

Ніла Танатар,

заввідділу організації та використання документального фонду ФПУ НБУВ,
канд. біол. наук

ІНФОРМАЦІЙНО-АНАЛІТИЧНІ РЕСУРСИ ІЗ СОЦІАЛЬНО-ПОЛІТИЧНОЇ ПРОБЛЕМАТИКИ: ЗАГАЛЬНИЙ ОГЛЯД

У статті розглянуто особливості створення інформаційно-аналітичних систем із соціально-політичної проблематики та подано їх загальну характеристику

Ключові слова: контент-аналіз, контент-моніторинг, інформаційні ресурси, інформаційно-аналітичні системи.

Актуальність та особливості інформаційно-аналітичного забезпечення управлінських структур. Інформація є одним з основних елементів соціального управління, тому значення інформаційного забезпечення органів державної влади важко переоцінити. Таку інформацію, необхідну владним структурам для конкретних впливів на об'єкт управління, варто розглядати як стратегічний ресурс, тобто вона повинна бути повною, актуальною, достовірною, охоплювати весь спектр проблем, що цікавлять керівні органи, збиратися регулярно й у необхідних обсягах. Виконання цих умов при формуванні інформаційних ресурсів допомагає управлінським структурам розробляти дієвий план, визначати реальні й точні цілі своїх рішень. Тому сьогодні ЗМІ та інформаційні потоки, які вони породжують, стали одними з найважливіших об'єктів дослідження суспільно-політичних процесів України.

Нині Національна бібліотека України імені В. І. Вернадського (далі – НБУВ), володіючи найбільшим в Україні документним ресурсом та потужним кадровим потенціалом, перетворилася на один з основних інформаційних центрів обслуговування вищих органів влади й управлінських структур [1]. Спеціалізовані бібліотечні підрозділи в змозі відслідковувати на інформаційному рівні глобальні та локальні суспільні процеси, здійснювати інформаційний супровід різних політичних кампаній, у тому числі загальнодержавного значення.

Корисним у даному контексті є досвід інформаційного забезпечення апарату вищих органів державної влади, насамперед Адміністрації Президента України, відділом організації та використання документального фонду Фонду президентів України НБУВ (далі – Відділ). Особливості

такого забезпечення полягають передусім в обслуговуванні вузького кола споживачів зі спеціальними та специфічними завданнями, а його ефективність, крім перелічених вимог, залежить від чіткого поставлення завдань і безперервної взаємодії зі споживачами інформаційних матеріалів. Унаслідок цього надзвичайно важливим і, крім того, дуже складним є окреслення інформаційних потреб таких споживачів. Від розв'язання цієї проблеми багато в чому залежить ефективність самого процесу підготовки інформації.

Не менш відповідальною процедурою є забезпечення й дотримання послідовності та безперервності всього циклу формування інформаційних масивів, починаючи від процедури встановлення джерел інформації, методик її відбору та класифікації, наповнення баз даних (далі – БД), до автоматизованого опрацювання та аналізу. Сам процес формування інформаційних потоків має бути детально продуманим та проробленим, інформація повинна відбиратися дуже ретельно і будь-якого надлишку для уникнення інформаційного шуму. Перевагу потрібно надавати власним БД, створюваним під конкретні завдання. Для коректного відбору інформації необхідно проводити постійне спостереження за рядом параметрів, які характеризують той чи інший соціально-політичний процес або явище. Методологія моніторингу повинна реалізовуватись у текстовій і статистичній формі, її принципи мають бути закладені в описі функціонування технічних систем.

Інформаційно-аналітичні системи за ключовими фрагментами публікацій: особливості формування та можливості використання. З урахуванням вказаних особливостей такої інформації та вимог конкретних замовників 1992 р. у Відділі було започатковано новий клас систем контент-моніторингу соціально-політичних процесів у суспільстві на базі ключових цитат публікацій [5]. Для наповнення БД було запропоновано використовувати не повні документи, а їх окремі фрагменти, які точно й стисло передають суть того чи іншого соціального явища, політичної події. Сама процедура контент-аналізу публікацій спрямована на виявлення та виділення з тексту фрагментів публікацій, які в межах досліджуваної проблеми мають риси найменшої цілісної одиниці (модуля) інформації.

