

УДК 004.774.6CMS:001.82

Олександр Кузнєцов,
наук. співроб. НБУВ

ВИЯВЛЕННЯ КЛЮЧОВИХ ПОНЯТЬ З ТЕКСТУ ЗА ДОПОМОГОЮ КОНТЕНТ-АНАЛІЗУ

Створено спектр визначень дослідних понять з текстів за допомогою контент-аналізу у вигляді графіка коефіцієнта тематичного напрямку (КТН) з подальшим аналізом.

Ключові слова: контент-аналіз, спектр визначень, тематичний напрям, кількісний аналіз, дослідження змісту.

Постановка завдання. Багатоцільність контент-аналізу вражає: кількісний і якісний аналіз текстів, визначення тематичного спрямування, оптимізація пошукових запитів, аналіз контенту сайтів для успішного їх просування, визначення плагіату, різні соціологічні дослідження, аналіз засобів масової інформації. Контент-аналіз може визначити досить на автора книги або статті (його профорієнтація, психологічний портрет, стиль письма) і зворотню задачу: написання книг або статей в стилі потрібного автора, якщо слід за когось написати.

Наведемо лише деякі визначення контент-аналізу з [1; 2].

Контент-аналіз – це методика об'єктивного якісного і систематичного вивчення змісту засобів комунікації (Д. Джері, Дж. Джері).

Контент-аналіз – це систематичне числове оброблення, оцінка та інтерпретація форми та змісту інформаційного джерела (Д. Мангейм, Р. Річ).

Контент-аналіз – це якісно-кількісний метод вивчення документів, який характеризується об'єктивністю висновків і суворістю процедури та полягає в квантифікаційному обробленні тексту з подальшою інтерпретацією результатів (В. Іванов).

Контент-аналіз полягає у знаходженні в тексті певних змістовних понять (одиниць аналізу), виявленні частоти їх вживаності та співвідношення з вмістом всього документа (Б. Краснов).

Контент-аналіз – це дослідницька техніка для отримання результатів шляхом аналізу змісту тексту про стан і властивості соціальної дійсності (Є. Таршиш).

Одним з важливих завдань контент-аналізу є виявлення понять з тексту. Тому метою нашої статті є визначення спектру понять на прикладі поняття «контент-аналіз».

Однозначне трактування понять необхідне перш за все в технічних системах. Поняття ж контент-аналізу, що має корені в психології та соціології, на сьогодні поки що не має однозначного визначення. Це породжує низку проблем, найважливіша з яких полягає в тому, що програмні системи, побудовані на основі різних підходів до контент-аналізу, – несумісні [1].

Кожне визначення контент-аналізу відображає напрямок його дослідження, завдання і є можливістю виявити сенс, загальні та унікальні ключові слова, що властиві конкретному методу дослідження, джерелам та досліджуванним об'єктам.

Як видно з перерахованих вище прикладів, визначення складають невеликі обсяги тексту з закінченою думкою [4].

Правомірність використання контент-аналізу в різних соціальних дослідженнях визначається тим, що текст аналізованого документа є продуктом діяльності людини. Тому він відображає вплив різноманітних соціальних і психологічних факторів. Звідси виникає можливість виявлення і вимірювання цих факторів шляхом реєстрації в тексті їх відповідних індикаторів або референтів. Таким чином, метою контент-аналізу є досягнення більш глибокої позатекстової реальності, тобто на основі такого аналізу ми робимо висновки про реальність історичної події або явища [5, с. 19].

Практично всі основні визначення поняття «контент-аналіз» надано у працях В. В. Осіна [1, с. 574–596] і В. М. Зверева [2, с. 119–141]. На думку англійського вченого кінця XIX ст. У. Кельвіна: «якщо ви в змозі виміряти і виразити, про що ви говорите в числах, то ви дещо про це знаєте, але якщо ви не можете виміряти це і виразити в числах, Ваші знання мізерні і незадовільні» [6]. Існує і більш загальне висловлювання: «в кожній науці стільки науки, скільки в ній математики».

