

УДК 004.912

Артур Федорчук,
наук. співробітник НБУВ

Катерина Лобузінa,
зав. відділу НБУВ

Ніла Танатар,
зав. відділу НБУВ

Створення інформаційних ресурсів на основі моніторингу змісту публікацій ЗМІ

Публікація присвячена питанням технології контент-моніторингу засобів масової інформації, створенню на її основі нових інформаційних ресурсів. Розглянуто історію виникнення контент-моніторингу та процес його впровадження на основі стандартних рішень, які пропонує система автоматизації бібліотек «ІРБІС-64».

К л ю ч о в і с л о в а: засоби масової інформації, контент-аналіз, контент-моніторинг, бібліографічна база даних, інформаційні ресурси.

A publication is devoted to the questions of technology content-monitoring of mass media and creation new information resources on the basis of these researches. History and advantages of introduction of technology is considered on the basis of standard decisions, the library automation system «IRBIS 64» offers.

K e y w o r d s: mass media, content-analysis, content-monitoring, bibliographic database, informative resources.

Засоби масової інформації (ЗМІ) на сьогоднішній день є одним з найпотужніших комунікативних засобів суспільства. Вони виступають сполучною ланкою, передавачем інформації від держави до суспільства і, навпаки, від суспільства до держави, крім того, втілюють і репрезентують громадську думку, виконують функції спостерігача та контролера законодавчої, виконавчої, судової влади. У громадянському суспільстві ЗМІ створюють своєрідну публічну сферу – загальну громадську трибуну, з якої оприлюднюються, лобіюються суспільні і приватні інтереси. Тому не буде перебільшенням сказати, що динамічний технологічний та методологічний розвиток ЗМІ у сучасному світі дедалі більше впливає не тільки на ступінь поінформованості окремих громадян, але й на характер, напрями та темпи розвитку всього суспільства.

Не менш важливою є роль засобів масової інформації і в українському суспільстві, у формуванні поглядів і уподобань громадян нашої країни. Наприклад, суспільно-політична інформація, що з'являється у засобах масової інформації, досить точно відображає стан, структуру політичної і суспільної системи України, взаємозв'язки основних інститутів влади і

громадянського суспільства, вона справляє доволі істотний вплив на стратегічні засади української державності, на визначення шляху її розвитку. Тому є всі підстави вважати, що вітчизняні ЗМІ, з одного боку, виступають індикатором, який дає змогу визначити ступінь зрілості та структурування нашого суспільства, а, з другого – у методологічному сенсі можуть бути тим дескриптором, за допомогою якого можливий опис, інтерпретація та оцінка перспектив суспільно-політичної системи загалом [4].

У зв'язку з наведеним вище стає зрозуміло, чому в останні роки саме ЗМІ приділяється велика увага з боку держави, громадськості, вони дедалі частіше стають об'єктом ретельного аналізу науковців.

Сьогодні практично не існує межі між мережевими ЗМІ та традиційною пресою, яка, в свою чергу, з кожним днем розширює свою присутність в мережі. Те, що відбувається в цій галузі, багато дослідників називають «четвертою комунікаційною революцією», яка характеризується розвитком супутникових і кабельних телекомунікацій, а також створенням комп'ютерних мереж та програмних засобів, здатних забезпечити накопичення і передачу величезних масивів

інформації в глобальному масштабі. Сьогодні Інтернет перетворюється на одне з найважливіших ЗМІ, що забезпечує, з одного боку, додатковий канал для розповсюдження інформації, а, з другого – надходження різноманітних даних для інших ЗМІ. Крім того, конвергенція (взаємне проникнення) інформаційних інтернет-ресурсів та ЗМІ створює передумови для розширення аудиторії традиційних ЗМІ, об'єднання її в часі і просторі з аудиторією Інтернету.

Міграція різних форм ЗМІ в Інтернет демонструє розвиток принципово нових видів інформаційних процесів, які інтегрують історичний досвід традиційних ЗМІ та нові можливості медіа. Іншими словами, Інтернет стає новим відкритим простором для збирання, зберігання і розповсюдження інформації, важливим чинником розвитку публічної сфери та громадянського суспільства. З погляду забезпечення якості будь-яких наукових досліджень суспільного стану, чинник взаємного проникнення електронних та традиційних ЗМІ висуває додаткові вимоги до технологій моніторингу інформаційних потоків, які вони породжують.

