

ОСОБЛИВОСТІ РУБРИКАЦІЇ ТЕКСТОВОГО КОМЕРЦІЙНОГО КОНТЕНТУ

© Висоцька В., 2015

Розглянуто застосування породжувальних граматик у лінгвістичному моделюванні. Опис моделювання синтаксису речення застосовують для автоматизації процесів аналізу та синтезу природномовних текстів.

Ключові слова: породжувальні граматика, структурна схема речення, інформаційна лінгвістична система, інформаційні ресурси, комерційний контент, контент-аналіз, контент-моніторинг, контентний пошук, система електронної контент-комерції.

This paper presents the generative grammar application in linguistic modelling. Description of syntax sentence modelling is applied to automate the processes of analysis and synthesis of texts in natural language.

Key words: generative grammar, structured scheme sentences, information linguistic system, generative grammar, structured scheme sentences, information linguistic system, information resources, commercial content, content analysis, content monitoring, content search, electronic content of commerce system.

Вступ. Загальна постановка проблеми

Необхідність реалізації процесів аналізу та синтезу природномовних текстів зумовила появу відповідних лінгвістичних моделей процесів їх опрацювання. Виникла потреба в розвитку багатьох мовознавчих дисциплін для потреб інформаційних наук, у розробленні автоматизованих систем опрацювання багатомовної інформації [1–9].

Лінгвістичний аналіз природномовних текстів складається з декількох послідовних процесів – графемного, морфологічного, синтаксичного та семантичного аналізу. Для моделювання на синтаксичному рівні мови застосовують формалізм породжувальних граматик, що запропонував Н. Хомські. Формальний аналіз граматичної структури фраз дає змогу виділити синтаксичні структури (складові), що є основною схемою фрази, незалежно від її значення. Напрацювання Н. Хомські та А. В. Гладкого застосовні до розроблення таких засобів опрацювання природної мови, як інформаційно-пошукові системи, системи машинного перекладу, анотування текстів, морфологічний, синтаксичний та семантичний аналіз текстів, навчально-дидактичні системи, до лінгвістичного забезпечення спеціалізованих програмних систем тощо.

У статті наведено способи застосування апарату породжувальних граматик до моделювання синтаксису речень для різних мов – англійської, німецької та української. Для цього виконано розбір синтаксичної структури речень, продемонстровано особливості процесу синтезу речень зазначених мов. Розглянуто вплив норм та правил мови на хід побудови граматик. Додатково було наведено приклад застосування граматики для ілюстрації породження речення зі схемою елементарної іменної групи відповідного типу. Також подано список прикметників, забезпечених індексами класу.

Аналіз останніх досліджень та публікацій

Аналіз процесу опрацювання Web-ресурсу дає змогу зменшити витрати на формування, публікацію, зберігання, управління, супровід і донесення готового актуального та релевантного комерційного контенту постійному користувачеві. Публікація комерційного контенту на Web-ресурсах надає додаткові можливості всім авторам релевантного контенту, який актуальний для

вужького кола цільової аудиторії, але який не публікують через збитковість традиційні видавництва. Автори контенту мають змогу публікуватися самостійно, формуючи тексти в редакторах і розміщуючи їх на Web-ресурсах для поширення з урахуванням авторського права. Відомим методом аналізу змісту текстової інформації (коментарів, форумів, е-листування, статей тощо) є *контент-аналіз*. Поняття контент-аналізу не має однозначного визначення [1], тому ІС, побудовані на основі різних підходів, несумісні. Застосування контент-аналізу тексту в СЕKK дає переваги для спрощення ведення е-бізнесу та вирішує низку проблем учасників бізнес-процесів. Зростання темпів виробництва контенту призводить до зниження загального рівня інформованості потенційного користувача, у контенті з'являється інформаційний шум, підвищується нерелевантність контенту, відбувається його дублювання, ускладнюється процес пошуку в разі відбору контенту з різних джерел інформації. Для узагальнення великих динамічних потоків контенту запропоновано метод контент-моніторингу. Вхідною інформацією для контент-моніторингу є текст природною мовою як послідовність символів, вихідна інформація – це таблиці розділів, речень і лексем аналізованого тексту. Контент-моніторинг є програмним засобом автоматизації знаходження найважливіших складових у потоках контенту. Складовою контент-моніторингу є контент-аналіз тексту, призначений для пошуку контенту в масиві даних за змістовими лінгвістичними одиницями. Застосування контент-аналізу для моніторингу інтернет-джерел даних дає змогу автоматизувати процес виявлення найважливіших складових у потоках контенту під час відбору даних з цих джерел. Це усуває дублювання контенту, інформаційний шум, паразитичний контент, надмірність результатів пошуку тощо.

