – P. 120–123.
46. Lytvyn, V. Designing architecture of electronic content commerce system / V. Lytvyn, V. Vysotska // *Computer Science and Information Technologies : Xth International Scientific and Technical Conference, 14–17 September 2015* : proceedings. – Lviv : IEEE, 2015. – P. 115–119.

Received 17.06.2019.
Accepted 18.09.2019.

УДК 004.9

ОСОБЕННОСТИ АРХИТЕКТУРЫ ИНТЕРНЕТ СИСТЕМЫ УПРАВЛЕНИЯ КОММЕРЧЕСКИМ КОНТЕНТОМ НА ОСНОВЕ МЕТОДОВ MACHINE LEARNING, WEB MINING И SEO-ТЕХНОЛОГИЙ

Высоцкая В. А. – канд. техн. наук, доцент, доцент кафедры «Информационные системы и сети», Национальный университет «Львовская политехника», Украина.

Демчук А. Б. – канд. техн. наук, ассистент кафедры «Информационные системы и сети», Национальный университет «Львовская политехника», Украина.

Лытвын В. В. – д-р техн. наук, профессор, заведующий кафедрой «Информационные системы и сети», Национальный университет «Львовская политехника», Украина.

АННОТАЦИЯ

Актуальность. Сегодня большинство корпораций постоянно переосмысливает бизнес с точки зрения возможностей Интернет, а именно его доступность, широкий охват и постоянно меняющиеся потребности пользователя. Web-ресурс электронной коммерции, который обеспечивает удобный пользовательский опыт, в частности, возможность быстро находить необходимые согласно его портб и вкуса товары, больше поддерживает конкурентные преимущества.

Целью исследования является разработка общей архитектуры интеллектуальной системы распространения коммерческого контента в Интернет-пространстве на основе обучения нейронной сети согласно истории поситийной айдитории для подачи уникального контента с использованием подхода персонализации и использование тегов.

Метод. Разработана модель информационной системы персонализации коммерческого контента согласно потребностям пользователя. Также разработан метод распространения коммерческого контента на основе подхода персонализации и использования тегов. При этом использовано обучения нейронной сети для создания тегов рекомендаций и доступных на рынке средств персонализации. Разработанный алгоритм персонализации позволяет связать каждого пользователя со списком продуктов, которые вероятнее всего его заинтересуют, а также может прогнозировать то, что клиенты могут хотеть видеть, даже если они еще не знают об этом. Разработанный метод можно использовать для обеспечения более релевантного набора контента. Также разработан метод дает возможность классифицировать соответствующий контент или показать его раньше в процессе перелистывания страниц во избежания потребителями выбора неправильного контента или затраты времени на прокрутки при поиске товара.

Результаты. Разработанная система предназначена для распространения продуктов информационных технологий (публикаций, книг, курсов, видео, файлов и т.д.) с помощью интернета.

Выводы. Внедрение этой системы позволит получать доступ к определенному рода контента широкой общественности пользователей, ведь сайт будет размещен во всемирной паутине, с другой стороны другая часть цели создания этой системы является коммерческая составляющая, а именно получение прибыли владельцем или администратором интеллектуальной системы, через механизмы e-коммерции.

КЛЮЧЕВЫЕ СЛОВА: коммерческий контент, персонализация, Web mining, Machine Learning, SEO-технология, метрики поиска, электронная коммерция, NLP, контент-мониторинг, контент-анализ, статистический лингвистический анализ, квантитативных лингвистика.

UDC 004.9

FEATURES OF THE ARCHITECTURE FOR INTERNET COMMERCIAL CONTENT MANAGEMENT SYSTEM BASED ON METHODS OF MACHINE LEARNING, WEB MINING AND SEO TECHNOLOGIES

Vysotska V. – PhD, Associate Professor of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

Demchuk A. – PhD, Assistant of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

Lytvyn V. – Dr. Sc., Professor, Head of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

ABSTRACT

Context. Today, most corporations are constantly rethinking business from the point of view of the Internet, namely its availability, broad reach and ever-changing needs of the user. The e-commerce web-site, which provides user-friendly experience, including the ability to quickly find the products that are necessary for its portables and taste, is more in favor of competitive advantage.

Objective of the study is to develop a general architecture of the intellectual system for the distribution of commercial content in the Internet space, based on the study of the neural network in accordance with the history of the psychedelic region to provide unique content using the approach of personalization and the use of tags.

