

Володимир КАСАТКІН,
стажист-дослідник СІАЗ НБУВ

ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ СУЧАСНОЇ БІБЛІОТЕКИ: ШЛЯХИ СТАНОВЛЕННЯ

У роботі розглядається сучасний стан розвитку інформаційно-аналітичної діяльності, процесів збору та аналізу інформації різних джерел та сучасне програмне забезпечення бібліотеки в контексті попиту на аналітичні дані.

Ключові слова: інформаційно-аналітична діяльність, аналітика, програмне забезпечення, системи моніторингу ЗМІ, повнотекстові бази даних, ІРБІС 64.

Одним з основних напрямів інформатизації, на який сьогодні спираються держструктури, бізнес, громадські організації та інші суспільні установи, всі категорії сучасних користувачів, – є інформаційно-аналітична діяльність, необхідна для прийняття ефективних управлінських рішень. Повна, точна і достовірна інформація про споживчий ринок, фірми-партнери, конкурентів, що випускається в галузі продукції, товарних ринків, ринку цінних паперів, інвестиційні можливості тощо, необхідна будь-якому підприємству для успішної діяльності на ринку, удосконалення ринкових відносин, на яких базується сучасне суспільство [2].

Наприкінці ХХ ст. аналітична діяльність набула широкого поширення і перетворилася у професійну діяльність. У багатьох країнах існують так звані «Фабрики думки», інформаційно-аналітичні відділи та служби в державних органах, компаніях, банках, політичних партіях. Загальною відмітною рисою даних служб є органічна інтегрованість у відповідні сфери діяльності, функціональний і організаційний симбіоз із їхніми соціальними інститутами і конкретними організаціями.

Стрімко розвиваються ринки аналітичної інформації, інтелектуального продукту, а також методичного та програмного забезпечення для його отримання.

Основою роботи в області інформаційного аналізу є глибока переробка матеріалів ЗМІ і фактографічних даних. Використання систем моніторингу дозволяє з мінімальними витратами оперативно отримувати інформацію з відкритих джерел, необхідну для відстеження активності конкурентів, оцінки публікацій про компанію, пошуку нових клієнтів та шляхів розвитку бізнесу, аналізу стану зовнішнього середовища організації.

Системи моніторингу ЗМІ в першу чергу призначені відбирати та сортувати матеріали, спрощуючи тим самим роботу аналітика. Таких програм на ринку близько сотні, але всі вони займаються, як правило, організацією зберігання середніх або великих обсягів інформації, володіють простими пошуковими можливостями та/або тематичним рубрикаторм, не пропонуючи будь-якого механізму якісного аналізу. І це не випадково, бо якісний аналіз тексту припускає оцінку таких нюансів, як емоції, загрози, характер відносин між об'єктами, тощо. У цих сферах людський мозок здатний накопичувати досвід і володіє інтуїцією краще. Хоча швидкість оцінки при цьому й невелика, ми виграємо в якості.

При вирішенні інтелектуальних завдань комп'ютерна програма покликає максимально полегшити працю людини: по-перше, забезпечити обмеженою вибіркою документів, відсіявши за встановленим фільтрам зайве, по-друге, надати зручне автоматизоване робоче місце (АРМ) для проведення оцінок. Деякі з програм уміють робити «інтелектуальний» пошук масивом даних, відбираючи документи з набору ключових слів, частоті їх вживання і відносному розташуванню, і роблять це з урахуванням морфології. Ці програми вже здатні на просто контент-аналіз, тобто на семантичну статистичну обробку [4].

Клас таких аналітичних (за термінологією їх авторів, хоча не всі вони повною мірою відповідають цій назві) програм російських та українських розробників має гарні шанси саме в країнах колишнього СРСР. Там, де доводиться мати справу з україномовними та російськомовними текстами та документами, у вітчизняних комп'ютерних фірм немає гідних конкурентів. Серед них виробники правових баз даних і систем розпізнавання текстів, пошукових систем і автоматичних класифікаторів, словників і перекладачів на іноземні мови.

У російсько-українському сегменті мережі Інтернет можна знайти не менше 10 баз даних, орієнтованих на збір і зберігання матеріалів ЗМІ. Найбільш відомі з них: «Артефакт» компанії «Інтегрум-Техно» (www.integrum.ru); система «Парк», створена «Парк.Ру» (is.park.ru); електронний кіоск Russian Story (www.russianstory.ru); Національна електронна бібліотека від НСН (nel.nns.ru); «Медіалогія» (www.medialogia.ru); УІС РОСІЯ (www.cir.ru); каталог ЗМІ від «СМІ.Ру» (www.smi.ru); InfoStream (infostream.ua); «Мониторикс» (monitorix.biz). Робота з усіма перерахованими джерелами, окрім «СМІ.РУ», платна, і всі їхні власники пропонують не тільки ознайомитися з матеріалами ЗМІ, а й надати деякі моніторингові послуги. А от компанія WPS (www.wps.ru) має свій електронний архів газет, журналів, транскриптів теле- і радіопередач в Інтернеті не тримає, використовуючи Internet як рекламний майданчик та інструмент для просування власних моніторингових продуктів.