Аналіз отриманих наукових результатів

Аналіз лексико-граматичної та семантико-прагматичної побудови тексту використовують для автоматичної рубрикації контенту та формування дайджестів, що приводить до формування тематично підібраних масивів контенту (таблиця) [1–9].

Основні етапи рубрикації комерційного контенту

Назва	Призначення етапу
Підготовка	Визначення тематики/мети/об'єкта аналізу, хронологічні та географічні межі, принципи відбору.
Збір даних	Формування класифікатора відбору ключових цитат та інструкції для кодувальника.
Формальний аналіз	Перетворення фрагментів тексту без аналізу його змісту. Морфологічні дані забезпечують доступ до змісту, опосередкованого через співвідношення одиниць змісту з одиницями виразу.
Змістовий аналіз	Аналіз елементів і логіко-семантичних відношень між ними для подання семантики контенту.
Синтаксичний аналіз	Автоматично за наявності лексико-граматичних та граматичних даних до кожного слова синтаксично прив'язують словоформи у реченні.
Морфемний аналіз	Сегментування тексту, де виділення префіксів можливе без знання частин мови, а суфіксів – ні: потрібні різні їх набори та процедури відсікання суфіксів для кожної частини мови окремо.
Класифікація	Автоматичне опрацювання фрагментів текстового контенту для розпізнавання змісту.
Кодування	Кодування фрагментів текстового контенту.
Архівація	Збереження фрагментів текстового контенту в базі даних.

Формалізм породжувальних граматик у ході процесу контент-моніторингу реалізує роботу з текстом як набором лінійно впорядкованих слів, словосполучень і речень. У разі використання варіації статистичних методів аналізу ігнорують лінгвістичну взаємопов'язаність і нелінійність природної мови. Цей етап дає змогу привести до єдиного вигляду досліджуваний контент під час заповнення шаблону для полегшення подальшої роботи з ним.

Оператор рубрикації α_5 комерційного контенту є відображенням контенту C_3 в новий стан C_4 через його валідацію, який відрізняється від попереднього стану його приналежністю до множини тематичного контенту $\alpha_5 : (C_3, U_{CT}, T) \rightarrow C_4$. За змістовий аналіз контенту відповідає процес видобування граматичних даних зі слова за допомогою графемного аналізу та корегування результатів морфологічного аналізу через аналіз граматичного контексту лінгвістичних одиниць (алг. 1).

Алгоритм 1. Рубрикація текстового комерційного контенту

Етап 1. Поділ комерційного контенту C_3 на блоки.

Крок 1. Подання на вхід блока побудови дерева блока комерційного контенту C_3 .

Крок 2. Створення нового блока в таблиці блоків.

Крок 3. Накопичення символів до символу нового рядка.

Крок 4. Перевірка на наявність крапки перед символом нового рядка. Якщо є, то перехід до кроку 5, якщо ні, то збереження послідовності у таблиці, розбір нового блока контенту C_3 та перехід до кроку 3.

Крок 5. Перевірка наявності кінця тексту для контенту C_3 . Якщо кінець тексту, то перехід до кроку 6, якщо ні, то зберігається накопичена послідовність у таблицю, розбір нового блока контенту C_3 та перехід до кроку 2.

Крок 6. Отримання на виході дерева блоків контенту C_3 у вигляді таблиці $U_{CT}^B \in U_{CT}$.

Етап 2. Поділ блока на речення зі збереженням структури контенту C_3 .

Крок 1. На вхід подається таблиця блоків $U_{CT}^B \in U_{CT}$. Створення таблиці речень $U_{CT}^R \in U_{CT}$ зі зв'язком за полем Код_розділу типу n -to-1 із таблицею блоків контенту C_3 .

Крок 2. Створення нового речення у таблиці речень $U_{CT}^R \in U_{CT}$.

Крок 3. Накопичення символів до крапки, крапки з комою або символу нового рядка.