Методологія контент-моніторингу

Все вищезазначене зумовлює нагальну потребу широкого використання та вдосконалення сучасних методик автоматизованого опрацювання великих масивів соціально-політичної інформації, яку несуть у собі сучасні ЗМІ. Традиційні інформаційно-пошукові системи у цих питаннях поступово втрачають свою актуальність, що зумовлено як швидко зростаючими обсягами інформаційних потоків, так і постійним оновленням інформації. Узагальнення значних інформаційних масивів, які безупинно генерують сучасні ЗМІ, потребує якісно нових підходів. Більшість авторитетних дослідників у світі сьогодні остаточно дійшли висновку, що завдання систематичного дослідження тенденцій і процесів у сучасному динамічному інформаційному середовищі може бути розв'язане лише завдяки моніторингу змісту інформаційних потоків. Цей перспективний напрям досліджень, що використовується для безперервного отримання якісних і кількісних зрізів інформаційних потоків протягом певного проміжку часу, у спеціальній літературі отримав назву контент-моніторинг (КМ) [12].

Інструментом контент-моніторингу, зазвичай виступає контент-аналіз (КА) текстів, який в останні роки широко використовується в різноманітних дослідженнях і набув різних інтерпретацій, методичних форм саме завдяки його використанню в

процесі моніторингу інформаційних потоків. У найпростішому вигляді ідею контент-моніторингу можна сформулювати як постійне виконання вузько окресленого своїми завданнями контент-аналізу безперервних інформаційних потоків.

Загально визнаним є розподіл методологій контент-аналізу на дві гілки: кількісну і якісну. Основа кількісного контент-аналізу – частота появи в документах визначених характеристик змісту. Якісний контент-аналіз ґрунтується на самому факті присутності або відсутності у тексті однієї або декількох характеристик змісту [11].

Методи проведення якісного контент-аналізу ґрунтуються на тому, що на будь-якій фазі кількісного контент-аналізу для оцінок результатів може бути залучений експерт. Якісний контент-аналіз, таким чином, покликаний забезпечити експерта необхідними технологічними засобами для висновків і додаткового аналізу. Експерт за допомогою таких засобів може виявити визначені властивості частини інформації і перевірити їх стосовно загального текстового потоку, а загальні властивості текстового потоку поширити на його конкретну тематичну частину.

Контент-моніторинг преси

Розглянемо використання методу КМ на конкретному прикладі моніторингу інформаційного потоку суспільно-політичного спрямування, породженого газетною періодикою України. Дане дослідження було започатковане у 1993 р. і проводиться відділом організації і використання документного фонду Фонду Президентів України Національної бібліотеки України імені В. І. Вернадського по сьогоднішній день. За цей час набуто значного досвіду моніторингу великих за обсягом інформаційних потоків на основі використання методу контент-аналізу та створення автоматизованих інформаційно-аналітичних систем.

Технологія КМ складається з низки послідовних процедур: упорядкування списку першоджерел для перегляду, розроблення методик відбору і класифікації матеріалів, їхнього автоматизованого опрацювання, аналізу занесеної до бази даних (БД) інформації та формування результатів моніторингу преси.

Автоматизована технологія контент-моніторингу має декілька важливих особливостей. По-перше, формування інформаційних масивів, що використовуються в процесі дослідження, є поєднанням двох автоматизованих процесів: аналітико-синтетичної переробки та багаторівневої процедури контент-аналізу текстів публікацій. По-друге, в якості

одиниці формування текстового інформаційного масиву і, відповідно, одиниці збереження та аналітичного дослідження використовується ключовий фрагмент публікації, індексація якого відбувається за допомогою багатофасетної класифікації, що має можливість динамічно змінюватися залежно від потреб дослідження. Унікальність запропонованої технології полягає у поєднанні змістовних та кількісних методів аналізу текстів публікацій.

Для контент-моніторингу використовуються монотематичні за наповненням системи з можливостями багатоаспектного використання інформації при аналізі та формуванні вихідних інформаційно-аналітичних матеріалів.