Для подальшого індексування цитат у рамках виділеного модуля визначаються окремі елементи проблеми і зв'язки між ними на основі зіставлення з конкретними значеннями класифікатора. Для цієї процедури запропоновано багатофасетні класифікатори з теоретично необмеженою кількістю фасет та значень кожної з них, що дає змогу формувати для кожного аспекту проблеми свою фасету з відповідним їй класом значень

та створювати інформаційні системи для складних і багатопланових тем. Процесу побудови класифікатора, який у схематичному вигляді повинен відтворювати модель досліджуваної проблеми, у кожному конкретному випадку передусім аналіз досліджуваної проблеми для виявлення її складових елементів і взаємозв'язків між ними. Як найпридатніші для формального опису документів при створенні інформаційних систем за ключовими фрагментами публікацій були обрані класифікатори фасетно-ієрархічного типу. У них кожному елементу досліджуваної проблеми відповідає своя фасета з фіксованим місцерозташуванням у фасетній формулі, а сукупності значень у межах окремого елемента – відповідний фасетний клас. При цьому класифікатор не є консервативним за структурою й змістом складових елементів, а розвивається й удосконалюється відповідно до розвитку проблеми. На рівні структури додаються нові елементи із сукупністю відповідних їм значень, паралельно фасетна формула доповнюється адекватними цим елементам фасетами й фасетними класами.

На основі розробленого класифікатора за допомогою фасетної формули здійснюється формальний опис конкретного фрагмента тексту, а сукупність фасетних формул усіх виділених фрагментів забезпечує формальний опис усього документа в контексті досліджуваної проблеми. Такий спосіб опрацювання першоджерел виконує функцію інформаційного «сита», через яке проходить лише релевантна досліджуваній темі інформація у вигляді цитат оригінального тексту. Занесений до БД вхідний документ містить повний бібліографічний опис публікації та всі виділені з тексту й заіндексовані за змістом фрагменти.

Спеціально розроблені програмні засоби дають змогу розбивати вхідний документ (так званий інформаційний об'єкт) на окремі незалежні фрагменти (інформаційні модулі), кожен з яких при цьому доповнюється бібліографічними даними, і формувати на основі всіх занесених до БД документів повну інформаційну матрицю – інформаційний відбиток досліджуваної проблеми. Крім того, кожен інформаційний модуль такої матриці в разі запуску спеціальних програм розбивається на три незалежні складові, кожна з яких за потребою може функціонувати самостійно. Перша складова – змістовна інформація (цитата документа), друга – його повний бібліографічний опис і третя – структурна характеристика виділеного фрагмента, передана за допомогою фасетної формули. Таким чином, весь масив інформації, поданий у вигляді повної публікаційної матриці, може аналізуватися як окремо за кожною з виділених складових, так і в їх сукупності. У випадку сортування цитат публікацій за заданими схемами отримуємо ті чи інші аналітичні огляди, на основі опрацювання бібліографічних записів

формуються бібліографічні списки, а різні маніпуляції з фасетними формулами надають можливість будувати різні статистичні та ілюстративні матеріали. При одночасному опрацюванні (за допомогою спеціально написаних програм) усіх складових відібраних цитат формуються аналітичні огляди у вигляді структурованих за певною схемою ключових фрагментів публікацій з повним або частковим їх бібліографічним описом, доповнених таблицями, графіками, діаграмами тощо. Такий спосіб формування та розбивання вхідної інформації забезпечує в подальшому створення різних варіантів текстових файлів і таблиць кількісних показників. Технологія надає можливість вибирати з усього масиву занесених до БД документів лише ту інформацію і з такою структурою, з якою вона необхідна замовнику. За допомогою програмних засобів фрагменти текстів з різних першоджерел перетворюються на структуровані текстові блоки.

Таким чином, автоматизована технологія контент-моніторингу базується на:

- використанні як одиниці формування текстового інформаційного масиву, ключового фрагмента публікації;
- індексації ключових фрагментів публікацій за допомогою багатофасетної класифікації;
- формуванні банку ключових фрагментів публікацій у процесі поєднання двох автоматизованих процесів: багаторівневої процедури контент-аналізу текстів публікацій та їх аналітико-синтетичного перероблення.

Особливість запропонованої технології насамперед полягає в застосуванні як кількісних, так і якісних методів аналізу текстів публікацій, а також у збереженні авторського змісту та забарвленості публікації.

Інформаційно-аналітичні ресурси на базі автоматизованої системи за ключовими фрагментами публікацій. Уперше таку систему було побудовано для аналізу процесу обговорення проекту нової Конституції України в пресі (жовтень 1992 р.) [2].

Необхідно було проаналізувати обговорення цього документа в публікаціях, відібраних у процесі прес-моніторингу газетної періодици: у цілому проекту, окремих його статей, а також пропозиції щодо їх змін, доповнення, вилучення та ін. Усе це потребувало точного передавання тексту оригіналу.

Принципи формування БД типових інформаційних систем, за винятком повнотекстових, не відповідали цій умові. Однак повнотекстові БД виявилися занадто громіздкими для проведення подібних досліджень. Тому для наповнення БД було запропоновано використовувати не цілі документи, а їх окремі фрагменти, які точно й стисло передавали зміст кожної пропозиції.