Бібліотеки є важливим джерелом інформатизації, продукування необхідної суспільству інформаційно-аналітичної продукції.

У наш час як в багатьох організаціях, приватних фірмах так і бібліотеках накопичується величезна кількість інформації в різноманітному форматі та з різних джерел, і дедалі більшого поширення набувають повнотекстові бази даних. Великі освітні центри організують в Інтернеті для студентів і співробітників бази наукових статей, авторефератів, багато організацій надають доступ до ресурсів електронних бібліотек, оргкомітети конференцій публікують тисячі повних текстів доповідей і ін.

Важливого значення набуває завдання створення повнотекстових баз даних на основі текстової інформації, накопиченої в електронному вигляді з метою спрощення їх подальшої фільтрації та аналізу. У загальнодоступних бібліотеках такий масив текстів документів звичайно виникає як результат сканування текстів за приватним замовленням користувачів. Із зростанням обсягів накопиченої інформації постає завдання смислового пошуку та експертного статистичного аналізу даних з метою надати користувачеві можливість правильно орієнтуватися в середовищі електронних фондів великого обсягу, а експерту – можливість виділяти підкласи текстів на задану тематику.

Нині у світі існують і активно розвиваються системи смислового пошуку в повнотекстових базах даних, які підтримуються провідними фірмами – виробниками серверів баз даних, наприклад, Oracle, Microsoft, IBM та ін. Такі системи будуються на основі багатовимірних сховищ, з яких дані витягуються і обробляються за допомогою алгоритмів для заздалегідь визначених суб'єкт-об'єктних відносин між ними. Великі пошукові сервери в Інтернеті (наприклад, Yahoo, Yandex) підтримують алгоритми пошуку текстів «схожих» з даним і розрахунку релевантності знайдених документів вихідного запиту. Спеціалізовані системи повнотекстового аналізу (наприклад, у Росії це «Слідопит», «Текстаналіст») дають змогу проводити автоматичну класифікацію та реферування текстів [3].

Для бібліотечної системи важливе визначення апаратного програмного забезпечення, що на пряму залежить від цілей, які ставить перед собою бібліотека.

На сьогодні існує велика кількість програмних продуктів, що призначені для автоматизації бібліотечних процесів. Серед західних систем для автоматизації можна виділити такі як: DYNIX, INNOPAC, GEAC, які наразі є досить коштовними навіть при купівлі базових пакетів послуг. З іншого боку, дуже цікавим набором функцій володіє система автоматизації бібліотек ІРБІС, розроблена в ДПНТБ Росії на базі пакета CDS/ISIS, розробленого ЮНЕСКО спеціально для бібліотек, музеїв, документальних архівів тощо. Вона застосовується в сотнях бібліотек Росії і країн СНД, серед яких бібліотеки різного профілю та рівня – від

сільських до національних і від шкільних до вузівських, і підтримує роботу з форматами RUSMARC, UNIMARC і USMARC.

Розробка четвертого покоління системи – **ІРБІС 64** – покликана значно розширити функціональні можливості, як-то:

- Забезпечення справжньої багатомовності на основі використання UNICODE.

- Підтримка повнотекстових баз даних з можливістю автоматичної класифікації і смислового аналізу текстів.

- Використання як сервісного засобу орфографічного контролю текстових даних.

Важливою функцією саме для інформаційного аналізу, безперечно, є підтримка повнотекстових баз даних та можливість автоматичної класифікації текстів, а також їх смислового аналізу, що дає змогу значно скоротити затрати часу на базову обробку текстів і пошук необхідної інформації. Основна ідея автоматизованої системи смислового аналізу текстів базується на тенденції будь-якої інформації об'єднуватися в тематичні групи «за інтересами». Це явище відображає той факт, що різні галузі знань мають свою внутрішню, властиву їм мову (термінологію).

У багатьох наукових роботах обґрунтована можливість використання математичної статистики для виділення природно-тематичних груп з однорідної маси текстового матеріалу. Коротко підходи до вирішення завдання смислової обробки текстів можуть бути сформульовані у вигляді таких етапів:

- створення повнотекстової бази даних з масиву текстів;

- природно-тематична класифікація текстів на основі виділення значущих термінів предметної галузі за ознакою перевищення загальноживованої частоти;

- надання інформації про розподіл слів у текстах у числовому вигляді та структуризація цієї інформації за допомогою алгоритмів вирішення слабоформалізованих завдань.

У результаті використання системи ІРБІС 64 повнотекстові електронні ресурси бібліотеки (або іншого певного масиву накопичених текстів у електронному форматі) стають частиною пошукового простору, доступного читачеві бібліотеки або ж аналітику, перед яким стоїть завдання вибору інформації і написання аналітичного матеріалу певної тематики.