Автоматизована технологія контент-моніторингу протягом багатьох років експлуатації постійно вдосконалювалася як у технологічно-методичному плані, так і в плані програмного забезпечення. В якості інструменту зберігання інформаційних масивів спочатку використовувалася програма *Abstrac*, яка дає змогу зберігати та управляти БД в форматі *MARC*. Для аналітико-синтетичної переробки, автоматичного аналізу та формування інформаційно-аналітичних матеріалів використовувалися власні програмні засоби.

Поліпшення комп'ютерного та програмного забезпечення, формування локальної мережі та поява нових завдань щодо створення електронних інформаційних ресурсів, забезпечення доступу до них читачів спричинили необхідність адаптації технології контент-моніторингу до нових загально-бібліотечних програмних засобів і систем автоматизації бібліотеки.

Система автоматизації бібліотек «ІРБІС»

Система автоматизації бібліотек (САБ) «ІРБІС» розроблена Державною публічною науково-технічною бібліотекою Росії (ДПНТБ). Вона є типовим інтегрованим вирішенням питань автоматизації бібліотечних технологій. Система може використовуватися в бібліотеках будь-якого типу і профілю, вона повністю відповідає міжнародним вимогам, що висуваються до такого роду систем та підтримує всі вітчизняні бібліографічні стандарти і формати [9, 13]. Крім того вона дає змогу описувати всі види видань, а інтерфейси системи максимально наближені до потреб користувача і легко освоюються.

У НБУВ упровадження САБ «ІРБІС» розпочалося 2004 р., коли у відділі формування музичних фондів використали цю систему в якості середови-

ща створення електронного каталогу [8]. Потім система «ІРБІС» використовувалася при створенні ілюстративних електронних колекцій стародруків, каталогів аркушевих образотворчих видань, електронного каталогу Інституту рукопису НБУВ, для формування різних бібліографічних баз даних [1; 5; 6; 7].

Нову версію системи автоматизації бібліотек – «ІРБІС-64» – почали впроваджувати у НБУВ з 2009 р. Завдяки їй відкрилися реальні перспективи автоматизації практично всіх галузей бібліотечно-інформаційної діяльності.

САБ «ІРБІС-64» повністю сумісна з російським, українським та міжнародними форматами представлення бібліографічної інформації на основі засобів двосторонньої конверсії даних. Система дає змогу створювати і підтримувати будь-яку кількість баз даних, що становлять електронний каталог (ЕК) або є проблемно-орієнтованими бібліографічними базами даних (БД) [2; 3].

Система включає засоби, що дозволяють використовувати як ілюстративний матеріал будь-які зовнішні щодо бібліографічного документа об'єкти: повні тексти, графіку, таблиці, аудіо- і відеоматеріали, а також ресурси Інтернет. Система працює в архітектурі клієнт-сервер, забезпечуючи взаємодію клієнтських застосувань (призначених для користувача АРМ) і сервера БД на основі протоколу *TCP/IP* як в локальних, так і в глобальних мережах [10].

САБ «ІРБІС-64» є сучасним проблемно-орієнтованим програмним продуктом. Її робота ґрунтується на клієнт-серверній платформі, що дає змогу організувати одночасну роботу із бібліографічною базою даних багатьох користувачів, налагодити клієнтські профілі відповідно до спеціалізованих завдань співробітників.

Система може підтримувати каталогізацію та опис документів будь-якого виду, у т. ч. документів спеціалізованих фондів: аудіо- та відеоматеріали, газети, електронні ресурси, карти, ноти. Вона має зручний інтерфейс для опрацювання періодичних видань, зокрема газетної продукції, а також автоматизовані механізми аналітичного розпису видань за їх змістом, з можливістю вводити реферати та анотації публікацій.

У ній також наявні розвинуті механізми підтримки систематизації, тематичного упорядкування зібраних документів: рубрикатори, бібліотечні класифікації, словники предметних рубрик.

Бібліографічний опис документів може супроводжувати будь-яка кількість зовнішніх об'єктів,

зокрема і ілюстративний матеріал, і повні тексти, і графіка, і аудіо- або відеоресурси Інтернету.

Вихідні формати бібліотечно-бібліографічних баз даних можна отримати як у форматі RTF для Microsoft Word, так і у форматі HTML, призначеному для онлайнної публікації опрацьованих матеріалів.

Формати представлення даних, робочі листи, допоміжно-довідковий апарат баз даних, пошукові словники відкриті та доступні для налагодження і перебудови відповідно до проблемно-орієнтованих завдань.