Крок 4. Перевірка на наявність скорочення. Якщо скорочення, то перехід до кроку 5, якщо ні, то збереження послідовності у таблиці $U_{CT}^R \in U_{CT}$, розбір нового речення та перехід до кроку 2.

Крок 5. Перевірка наявності кінця тексту блока для контенту C_3 . Якщо кінець тексту, то перехід до кроку 6, якщо ні, то збереження послідовності у таблиці $U_{CT}^R \in U_{CT}$, розбір нового речення та перехід до кроку 2.

Крок 6. Отримують на виході дерево речень у вигляді таблиці $U_{CT}^R \in U_{CT}$.

Крок 7. Перевірка наявності кінця тексту для контенту C_3 . Якщо кінець тексту, то перехід до кроку 8, якщо ні, то розбір нового блока та перехід до кроку 1.

Крок 8. Отримання на виході дерева речень у вигляді таблиць $U_{CT}^R \in U_{CT}$.

Етап 3. Поділ речень на лексеми із вказанням належності до речень $U_{CT}^L \in U_{CT}$.

Крок 1. Формування на основі таблиці речень таблиці лексем $U_{CT}^L \in U_{CT}$ із полями Код_лексеми (унікальний ідентифікатор), Код_речення (число, що дорівнює коду речення з лексемою), Номер_лексеми (число, що дорівнює номеру лексеми в реченні), Текст (текст лексеми).

Крок 2. Подання на вхід для розбору на лексеми речення з таблиці речень $U_{CT}^R \in U_{CT}$.

Крок 3. Створення нової лексеми в таблиці лексем $U_{CT}^L \in U_{CT}$.

Крок 4. Накопичення символів до крапки, пропусків або кінця речення та збереження в таблиці лексем.

Крок 5. Перевірка кінця речення. Якщо так, то перехід до кроку 6, якщо ні, то збереження накопиченої послідовності у таблицю $U_{CT}^L \in U_{CT}$, розбір нової лексеми та перехід до кроку 3.

Крок 6. Проведення синтаксичного аналізу на підставі вихідних даних (алг. 2).

Крок 7. Проведення морфологічного аналізу на основі даних, одержаних на виході.

Етап 4. Визначення тематики комерційного контенту $U_{CT}^T \in U_{CT}$.

Крок 1. Побудова ієрархічної структури властивостей $U_{CT}^T \in U_{CT}$ кожної лексичної одиниці тексту, що містить граматичну та семантичну інформацію.

Крок 2. Формування лексикону з ієрархічною організацією типів властивостей, де кожен тип-нащадок успадковує і перевизначає властивості предка.

Крок 3. Уніфікація – базовий механізм побудови синтаксичної структури.

Крок 4. Визначення ключових слів *KeyWords* комерційного контенту $C_4 = \alpha_5(\alpha_4(C_2, U_K), U_{CT})$ при $U_{CT} = \{U_{CT1}, U_{CT2}, U_{CT3}, U_{CT4}\}$, де U_{CT} – колекція умов рубрикації, U_{CT1} – множина тематичних ключових слів зі словника, U_{CT2} – множина частот вживання ключових слів у комерційному контенті, U_{CT3} – множина залежностей вживання ключових слів різних тематик (коефіцієнти визначає модератор за належністю ключового слова до певної тематики в межах $[0,1]$), U_{CT4} – множина частот вживання тематичних ключових слів у контенті (алг.2).

Крок 5. Визначення $U_{CT}^T \in U_{CT}$ з *TKeyWords* – тематичні ключові слова в множині *KeyWords* для *Topic* – тема контенту та *Category* – категорія контенту.

Крок 6. Визначення *FKeyWords* – частоти вживання ключових слів та *QuantitativyTKey* – частоти вживання тематичних ключових слів у тексті комерційного контенту.

Крок 7. Визначення *Comparison* – порівняння появи ключових слів різних тематик. Розрахунок *CofKeyWords* – коефіцієнт тематичних ключових слів контенту, *Static* – коефіцієнт статистичної важливості термів, *Addterm* – коефіцієнт наявності додаткових термів. Порівняння множини ключових слів контенту з ключовими поняттями тем. Якщо є збіг, то перехід до кроку 9, якщо ні, то перехід до кроку 8.

