

К.Ю. Острівська, І.В. Стovпченко, А.І. Андрєєв

**РОЗРОБКА СИСТЕМИ ІНТЕГРАЦІЇ
ІНТЕРНЕТ-КОНТЕНТУ ДЛЯ ДОСЛІДЖЕННЯ
КОНКУРЕНТНОГО СЕРЕДОВИЩА РИНКУ ПРАЦІ**

Анотація. Сформульовано критерії оцінювання систем інформаційного моніторингу, до яких належать як критерії функціональної частини, так і критерії оцінювання інтерфейсу. Спроектовано та розроблено повнофункціональну систему інформаційного моніторингу середовища ринку праці за обраною технологією, яка пропонується як вбудовуваний модуль до сайту.

Ключові слова: інформаційний моніторинг, інтерфейс, Internet, ринок праці, модуль, контент.

У сучасних умовах в бізнесі істотне значення набуває аналіз можливих ситуацій, визначення їх тенденцій, оцінка можливих ризиків при прийнятті рішень. Недостатньо володіти інформацією у вигляді маркетингових досліджень або бізнес-довідок про підприємства потенційних партнерів і конкурентів. Важливість вирішення цих завдань стала причиною виникнення інформаційного моніторингу — інструменту інформаційно-аналітичної підтримки прийняття рішень.

Сучасний ринок — це найвища ступінь ризику і жорстка конкуренція. Для того, щоб вижити і успішно функціонувати, потрібно оперативно реагувати на події і прораховувати свої дії на кілька кроків вперед, для чого потрібна інформація. Ефективним засобом отримання такої інформації є інформаційний моніторинг, який дозволяє вчасно отримувати потрібну інформацію, аналізувати її та приймати рішення.

У свою чергу для здійснення інформаційного моніторингу необхідно є система інформаційного моніторингу, яка використовує технології інтеграції Інтернет-контенту. Тож для вибору оптимальної системи необхідним є проведення комплексного аналізу усіх технологій інтеграції та обрання доцільнішої.

Концепції та методи аналізу

Основи процедури контент-аналізу було позроблено Гарольдом Ласвеллом. На сучасному етапі існує багато думок щодо почерговості етапів аналізу контенту та ступеня їх важливості.

Виділяють такі стадії аналізу:

1. Підготовка програми аналізу документів (задання, гіпотези, поняття, об'єкт аналізу і тощо).

На цьому етапі, як правило, формулюється так звана емпірична теорія дослідження. Тобто, в ході підготовки до проведення аналізу, систематизуються гіпотези, існуючі в контексті даної проблематики та відкидаються ті з них, які не піддаються верифікації на даних інформаційного масиву.

2. Відбір джерел аналізу.

Необхідно визначити коло джерел, які містять у собі матеріали по заданій темі. Далі важливо встановити додаткові умови відбору матеріалу: визначити тип джерела (телебачення, преса, рекламні матеріали, радіо та тощо). Потім потрібно визначити вид повідомлення (публіцистичні статті в електронному або в друкованому вигляді, інформаційні замітки, рекламні плакати) роль агента комунікації (відправник або одержувач повідомлення). Визначаються мінімальні та максимальні граници об'єму тексту, їх протяжності, частота, час, місце і засіб трансляції повідомлень цільовій аудиторії. Існують і інші критерії відбору повідомлень, їх кількість і вибір варіюється залежно від поставлених завдань дослідження. Можна виділити такі проміжні етапи:

2.1. Визначення кола і обсягу документів, що є носіями необхідної інформації (найменування, періодичність виходу, період, тиражі);

2.2. Побудова вибірки: які документи та за якими критеріями будуть аналізуватись;

2.3. Перевірка побудови вибіркової сукупності;

3. Визначення емпіричних моделей аналізу, проведення вибірки (тобто підбір комунікаційних органів, вибір матеріалів за ті чи інші періоди часу, визначення видів повідомлень, типу вибірки);

У разі обмеженої кількості матеріалу по заданій темі, вибіркова сукупність може бути еквівалентна генеральній. Класичне трактування методу аналізу контенту передбачає можливість

скорочення вибіркової сукупності повідомлень при їх схожості й однорідності. Це допустимо, якщо обсяг генеральної сукупності дуже великий. Вибірка при дослідженні великих сукупностей даних випадкова. Безумовно, необхідно розрахувати її обсяг так, щоб вона залишалася репрезентативною, важливо визначити допустиму похибку вибірки. Варто додати, що часто обсяг вибіркової сукупності визначається дослідниками, виходячи з понять здорового глузду, доступності матеріалу, терміновості дослідження, а не розрахунком допустимої помилки вибірки та репрезантивності масиву джерел.