Система ІРБІС 64 складається з таких основних компонентів:

АРМ «Адміністратор» – робоче місце фахівця, що виконує системні операції над базами даних у цілому, спрямовані на підтримку їх в актуальному і збереженому стані;

АРМ «Каталогізатор» – робоче місце бібліотечного працівника, що виконує функції з формування (поповнення й коригування) Електронного каталогу й інших БД системи. Також може використовуватися

спеціалістом з відбору інформації для наповнення бази даних за певною тематикою;

АРМ «Комплектатор» – робоче місце бібліотечного працівника, що виконує функції з комплектування та обліку фондів бібліотеки на основі ведення спеціальної бази даних;

АРМ «Читач» – робоче місце з кінцевим користувачем електронного каталогу, призначене для всеосяжного пошуку в електронному каталозі, перегляду/друку знайденої інформації та формування замовлення на видачу знайденої літератури;

АРМ «Книговидача» – робоче місце бібліотечного працівника, що виконує функції з видачі та повернення літератури;

АРМ «Книгозабезпеченість» – модуль для вирішення завдання книгозабезпеченості для вузівських бібліотек.

Крім пошуку за словами з текстів, користувач отримує додаткову можливість знаходити тексти, близькі до заданого зразку. Як текст зразка може виступати запит користувача на мові оригіналу, довільний повний текст (зовнішній по відношенню до бази даних) або текст з бази даних. Крім кінцевого користувача (читача, аналітика) система орієнтована на користувача-експерта, у завдання якого входить попередня робота зі створення повнотекстової бази даних і природно-тематична класифікація текстів. Таким експертом може виступати як спеціаліст з первинного відбору інформації так і інженер з настройки системи. Крім того, система дозволяє експертам і фахівцям-класифікаторам працювати з повнотекстовими колекціями бібліотеки в режимі інтелектуального смислового аналізу текстів. Можливо також застосування системи для проведення смислового статистичного аналізу електронного каталогу бібліотеки з метою отримання і формалізації прихованого знання [6].

Для практичної реалізації повнотекстових баз даних і смислової обробки текстів у рамках ІРБІС 64 введено такі доопрацювання програмного забезпечення:

- АРМ «Читач для повнотекстових баз даних» (для відображення результатів пошуку складає HTML-сторінку за допомогою набору форматів);

- АРМ «Адміністратор» включено підсистему створення повнотекстових баз даних, індексації та класифікації повних текстів на основі алгоритмів системи смислового аналізу текстів [7].

Дана система важлива для підвищення ефективності інформаційно-аналітичної діяльності бібліотечних установ. Її значення для відповідних спеціальних підрозділів бібліотек в можливості оптимізувати управління базами використовуваної інформації, удосконалення структур оперативних інформаційних ресурсів. Це важливо з огляду на те, що такого роду підрозділи, як наприклад СІАЗ НБУ імені В. І. Вернадського, серед своїх замовників мають управлінські

та інші структури для обслуговування яких важливим є факт часу, організація замовленого матеріалу в стислі строки.

Список використаних джерел

1. *Бродовський О. І.* Нове покоління системи автоматизації бібліотек ІРБИС – ІРБИС64: від електронного каталогу до повнотекстових баз даних [Електронний ресурс]. / О. І. Бродовський, К. О. Сбойчаков. – Режим доступу: <http://www.nbu.gov.ua/Articles/crimea/2004/doc/179.pdf>. – Назва з екрана.

2. Информационно-аналитическая деятельность [Электронный ресурс]. – Режим доступа: [http://www.scip.org.ua/2009/04/23/информационно-аналитическая-деятельность-information-analytical-activity\(04/23/2009\)](http://www.scip.org.ua/2009/04/23/информационно-аналитическая-деятельность-information-analytical-activity(04/23/2009)). – Загл. с экрана.

3. *Сбойчаков К. О.* Автоматизированная система смысловой обработки текстов при создании электронных фондов библиотеки: автореф. [Электронный ресурс]. / К. О. Сбойчаков. – Режим доступа: [http://www.zipsites.ru/slovari_enc/dissertacii_referaty/sboichakov_avtomatizirovannaya_sistema_obrabotki_tekstov\(2003\)](http://www.zipsites.ru/slovari_enc/dissertacii_referaty/sboichakov_avtomatizirovannaya_sistema_obrabotki_tekstov(2003)). – Загл. с экрана.

4. *Опарин А.* Системы мониторинга и анализа СМИ [Электронный ресурс]. / Опарин А. – Режим доступа: [http://www.pcweek.ru/themes/detail.php?ID=66333\(16.12.2003\)](http://www.pcweek.ru/themes/detail.php?ID=66333(16.12.2003)). – Загл. с экрана.

5. *Сляднева Н. А.* Информационная аналитика – эзотерическое искусство или современная профессия? [Электронный ресурс]. / Н. А. Сляднева. – Режим доступа: <http://www.fact.ru/www/arhiv7s6.htm>. – Загл. с экрана. – (Дата доступа 25.07.2003).

6. *Сбойчаков К. О.* Распределение ключевых слов по рубрикам ГРНТИ в базе данных электронного каталога ГПНТБ России // Библиотека и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: тр. конф. – М., 2004.

7. Сайт документации для системы автоматизации библиотек ИРБИС. [Электронный ресурс]. – Режим доступа: http://www.wiki.elnit.org/index.php/Викидокументация_по_системе_автоматизации_библиотек_ИРБИС. – Загл. с экрана.