Крок 8. Формування нової рубрики з набором ключових понять аналізованого контенту C_4 .

Крок 9. Присвоєння визначеній рубриці аналізованого комерційного контенту C_4 .

Крок 10. Розрахунок *Location* – коефіцієнта розташування контенту C_4 в тематичній рубриці.

Етап 4. Заповнення бази пошукових образів для атрибутів *Topic* – тема контенту, *Category* – категорія контенту, *Location* – коефіцієнт розташування контенту в тематичній рубриці, *CofKeyWords* – коефіцієнт тематичних ключових слів у контенті, *Static* – коефіцієнт статистичної важливості термів, *Addterm* – коефіцієнт наявності додаткових термів, *TKeyWords* – тематичні ключові слова, *FKeyWords* – частота вживання ключових слів, *Comparison* – порівняння появи ключових слів різних тематик, *QuantitativyTKey* – частота вживання тематичних ключових слів у тексті комерційного контенту C_4 .

Побудова тексту контенту C_4 визначається темою, вираженою інформацією, умовами спілкування, завданням повідомлення та стилем викладення. Із семантичною, граматичною та композиційною структурою контенту C_4 пов'язані його стильові/стилістичні характеристики, залежні від індивідуальності автора та підпорядковані тематичній/стильовій домінанті тексту. Процес рубрикації контенту C_4 у вигляді діаграми варіантів подано на рис. 1.

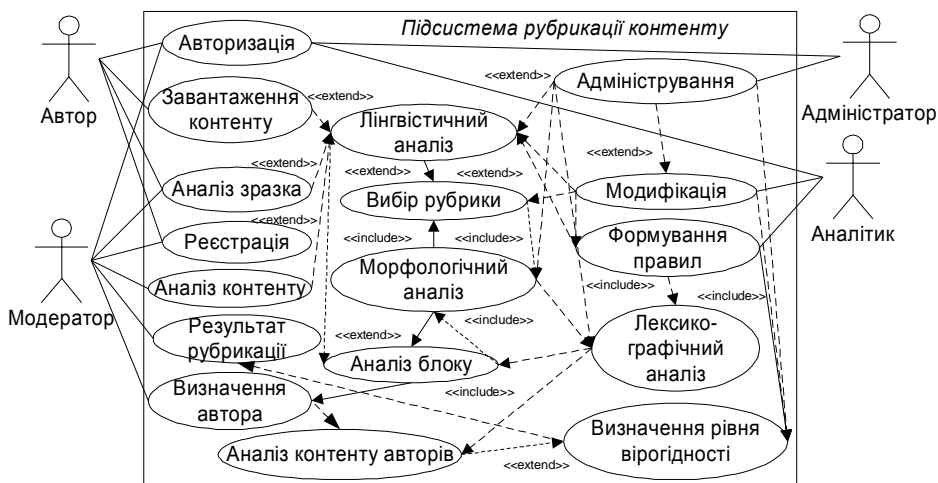


Рис. 1. Діаграма варіантів використання для процесу рубрикації контенту в СЕКК

Основні етапи визначення морфологічних ознак U_{CT} одиниць тексту C_4 : визначення граматичних класів слів – частин мови і принципів їх класифікаційного виділення; виокремлення частини семантики слова як морфологічної; обґрунтування набору морфологічних категорій та їх природи; опис сукупності формальних засобів, закріплених за частинами мови та їх морфологічними категоріями. Процес рубрикації $C_4 = \alpha_5(\alpha_4(C_2, U_K), U_{CT})$ через автоматичне індексування складових комерційного контенту C_3 розділено на послідовні блоки: морфологічний аналіз, синтаксичний аналіз, семантико-синтаксичний аналіз лінгвістичних конструкцій та варіювання змістового запису текстового контенту (рис. 2).