4. Розробка методики даного конкретного аналізу.

Можна виділити такі проміжні етапи:

4.1. Класифікація соціальних ситуацій відповідно досліджувальних проблем;

4.2. Визначення одиниць аналізу;

4.3. Перевірка надійності методики.

5. Пілотажне дослідження, перевірка надійності методики.

6. Збір первинної емпіричної інформації.

7. Кількісна обробка зібраних даних.

8. Інтерпретація здобутих результатів, висновки дослідження.

Вибір категорій аналізу

Категорія аналізу контенту — це поняття, яке показує один з аспектів загального напрямку конкретного дослідження документів.

Типовий комплекс контент-моніторингу містить такі автоматизовані підсистеми (рис. 1):

- збору і обробки інформації з мережі Інтернет;
- забезпечення доступу до повнотекстових баз даних;
- аналізу та узагальнення інформації.

У відповідності своєму основному призначенню, перша підсистема забезпечує:

- збір та обробку інформації з різноманітних веб-ресурсів та її форматування;
- виявлення ключових слів та понять;
- автоматичну рубрикацію;
- виявлення змістового дублювання документів;
- вибіркове розповсюдження інформації.



Головна задача другої підсистеми — формування баз даних та забезпечення доступу до неї користувачів, а саме:

- формування оперативних та ретроспективних баз даних;
- ротація баз даних;
- генерація вихідних форм;
- персоналізація роботи користувачів, збереження їх персональних запитів та джерел, ведення статистики роботи;
- забезпечення пошуку в базах даних;
- інформаційна взаємодія з базами даних інших підсистем.

Рисунок 1 - Основні технологічні процеси системи контент-моніторингу

Підсистема аналізу та узагальнення інформації забезпечує:

- формування інформаційних портретів та дайджестів;
- виявлення тематичних сюжетів та побудову таблиць взаємозв'язків понять;
- розрахунок рейтингів понять.

Загальна схема роботи аналізатора (парсера), до якого входить і синтаксичний, і лексичний аналізатор, зображена на рис. 2.

У системах інформаційного моніторингу більш вузько направленого характеру можуть використовуватися більш спрощені способи інтеграції контенту. Для того, щоб виділити в усьому об'ємі інформації потрібні дані використовується синтаксичний аналіз або його ще називають парсингом.

Синтаксичний аналізатор або парсер — це програма або частина програми, що виконує синтаксичний аналіз.

При парсингу вихідний текст перетворюється в структуру даних, зазвичай — в дерево, яке відображає синтаксичну структуру вхідної послідовності і добре підходить для подальшої обробки.



Рисунок 2 - Схема роботи аналізатора (парсера)

Алгоритми аналізу

Інформація, отримана за допомогою регулярних виразів, сортується і записується в базу даних. На екрані користувач системи бачить таблицю з отриманими результатами, в якій присутні колонки з назвою компанії, містом компанії-роботодавця (або декілька міст, якщо компанія має філії), заробітну плату в єдиному шаблоні (національна валюта), статі, посади, віком, освітою та досвідом роботи. Інформація представлена у зручному форматі і наочно демонструє ті чи інші дані по вакансіям.

Інтерфейс надає можливість отримати два типи звітів по результатам за вказаним запитом:

Звіт у вигляді графіків та діаграм, що будується динамічним шляхом за допомогою спеціальних SQL-запитів та окремою бібліотекою Google Charts.

Звіт у вигляді Excel-файлу, в якому знаходяться оброблені дані, представлені у форматі електронних таблиць. Інформацію в цьому файлі зручно сортувати, фільтрувати та будувати власні звіти.