Method. The model of information system of commercial content personalization for the user needs is developed. Also the method of distributing commercial content based on the approach of personalization and the tags use is developed. In this case, the neural network training is used to create a recommendation tag and marketable personalization tools. The customization algorithm allows you to associate each user with a list of products that they are most likely to be interested in, and can predict what customers might want to see even if they do not yet know about it. The developed method can be used to provide a more relevant set of content. Also, the developed method gives the opportunity to classify the relevant content or show it earlier in the process of rolling the pages to avoid consumers choosing the wrong content or spending time scrolling when looking for a product.

Results. The developed system is intended for distribution of information technology products (publications, books, courses, videos, files, etc.) through the Internet.

Conclusions. Implementation of this system will allow access to certain types of content to the general public, since the site will be placed on the World Wide Web; on the other hand, another part of the purpose of creating this system is a commercial component, namely, the receipt of profits by the owner or administrator of the intellectual system, through the mechanisms of e-commerce.

KEYWORDS: commercial content, personalization, Web mining, Machine Learning, SEO technology, metrics search, e-commerce, NLP, content monitoring, content analysis, statistical linguistic analysis, quantitative linguistics.

REFERENCES

1. Mobasher B. Data mining for web personalization, *The adaptive web*, 2007, Vol. 4321, pp. 90–135.
2. Dinucă C., Ciobanu D. Web Content Mining. In: University of Petroșani, *Economics*, 2012, Vol. 12, pp. 85–92.
3. Xu G., Zhang Y., Li L. Web content mining, *Web Mining and Social Networking*, 2011, Vol. 6, pp. 71–87.
4. Khribi M. K., Jemni M., Nasraoui O. Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval, *Advanced Learning Technologies : International Conference, 1–5 July 2008 : proceedings. Santander, Cantabria, Spain, IEEE*, 2008, pp. 241–245.
5. Ferretti S., Mirri S., Prandi C., Salomoni P. Automatic web content personalization through reinforcement learning, *Journal of Systems and Software*, 2016, Vol. 121, pp. 157–169.
6. Lavie T., Sela M., Oppenheim I., Inbar O., Meyer J. User attitudes towards news content personalization, *International journal of human-computer studies*, 2010, Vol. 68(8), pp. 483–495.
7. Fredrikson M., Livshits B. Repriv: Re-imagining content personalization and in-browser privacy, *Symposium on Security and Privacy: Conference, 22–25 May 2011 : proceedings*. Berkeley, CA, USA, IEEE, 2011, pp. 131–146.
8. Chang C., Chen P., Chiu F., Chen Y. Application of neural networks and Kano's method to content recommendation in web personalization, *Expert Systems with Applications*, 2009, Vol. 36(3), pp. 5310–5316.
9. Partovi H., Brathwaite R., Davis A., McCue M., Porter B., Giannandrea J., Li Z. (US) Pat. US7,571,226B1 US Content personalization over an interface with adaptive voice character, U.S. ; TellMe Networks, Inc., Mountain View, CA (US), No.: 09/523,853 ; Marz 14, 2009; August 4, 2009, Patent and Trademark Office, 20 p.
10. Kane F. J., Hicks C. (US) Pat. US2009/0171968A1 US Widget-assisted content personalization based on user behaviors tracked across multiple web sites, Amazon Technologies Inc (US), No.: 11/966,817 ; December 28, 2007; July 2, 2009, Google Patents, 24 p.