Рис. 2. Діаграма послідовності для процесу рубрикації контенту в СЕКК

Використано такі способи вираження граматичного значення: синтетичний, аналітичний, аналітико-синтетичний та суплетивний. Граматичні значення узагальнені через однотипні характеристики та підлягають поділу на часткові значення. Для позначення класів однотипних граматичних значень використано поняття граматичної категорії. До морфологічних значень належать категорії роду, числа, відмінка, особи, часу, способу, стану, виду, об'єднані у парадигми для класифікації частин тексту. Об'єктом морфологічного аналізу є структура слова, форми словозміни, способи вираження граматичних значень. Морфологічні ознаки одиниць тексту – це інструменти дослідження зв'язку між лексикою, граматикою, використанням їх у мовленні, парадигматикою (відмінкові форми відмінюваних слів) і синтагматикою (лінійні зв'язки слів, сполучення). Реалізація автоматичного кодування слів тексту, тобто приписування їм кодів граматичних класів, пов'язана з граматичною класифікацією. Морфологічний аналіз містить такі етапи: виділення основи у словоформі; пошук основи у словнику основ; порівняння структури словоформи з даними у словниках основ, коренів, префіксів, суфіксів, флексій. У процесі аналізу ідентифікують значення слів та синтагматичних відношень між словами контенту. Інструментами

аналізу є словники основ/флексій/омонімів та статистичних/синтаксичних словосполучень, зняття лексичної омонімії, семантичний аналіз іменних безприйменникових конструкцій, таблиці семантико-синтаксичного сполучення іменників/прикметників та компонентів прийменникових конструкцій, алгоритми аналізу для визначення послідовностей перевірок і звертань до словника і таблиць; система поділу слів тексту на флексію й основу; тезаурус еквівалентностей для заміни еквівалентних слів одним/кількома номерами понять, які слугують ідентифікаторами змісту замість основ слів; тезаурус у вигляді ієрархії понять для забезпечення пошуку для цього поняття загального/асоційованого з ним поняття; система обслуговування словників. Процес індексування залежить від дескрипторного словника або інформаційно-пошукового тезауруса. Дескрипторний словник має структуру таблиці з трьома колонками: основи слів; набори дескрипторів, приписані кожній основі; граматичні ознаки дескрипторів. Індексування складається з виділення інформативних словосполучень з тексту; розшифрування абрєвіатури; заміна слів з основами-дескрипторами на код дескриптора; зняття омонімії.

Формування дайджестів комерційного контенту. Дайджест – це короткий зміст публікації у СЕКК, для формування якого використовують контент-аналіз з урахуванням частотних ваг слів зі сформованого словника понять. Оператор формування дайджестів комерційного контенту $\alpha_6 : (C_4, U_D, T) \rightarrow C_5$ є відображенням комерційного контенту C_4 в новий стан C_5 , який відрізняється від попереднього стану появою нової частини контенту у вигляді його короткого змісту, що доповнює попередній стан. Процес формування дайджестів складається з алгоритмів формування словника понять (алг. 2) та створення дайджесту (алг. 3).

Алгоритм 2. Формування словника понять.

Етап 1. Формування словника понять.

Крок 1. Послідовне виділення всіх лінгвістичних одиниць з вхідного контенту.

Крок 2. Побудова алфавітно-частотного словника.

Крок 3. Нормалізація слів через автоматичний морфологічний аналіз.

Крок 4. Модифікація алфавітно-частотного словника.

Крок 5. Приписування словам ваги W (частоти появи).

Крок 6. Вилучення зі словника незначних слів ($W \leq k$, де k – значення порога вилучення).

Етап 2. Вибір тематичного словника відповідно до запиту.

Етап 3. Коригування алфавітно-частотного словника з урахуванням термів тематичного словника (коригування значень ваг окремих одиниць).

Етап 4. Вибір $N = n$ слів із більшою вагою (вагомих) із алфавітно-частотного словника, де $n = const$ і задається модератором.

Алгоритм 5. Створення дайджесту.

Етап 1. Вибір контенту з урахуванням його ваги.

Крок 1. Завдання розміру дайджесту C_4 .

Крок 2. Виконання алгоритму 1.

Крок 3. Послідовне визначення ваги контенту як суми значень ваг окремих лінгвістичних одиниць, тобто $W = \sum_i w_i$.

Крок 4. Сортування вхідного потоку контенту за величинами ваг.

Крок 5. Визначення змістових дублів контенту за статистичним критерієм унікальності тексту $U_D \geq 0,9$ (алг. 1).

Крок 6. Фільтрування контенту, непридатного для формування дайджестів (якщо $W \leq l$, де l – значення порога вилучення контенту, за допомогою правил структуризації та модерації контенту із самонавчанням) та статистично змістових дублів.