Для індексування відібраних цитат було розроблено двофазетний класифікатор, створені програмні засоби конвертування вхідної інформації у структуровані згідно зі статтями проекту нової Конституції вихідні текстові файли у вигляді постатейно об'єднаних цитат публікацій. Текстова складова доповнювалася бібліографічними описами, окремо формувалися табличні матеріали, які також передавали структуру винесеного на всенародне обговорення проекту.

Система користувалася попитом у розробників проекту нової Конституції, і основні принципи її побудови було взято за основу при створенні ряду інших систем соціально-політичної тематики. У зв'язку з Постановою ВР України від 17.06.1993 р. «Про проведення референдуму про довіру (недовіру) Президентів та парламенту 26.09.1993 р.» у рамках аналогічної системи здійснювався контент-моніторинг з теми «Преса про референдум» (липень – вересень 1993 р.). В обох випадках система була вузькопроблемною, відслідковувала лише один соціально-політичний процес.

Виходячи з нових більш складних завдань, були переглянуті принципи побудови інформаційно-аналітичної системи, доопрацьовані методики аналізу й відбору матеріалу в процесі прес-моніторингу газет, його індексації, розширені можливості класифікаційних схем [12]. У результаті проведеної роботи монотематичну за наповненням систему можна було використовувати як багатоаспектну. Це дало змогу в рамках системи відслідковувати події, в яких поряд з політичними процесами фіксувалися конкретні об'єкти дослідження. Так, у випадку передвиборних кампаній завдяки вдосконаленню процедур індексування та сортування занесених до БД фрагментів публікацій можна було аналізувати передвиборний процес із усіма його нюансами, починаючи від обговорення закону про вибори до різноманітних засобів впливу на виборців і претендентів на звання депутата чи Президента, аж до порушень закону та терористичних актів. Така система дає змогу також відслідковувати поведінку виборчих блоків, політичних партій, кандидатів, представників вищих ешелонів влади, вивчати настрої регіонів тощо.

Тому, починаючи з 1994 р., зусилля співробітників Відділу були перш за все спрямовані на дослідження в процесі прес-моніторингу ЗМІ найважливіших для України в той чи інший період її сучасного розвитку соціально-політичних подій – виборів до ВР і Президента України. На основі контент-моніторингу матеріалів газетної періодики формувалась інформаційно-аналітична система з таких тем [7, 9–11]:

- «Преса про вибори до ВР України-1994» (1994 р.);
- «Преса про вибори до ВР України-1998» (1997–1998 рр.);
- «Преса про вибори до ВР України-2002» (2002 р.);
- «Преса про вибори до ВР України-2006» (2006 р.);
- «Преса про вибори до ВР України-2007» (2007 р.);
- «Преса про вибори Президента України-1994» (1994 р.);
- «Преса про вибори Президента України-1999» (1999 р.);
- «Преса про вибори Президента України-2005» (2004–2005 рр.);
- «Преса про вибори Президента України-2010» (2009–2010 рр.).

Започаткована для всебічного аналізу політичних баталій під час виборчих кампаній 1994 р. інформаційна система завдяки програмному інструментарію багатоаспектного використання БД переросла в більш глобальну, яка протягом багатьох років давала можливість відслідковувати імідж політичної еліти та політичних партій України [8]:

- «Преса про політичних діячів України» (1994 р. – по сьогодні);
- «Преса про політичні партії України» (1996–2007 рр.).

У процесі вдосконалення та розвитку технології системи стало можливим аналізувати інформацію про непов'язані між собою політичні події та явища, до того ж у різних аспектах та ракурсах. При цьому передбачені різноманітні варіанти використання та подання такої інформації. Залежно від потреб замовника, система контент-моніторингу ЗМІ може функціонувати як політематична чи як монотематична. З 1996 р. у рамках такої системи, крім вищезгаданих тем, в окремі періоди відслідковувалися законотворчі акти, соціально-політична ситуація в різних регіонах України, фіксувалися висвітлені пресою поточні актуальні події, відносини України з іншими державами тощо.

Основні з таких тем [3, 4, 6]:

- «Політична еліта України про "Закон про владу"» (1994–1995 рр.);
- «Преса про соціально-політичну ситуацію в Криму» (1995–1996 рр.);
- «Преса про українсько-російські відносини» (1996 р.);
- «Обговорення проекту нової Конституції (1996 р.) в пресі» (1996 р.);
- «Преса про вибори президента Росії» (1996 р.);
- «Преса про п'ятиріччя незалежності України» (1996 р.);
- «Преса про соціально-політичну ситуацію на Донбасі» (1996 р.);
- «Актуальні події в дзеркалі преси» (1997–1998 рр.);
- «Всеукраїнський референдум 2000 року» (2000–2001 рр.);
- «Реакція преси на президентський та альтернативний законопроект "Про внесення змін до Конституції України"» (2000–2001 рр.);
- «Бюджет-2001: прес-моніторинг» (2000 р.);
- «Бюджет-2002: прес-моніторинг» (2001 р.);

- «Стосунки між парламентом та урядом: прес-моніторинг» (2001 р.);
- «Обговорення законопроектів у пресі» (2001 р.).