11. Mirri S., Prandi C., Salomoni P. Experiential adaptation to provide user-centered web content personalization, *Advances in Human oriented and Personalized Mechanisms, Technologies, and Services : The Sixth International Conference, October 27 – November 1, 2013: proceedings*. Venice, Italy, IARIA, 2003, pp. 31–36.
12. Fernandez-Luque L., Karlsen R., Bonander J. Review of extracting information from the Social Web for health personalization, *Journal of medical Internet research*, 2011, Vol. 13(1), P. 15.
13. Hauser E. (US) Pat. US8,019,777B2 US Digital content personalization method and system /; CRICKET MEDIA Inc (US). No.: 12/795,419 ; June 7, 2010; September 13, 2011, Patent and Trademark Office, 15 p.
14. Ho S. Y., Bodoff D., Tam K. Y. Timing of adaptive web personalization and its effects on online consumer behavior *Information Systems Research*, 2011, Vol. 22(3), pp. 660–679.
15. Uchyigit G., Ma M. Y. Personalization techniques and recommender systems. Singapore, World Scientific, 2008, 322 p.
16. Kothari N., Harder M., Howard R., Sanabria A., Schackow S. Pat. US2006/0020883A1 Web page personalization / (US) ; Microsoft Technology Licensing LLC (US). No.: 10/857,724 ; May 28, 2004; Januar 26, 2006, Patent and Trademark Office. – 18 p.
17. Zhang H., Song Y., Song H. T. Construction of ontology-based user model for web personalization, *Lecture Notes in Computer Science*, 2007, Vol. 4511, pp. 67–76.
18. Chien H. (US) Pat. US 8,254,892 B2 US Methods and apparatus for anonymous user identification and content personalization in wireless communication; AT&T Mobility II LLC (US). No.: 12/468,708 ; September 10, 2009; August 28, 2012, Patent and Trademark Office, 9 p.
19. Linden G. D., Smith B. R., Zada N. K. (US) Pat. US7,970,664B2 US Content personalization based on actions performed during browsing sessions; Amazon Technologies Inc (US). No.: 11/009,732 ; December 10, 2004; June 28, 2011, Patent and Trademark Office. –36 p.
20. Mehtaa P., Parekh B., Modi K., Solanki P. Web personalization using web mining: concept and research issue, *International Journal of Information and Education Technology*, 2012, Vol. 2(5), pp. 510–512.
21. Zhezhnych P., Markiv O. Linguistic Comparison Quality Evaluation of Web-Site Content with Tourism Documentation Objects, *Advances in Intelligent Systems and Computing*, 2018, Vol. 689, pp. 656–667.
22. Basyuk T. The main reasons of attendance falling of internet resource, *Computer Sciences and Information Technologies : Xth International Scientific and Technical Conference, 14–17 September 2015 : proceedings*. Lviv, IEEE, 2015, pp. 91–93.
23. Gozhyj A., Chyrun L., Kowalska-Styczen A., Lozynska O. Uniform Method of Operative Content Management in Web Systems, *CEUR Workshop Proceedings*, 2018, Vol. 2136, pp. 62–77.
24. Kravets P. The control agent with fuzzy logic, *Perspective Technologies and Methods in MEMS Design : Vth International Conference, 20–23 April 2010 2015 : proceedings*. Lviv, IEEE, 2015, pp. 40–41.
25. Davydov M., Lozynska O. Linguistic Models of Assistive Computer Technologies for Cognition and Communication, *Computer Science and Information Technologies : XIth International Scientific and Technical Conference, 6–10 September 2016 : proceedings*. Lviv, IEEE, 2016, pp. 171–175.
26. Peleshko D., Ivanov Y., Sharov B., Izonin I., Borzov Y. Design and implementation of visitors queue density analysis and registration method for retail videosurveillance purposes, *Data Stream Mining & Processing : First International Conference, 23–27 August 2016 : proceedings*. Lviv, IEEE, 2016, pp. 159–162.
27. Ivanov Y., Peleshko D., Makoveychuk O., Izonin I., Malets I., Lotoshunska N., Batyuk D. Adaptive moving object segmentation algorithms in cluttered environments, *The Experience of Designing and Application of CAD Systems in*