Для того, щоб побудувати графіки та діаграми, створені SQL-запити до бази даних результатів пошуку, результатом яких є правильна таблична структура, з якої формується асоціативний

массив засобами мови PHP. Асоціативний масив із цією структурою приводиться спеціальними функціями до коректного формату (правильність кодування, округлення дробових чисел) і передається бібліотеці Google Charts, що написана на мові JavaScript. Функції цієї бібліотеки отримують правильну структуру даних з асоціативного масиву і згідно цим даних будують певні графіки, діаграми та гістограми.

Для створення Excel-файлу з обробленими даними був написаний окремий скрипт на мові PHP, який отримує за допомогою SQL-запитів масив з інформацією та ітеративними методами достроково записує цю інформацію у файл. Після створення файлу система пропонує зберегти його на локальному комп'ютері.

Навігаційна схема системи інформаційного моніторингу ринку праці зображена на рис. 3.

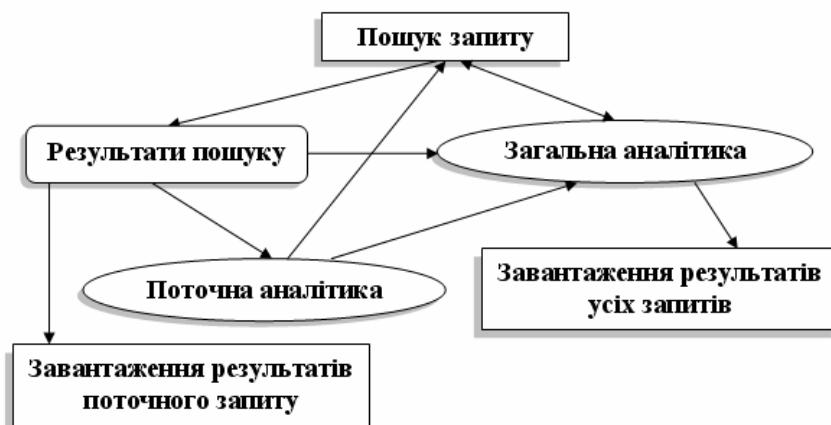


Рисунок 3 - Навігаційна схема системи інформаційного моніторингу ринку праці

Висновки

1. Проаналізовано стан інформаційного моніторингу на сьогоднішній день, сфери його використання та його завдання. Його поділяють за сферами застосування, засобами, що використовуються для збору вихідної інформації, потенційними користувачами, засобами вимірювань, способами поширення інформації, часом реалізації та широтою охоплення.

2. Проаналізовано та порівняно існуючі системи інформаційного моніторингу в мережі Інтернет, які є оптимальним рішенням завдання орієнтації в інформації з Інтернет в даний час є інформаційні мережеві служби нового типу. Але більшість із них платні та доволі дорогі, а ті, що безкоштовні — не досить

функціональні та усі вони направлені на моніторинг в основному новин та статей.

3. Сформульовано критерії оцінювання систем інформаційного моніторингу, до яких належать як критерії функціональної частини, так і критерії оцінювання інтерфейсу.

4. Спроектовано та розроблено повнофункціональну систему інформаційного моніторингу середовища ринку праці за обраною технологією, яка пропонується як вбудовуваний модуль до сайту.

ЛІТЕРАТУРА

1. Наталія Костенко, Валерій Іванов. Досвід контент-аналізу: моделі та практики: Монографія. — К.:Центр вільної преси, 2003. — 62 с.
2. Ахо А. Теория синтаксического анализа, перевода и компиляции. Т. I. Синтаксический анализ. / А. Ахо, Дж. Ульман. — М.: Мир, 1978. — 650 с.
3. Ландэ Д.В. Поиск знаний в Internet. Профессиональная работа / Д.В. Ландэ — М.: "Вильямс", 2005. — 272 с.
4. Мещеряков С.В. Эффективные технологии создания информационных систем. / С.В. Мещеряков, В.М. Иванов — СПб.: Политехника. — 2005. — 309 с.
5. Смит П. Коммуникации стратегического маркетинга / П.Смит, К. Берри. — М.: Юнити-Дана, 2001. — 282 с.