Висновки. Набутий досвід показав, що запропонована система контент-моніторингу дає змогу максимально повно й концентровано зображати будь-який соціально-політичний процес, відображений у ЗМІ, у сформованих інформаційних масивах з подальшим їх аналізом та синтезом.

Вищенаведений перелік досліджуваних протягом 18-ти років актуальних для українського суспільства подій шляхом контент-моніторингу газетної періодики (у друкованому, а останніми роками в електронному форматі) демонструє унікальний за змістом і структурою та значний за обсягом (БД нараховують понад 200 тис. записів) інформаційно-аналітичний ресурс із соціально-політичної тематики, яким на сьогодні володіє НБУВ.

До недавнього часу формування БД системи проводилося у форматі Mark та на базі програми Absmark. Для подальшого інформаційно-аналітичного опрацювання БД використовувалися власні програмні продукти. Враховуючи нові завдання НБУВ з інформатизації бібліотеки і загальнодоступності її фондів та ресурсів, з 2010 р. наповнення та функціонування системи контент-моніторингу ЗМІ здійснюється в середовищі Irbis. Це дає змогу доповнювати БД за ключовими фрагментами публікацій повнотекстовими БД.

Список використаної літератури:

1. *Горовий В.* Бібліотеки як сучасні центри української інформатизації / В. Горовий // Наук. пр. Нац. б-ки України ім. В. І. Вернадського. – К. : НБУВ, 2009. – Вип. 25. – С. 23–35.
2. *Заєць А.* Зі шпальт газет – до проекту: за матеріалами обговорення проекту нової Конституції України у пресі / А. Заєць, Н. Танатар // Голос України. – 1993. – 4 черв. – № 103. – С. 12.
3. *Танатар Н. В.* Бюджет-2001: контент-анализ материалов прессы / Н. В. Танатар // Финансовые риски. – 2000. – № 4. – С. 5–11.
4. *Танатар Н. В.* Бюджет-2002: контент-анализ материалов прессы / Н. В. Танатар // Финансовые риски. – 2001. – № 4. – С. 5–11.
5. *Танатар Н. В.* Інформаційно-аналітичні системи за ключовими фрагментами публікацій як основа для створення електронної бібліотеки соціально-політичного спрямування / Н. В. Танатар // Наук. пр. Нац. б-ки України ім. В. І. Вернадського. – К. : НБУВ, 2001. – Вип. 6. – С. 190–200.

6. *Танатар Н. В.* Обсуждение всеукраинского референдума по народной инициативе в прессе / Н. В. Танатар, Н. О. Моршна // Финансовые риски. – 2000. – № 1. – С. 11–16.

7. *Танатар Н. В.* Сучасні інформаційні технології прес-моніторингу передвиборних кампаній / Н. В. Танатар, А. Г. Федорчук // Українська періодика: історія і сучасність: шоста Всеукр. наук.-теорет. конф. 11–13 трав. 2000 р. : доп. та повідомл. – Л., 2000. – С. 342–344.

8. *Танатар Н. В.* Усе – про політичну еліту / Н. В. Танатар, А. Г. Федорчук // Вісн. НАН України. – 1997. – № 1–2. – С. 91–92.

9. *Федорчук А. Г.* Використання інтернет-ресурсів для контент-моніторингу передвиборних кампаній / А. Г. Федорчук, Н. В. Танатар // Наук. пр. Нац. б-ки України ім. В. І. Вернадського. – К. : НБУВ, 2008. – Вип. 21. – С. 227–236.

10. *Федорчук А. Г.* Президентські вибори 1999 року: контент-аналіз матеріалів преси / А. Г. Федорчук, Н. В. Танатар // Українська періодика: історія і сучасність: шоста Всеукр. наук.-теорет. конф. 11–13 трав. 2000 р. : доп. та повідомл. – Л., 2000. – С. 351–354.

11. *Федорчук А. Г.* Роль преси в передвиборних перегонах-2006: контент-моніторинг на базі газетних фондів НБУВ / А. Г. Федорчук, Н. В. Танатар // Наук. пр. Нац. б-ки України ім. В. І. Вернадського. – К. : НБУВ, 2007. – Вип. 18. – С. 156–168.

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.