Всі ці переваги програмного бібліотечного продукту нового покоління створюють передумови для професійної організації бібліографічної та інформаційно-аналітичної діяльності бібліотеки.

У 2010 р. САБ «ІРБІС-64» була впроваджена відділом організації та використання документного фонду Фонду Президентів України в якості середовища технології контент-моніторингу ЗМІ.

Бібліографічна база даних

«Інформаційно-аналітичний огляд преси»

На основі стандартних рішень, які пропонує «ІРБІС-64», було розроблено спеціалізовану базу даних «Інформаційно-аналітичний огляд преси», адаптовану до потреб Відділу та забезпечено її експлуатацію в мережевому режимі.

База даних включає записи двох основних видів:

1. «ОПИС ГАЗЕТИ»

- Назва газети
- Колектив
- Вихідні дані
- Посилання на Інтернет-версію видання
- Опрацьовані надходження
- Мова тексту
- Країна видання

Такий різновид запису дає змогу створити список усіх видань, що опрацьовуються у процесі аналітичного огляду і зберегти посилання на Інтернет-версію видання, за яким можна автоматично, прямо із бібліографічного опису, перейти до сторінок онлайнної електронної публікації.

2. «АНАЛІТИЧНИЙ ОПИС»

Основний опис

- Газета
- Мова публікації
- Автор
- Інші автори
- Заголовок
- Кореспонденти, редактори
- Колективи (авторська відповідальність)

Систематизація

- Анотація (текст публікації)
- Рубрики
- Географія (регіони)
- Ключові слова

Файли

- Ім'я файла
- Інтернет-адреса джерела інформації

Цей запис призначений для збереження необхідної інформації про матеріал публікації, він має інструменти для аналітичного опрацювання інформації та її систематизації, а також можливість упорядкувати повнотекстовий ресурс за матеріалами преси.

До бібліотечно-інформаційної бази даних адаптовані пошукові словники, що надають необхідні точки доступу до опрацьованих матеріалів:

1. ГАЗЕТА (джерело публікації)
2. ГАЗЕТА (загальний опис)
3. Дата виходу газети
4. Мова видання
5. АВТОР
6. НАЗВА публікації
7. Ключові слова
8. Місце видання
9. РУБРИКА
10. Об'єкт
11. Суб'єкт
12. Тема
13. Преса
14. Регіон
15. Тональність
16. Тип документа
17. Статистика (введення нових записів)
18. Статистика (редагування записів)

Для кожного користувача розроблено персональний профіль, де розділено доступ до інформаційних ресурсів, які він має опрацьовувати, та зареєстровано прізвище співробітника, за яким автоматично ведеться статистика його роботи із записами бази даних.

Для опрацьованих матеріалів налагоджено вихідну форму у форматі RTF, яка виводить аналітичний матеріал в упорядкованому ієрархічному вигляді відповідно до введених тематичних та географічних рубрик і може бути представлена замовнику за будь-який період часу.

Вдалося також вирішити питання конвертації БД, що свого часу були створені в форматі MARC за допомогою програми Absmarg та використовувались як інформаційні масиви в технології контент-моніторингу відділу.

Висновки

Поєднання можливостей інтегрованого середовища САБ «ІРБІС-64» та технології контент-моніторингу зробило процес дослідження більш технологічним і зручним для аналітиків, допомогло отримати унікальний інформаційно-аналітичний продукт широкого спектру використання та представлення. Сьогодні користувачеві (читачеві) можуть бути надані як структуровані, відповідно до запиту, тематичні інформаційно-аналітичні матеріали за будь-який відрізок часу, аналітичні та статистичні результати досліджень, так і доступ до БД у середовищі САБ «ІРБІС-64» з можливостями різноманітного пошуку та відбору потрібної інформації.

Спеціалізована база даних забезпечує зберігання інформаційних масивів контент-моніторингу ЗМІ і дає змогу вирішувати такі основні завдання:

- створення зручних у користуванні повнотекстових баз даних зі збереженням принципів фасетної індексації та анотації у вигляді ключового фрагмента публікації;
- забезпечення поєднання в одному дослідженні традиційних та мережевих ЗМІ, полегшення процесу пошуку, аналізу та бібліографічного опису публікацій як суто мережевих ЗМІ, так і інтернет-версій періодичних видань;
- задоволення потреб користувачів щодо високоякісної інформації з визначеної суспільно-політичної тематики;
- можливість підготовки різноманітних інформаційно-аналітичних матеріалів, структурованих відповідно до мети дослідження.