Крок 7. Вибір $V = q$ контенту із більшою вагою, де $q = const$ і задається модератором.

Етап 2. Побудова тексту дайджесту з відібраного контенту.

Крок 1. Побудова словника з відібраного контенту (алг. 3.8).

Крок 2. Застосування контент-аналізу до тексту (алг. 3.2).

Крок 3. Фільтрування речень, що не відповідають семантичним правилам структуризації та модераторії контенту.

Крок 4. Автоматичне формування гіпертекстового подання дайджесту, його змісту і гіперпосилання на вихідні джерела.

Етап 3. Редагування сформованого тексту дайджесту c_{i4} , де $C_5 = \{c_{i4}, C_4\}$.

Крок 1. Перевірка обсягу c_{i4} сформованого контенту C_5 . Якщо $c_{i4} < C_4$, то виконання кроку 2, інакше виконання етапу 4.

Крок 2. Видалення з вхідного потоку контенту C_4 зі сформованим дайджестом c_{i4} .

Крок 3. Виконання етапів 1–2.

Крок 4. Дописування до сформованого дайджесту отриманого та перехід до кроку 1.

Етап 4. Форматування тексту дайджесту як окремих контент C_5 та збереження в бази даних дайджестів із посиланням на джерело.

Процес формування дайджестів комерційного контенту C_5 формує множину коротких анотацій та основних положень контенту за певний період. Це зручно для швидкого ознайомлення з основним змістом певної тематики/рубрики, дослідження та пошуку необхідного комерційного контенту.

Процес розподілу комерційного контенту. Процес розподілу контенту (рис. 3) реалізує розподілення навантаження між авторами/модераторами СЕКК за активного збільшення постійної аудиторії читачів та обсягу затребуваного комерційного контенту.

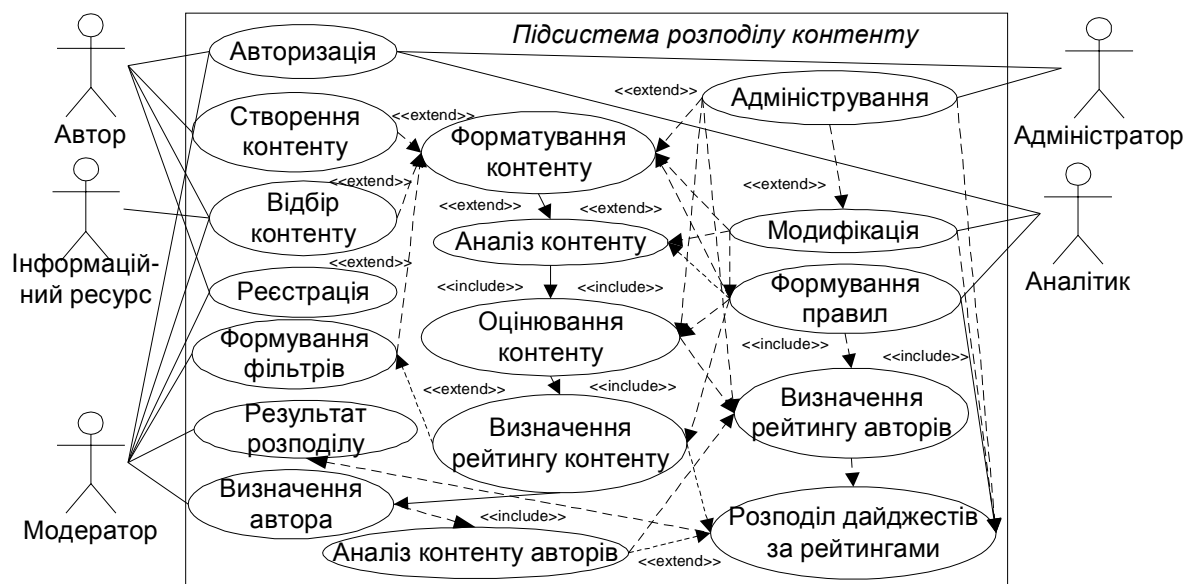


Рис. 3. Діаграма варіантів використання для процесу розподілу контенту СЕКК

Релевантний контент розсилають користувачам та завантажують у БД. Вибіркове поширення контенту описано $C_6 = \alpha_7(C_5, U_{DS})$, де U_{DS} – множина умов вибіркового поширення контенту. На рис. 4 діаграма кооперації ілюструє взаємодію модератора із модулями оцінювання матеріалу, рейтингування авторів та розподілу дайджестів. Оцінювання кожного контенту проводять в комплексі з іншим тематичним контентом (рис. 5). Одним із важливих критеріїв розподілу дайджестів між авторами є процент унікальності опублікованого контенту кожного автора. Системами визначення унікальності контенту є Praide Unique Content Analyser 2, FIndCopy та Miratools.