- Microelectronics : Conference, 24 February 2015 : proceedings.* Lviv, IEEE, 2015, pp. 97–99.
28. Vitynskyi P., Tkachenko R., Izonin I., Kutucu H. Hybridization of the SGTM Neural-like Structure through Inputs Polynomial Extension, *Data Stream Mining & Processing : Second International Conference, 21–25 August 2018 : proceedings.* Lviv, IEEE, 2018, pp. 386–391.
29. Tkachenko R., Izonin I., Vitynskyi P., Lotoshynska N., and Pavlyuk O. Development of the Non-Iterative Supervised Learning Predictor Based on the Ito Decomposition and SGTM Neural-Like Structure for Managing Medical Insurance Costs, *Data*, 2018, Vol. 3(4), pp. 1–14.
30. Mykych K., Burov Y. Algebraic model for knowledge representation in situational awareness systems, *Computer Sciences and Information Technologies : 11th International Scientific and Technical Conference, 6–10 September 2016 : proceedings.* Lviv, IEEE, 2016, pp. 165–167.
31. Mykych K., Burov Y. Uncertainty in situational awareness systems, *Modern Problems of Radio Engineering, Telecommunications and Computer Science : 13th International Conference, 623–26 Februar 2016 : proceedings.* Lviv, IEEE, 2016, pp. 729–732.
32. Mykych K. Algebraic Framework for Knowledge Processing in Systems with Situational Awareness, *Advances in Intelligent Systems and Computing*, 2017, Vol. 512, pp. 217–227.
33. Mykych K., Burov Y. Research of uncertainties in situational awareness systems and methods of their processing, *Eastern European Journal of Enterprise Technologies*, 2016, Vol. 1(79), pp. 19–26.
34. Lytvyn V., Vysotska V., Veres O., Rishnyak I., Rishnyak H. The Risk Management Modelling in Multi Project Environment, *Computer Science and Information Technologies : 12th International Scientific and Technical Conference, 5–8 September 2017 : proceedings.* Lviv, IEEE, 2017, pp. 32–35.
35. Vysotska V. Linguistic Analysis of Textual Commercial Content for Information Resources Processing, *Modern Problems of Radio Engineering, Telecommunications and Computer Science : International Scientific and Technical Conference, 23–26 February 2016 : proceedings.* Lviv, IEEE, 2016, pp. 709–713.
36. Su J., Vysotska V., Sachenko A., Lytvyn V., Burov Y. Information resources processing using linguistic analysis of textual content, *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications : 9th International Conference, 21–23 September 2017 : proceedings.* Bucharest, IEEE, 2017, pp. 573–578.
37. Lytvyn V., Vysotska V., Veres O., Rishnyak I., Rishnyak H. Content Linguistic Analysis Methods for Textual Documents Classification, *Computer Science and Information Technologies : 11th International Scientific and Technical Conference, 6–10 September 2016 : proceedings.* Lviv, IEEE, 2016, pp. 190–192.
38. Bisikalo O. V., Vysotska V. A. Identifying keywords on the basis of content monitoring method in ukrainian texts, *Radio Electronics, Computer Science, Control*, 2016, Vol. 1(36), pp. 74–83.
39. Bisikalo O. V., Vysotska V. A. Sentence syntactic analysis application to keywords identification Ukrainian texts, *Radio Electronics, Computer Science, Control*, 2016, Vol. 3(38), pp. 54–65.
40. Lytvyn V., Bobyk I., Vysotska V. Application of algorithmic algebra system for grammatical analysis of symbolic computation expressions of propositional logic, *Radio Electronics, Computer Science, Control*, 2016, Vol. 4(39), pp. 54–67.
41. Aliksieieva K., Berko A., Vysotska V. Technology of commercial web-resource management based on fuzzy logic, *Radio Electronics, Computer Science, Control*, 2015, Vol. 3(34), pp. 71–79.
42. Korobchinsky M., Vysotska V., Chyrun L., Chyrun L. Peculiarities of Content Forming and Analysis in Internet Newspaper Covering Music News, *Computer Science and Information Technologies : 12th International Scientific and Technical Conference, 5–8 September 2017 : proceedings.* Lviv, IEEE, 2017, pp. 52–57.
43. Naum O., Chyrun L., Kanishcheva O., Vysotska V. Intellectual System Design for Content Formation, *Computer Science and Information Technologies : 12th International Scientific and Technical Conference, 5–8 September 2017 : proceedings.* Lviv, IEEE, 2017, pp. 131–138.
44. Lytvyn Vasyl, Vysotska Victoria, Dosyn Dmytro, Holoschuk Roman, Rybchak Zoriana Application of Sentence Parsing for Determining Keywords In Ukrainian Texts, *Computer Science and Information Technologies : 12th International Scientific and Technical Conference, 5–8 September 2017 : proceedings.* Lviv, IEEE, 2017, pp. 326–331.
45. Vysotska V., Hasko R., Kuchkovskiy V. Process analysis in electronic content commerce system, *Computer Science and Information Technologies : Xth International Scientific and Technical Conference, 14–17 September 2015 : proceedings.* Lviv, IEEE, 2015, pp. 120–123.
46. Lytvyn V., Vysotska V. V. Designing architecture of electronic content commerce system, *Computer Science and Information Technologies : Xth International Scientific and Technical Conference, 14–17 September 2015 : proceedings.* Lviv, IEEE, 2015, pp. 115–119.