Список використаної літератури

1. Бодак О. П. Електронні інформаційні ресурси Інституту рукопису НБУВ: історія створення та перспективи розвитку / О. П. Бодак // Рукописна та книжкова спадщина України. – К., 2009. – Вип. 13. – С. 449–457.
2. Бродовский А. И. Интегрированная библиотечно-информационная система ИРБИС – современное средство для автоматизации малых и средних библиотек / А. И. Бродовский // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: материалы конф.: Судак, Автоном. Респ. Крым, Украина, 6–14 июня 1998г. / 5-я юбилейная международная конф. «Крым-98». – М., 1998. – Т. 1. – С. 114–118; Научные и технические библиотеки. – 1999. – № 1. – С. 87–94.
3. Бродовский А. И. Новое поколение системы автоматизации библиотек ИРБИС – ИРБИС64: от электронного каталога к полнотекстовым базам данных / А. И. Бродовский, К. О. Сбойчаков // Научные и технические библиотеки: ежемес. сб. по вопросам теории и практики библиотечного дела / ГПНТБ России. – 2005. – № 2. – С. 107.
4. Беззуб'як М. Й. Дослідження політичних відносин методом контент-аналізу: автореф. дис. ... канд. політ. наук: 23.00.01 / М. Й. Беззуб'як; Інститут політичних та етнопольових досліджень НАН України. – К., 2003. – 18 с.
5. Ілюстрований електронний каталог «Радянський лубок із фондів Національної бібліотеки України імені В. І. Вернадського» / К. Лобузїна, О. Донець // Бібл. вісн. – 2009. – № 1. – С. 16–20.
6. Ковальчук Г. И. Создание цифровых коллекций книжных памятников в Национальной библиотеке Украины им. В. И. Вернадского / Г. И. Ковальчук, К. В. Лобузїна // Матеріали Чацвєртых Міжнародных Книгзнаучьих чытанняў: «Новыя тэхналогіі ў захаванні документальных помнікаў» (Мінск, 20–21 лістапада 2008 г.). – Мінск, 2008. – С. 120–124.
7. Ковальчук Г. И. Электронная коллекция старопечатных книг Национальной библиотеки Украины имени В. И. Вернадского / Г. И. Ковальчук, К. В. Лобузїна // Машиночитаемая каталогизация старопечатной книги: материалы обучающих семинаров, прошедших в Российской национальной библиотеке в рамках сотрудничества с Консорциумом европейских научных библиотек (CERL) в 2006–2008 гг. – СПб.: РНБ, 2009. – С. 130–137.
8. Лобузїна К. В. Музичний фонд НБУВ : створення електронного каталогу / К. В. Лобузїна, Л. В. Івченко // Бібл. вісн. – 2007. – № 4. – С. 12–22.
9. Руководство по UNIMARC / под ред. Браїена П. Холта, при участии Салли Х. МакКаллум и А. Б. Лонга; пер. на рус. яз. авторского коллектива под руководством А. И. Земского, Я. Л. Шрайберга. – М., 1992. – 319 с.
10. Система автоматизации библиотек ИРБИС-64. Общее описание системы. – М.: ГПНТБ России, 2004. – 259 с.
11. Федорчук А. Г. Контент-мониторинг информационных потоков / А. Г. Федорчук // Библиотеки национальных академий наук: проблемы функционирования, тенденции развития: науч.-практ. и теорет. сб. – К., 2005. – Вып. 3. – С. 141–150.
12. Федорчук А. Г. Теоретико-методичні засади аналізу інформаційного потоку соціально-політичного спрямування / А. Г. Федорчук, Н. В. Танатар // Бібліотекознавство, документознавство, інформологія. – 2004. – № 2. – С. 33–38.
13. UNIMARC Manual. Bibliographic Format: Руководство по применению UNIMARC для библиографических данных: фрагменты 2-го изд.: изм. и доп. к 1-му изд. / Международная федерация библиотечных ассоциаций и учреждений (ИФЛА); ГПНТБ России. – М., 1998. – 52 с.