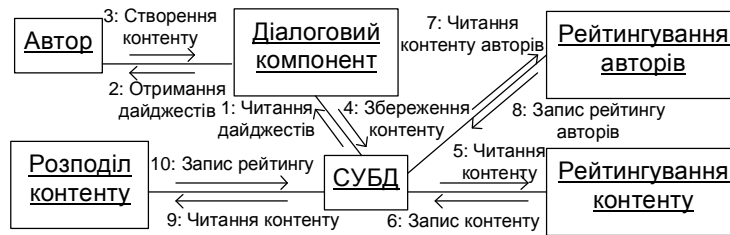


Рис. 4. Діаграма кооперації для процесу розподілу комерційного контенту



Рис. 5. Діаграма послідовності для процесу розподілу комерційного контенту

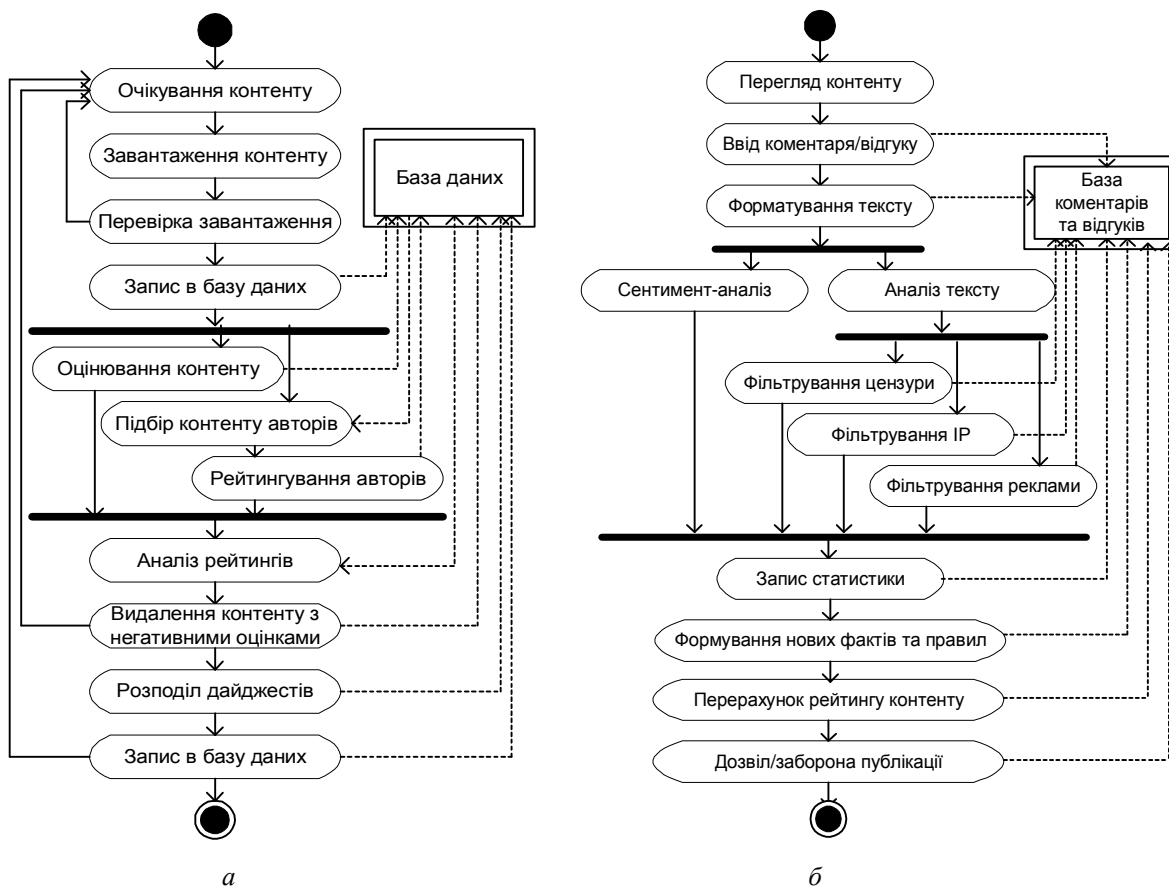


Рис. 6. Діаграма діяльності для процесу: а – розподілу контенту; б – управління контентом

Спочатку підсистема отримує готові дайджести з джерел через RSS. Далі дайджести розподіляють між авторами за їх рейтингом: першими одержують дайджести для опрацювання автори з найвищим рейтингом (рис. 6). Після виконання усіх дій підсистема переходить у стан очікування до появи нового контенту. Рейтинг авторів вказує на продуктивність/результативність роботи кожного з них окремо. Впливають на нього такі критерії U_{DS} , як процент унікальності контенту (якість роботи автора), кількість переглядів контенту (вага пошукових та прямих переходів), оцінка користувача (активність користувачів) та час перебування на сторінці з контентом (міра зацікавленості користувачів у контенті). Підсистема рейтингування оцінює якість роботи за сукупністю критеріїв, що забезпечує об'єктивність та стимулювання якісної роботи. Частка такої роботи модератора, як розподіл однотипних даних, їх сортування, оцінювання та аналіз, зменшується. Це зменшує обсяг застосованих ресурсів, пришвидшує створення контенту та покращує характеристики контенту завдяки об'єктивності оцінювання якості виконаних завдань. За процентом унікальності тексту оцінюють якість роботи автора цього контенту та вносять отриману оцінку в таблицю рейтингів.

Висновки і перспективи подальших наукових розвідок

Зазвичай розподіл виконують модератори. Підсистема розподілу контенту зменшує час та ресурси для подальшого функціонування СЕКК. Процес розподілу передбачає декілька етапів: формування списку об'єктів розподілу (наприклад, статей, програмного забезпечення, книг або дайджестів); визначення критеріїв/ознак розподілу контенту з отриманого списку (процент унікальності контенту; кількість звернень до контенту; користувацька оцінка; час перегляду); рейтингування авторів контенту; оцінювання параметрів контенту з метою використання у процесі розподілу. Наведені критерії не є однаковими за значенням та важливістю під час аналізу їх загалом та обчислення зведеної оцінки якості роботи авторів контенту. Контент містить тему та дайджест. Підсистема розподілу контенту вибірково розсилає дайджести між авторами згідно з рейтингуванням якості їх роботи. Збільшення обсягу контенту приводить до точнішого оцінювання якості та продуктивності кожного автора контенту. Збільшення кількості критеріїв оцінки дає змогу охопити ширший спектр аспектів роботи автора/модератора.

1. Берко А. Системи електронної контент-комерції / А. Берко, В. Висоцька, В. Пасічник. – Л: Нац. ун-т “Львівська політехніка”, 2009. – 612 с. 2. Иванов В. Контент-аналіз: Методологія і методика дослідження ЗМК / В. Иванов. – Київ, 1994. – 112 с. 3. Иванов С. Статистический анализ документальных информационных потоков / С. Иванов, Н. Круковская // Научно-техническая информация. – 2004. – № 2. – С. 11–14. 4. Клифтон Б. Google Analytics / Б. Клифтон. – М.: ООО “И. Д. Вильямс”, 2009. – 400 с. 5. Ландэ Д. Основы моделирования и оценки электронных информационных потоков / Д. Ландэ, В. Фурашев, С. Брайчевский, О. Григорьев. – К.: Інжиніринг, 2006. – 348 с. 6. Пасічник В. Математична лінгвістика / В. Висоцька, В. Пасічник, Ю. Щербина, Т. Шестакевич. – Л: Новий Світ, 2012. – 359 с. 7. Солтон Д. Динамические библиотечно-информационные системы / Д. Солтон. – М.: Мир, 1979. – 560 с. 8. Федорчук А. Контент-мониторинг информационных потоков // БНАН. – Киев. – 2005. – № 3. Режим доступу: www.nbuv.gov.ua/articles/2005/05fagmip.html. 9. Voiko B. Content Management Bible. – Hoboken, 2004. – 1176 p.