Для дослідження автоматично знайдено близько ста визначень [1; 2]. За визначення вважається абзац або речення обсягом більше ста знаків, де згадується про контент-аналіз, зі структурою «контент-аналіз – це ...».

Кожне окреме визначення контент-аналізу порівнюється з іншими за частотним складом з підрахунком коефіцієнта тематичного спрямування (КТН) за допомогою програми «Аналіз текстів». При цьому поповнюється словник стоп-слів і словник лексимів до тих пір, поки кількість слів масиву унікальних слів кожного файлу не перевищує п'яти.

Показання коефіцієнта тематичного спрямування та номер визначення поняття «контент-аналізу» заносимо в таблицю (табл. 1):

Таблиця 1

№ визначення	КТН
1	0,593182727546343
2	0,578939268964588
3	0,56777890672908

101	0,709023059726348

Будуємо графік зміни КТН, упорядкувавши другий стовпчик табл. 1 – КТН за зменшенням.

Помітивши лінійний характер кривої зміни показника КТН, додамо лінію тренда лінійного характеру з рівнянням.

Отже, якщо в текстах досліджується яке-небудь поняття, то графік КТН за зменшенням буде виглядати як пряма лінія (в ідеалі). Чим крутіше нахил прямої, тим ширше поняття. Наявність на графіку КТН спотворень типу «сходинка» свідчить про надмірність кількості знайдених в тексті визначень: визначені поняття з однаковими значеннями КТН можуть бути виключені з аналізу поняття.

Так, замість чотирьох значень КТН доцільно залишити одне, наприклад: А – (0,86168).

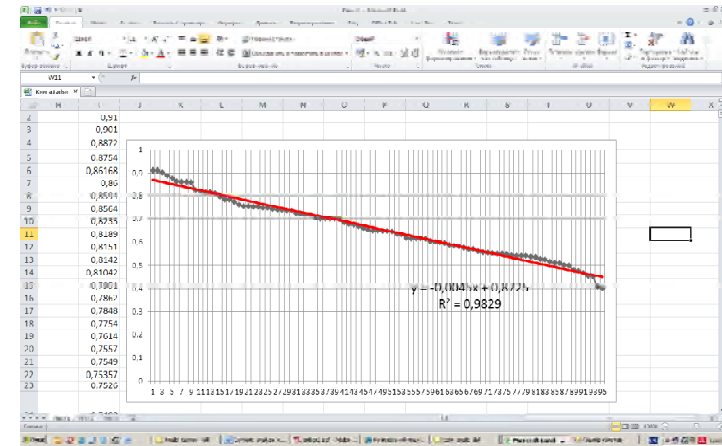


Рис. 1. Загальний графік розподілу КТН

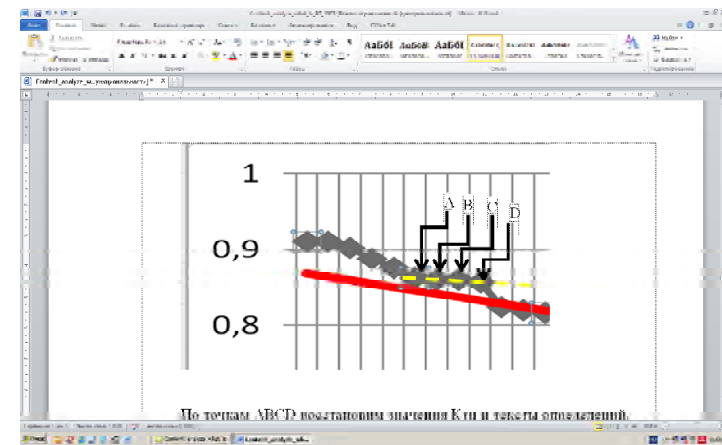


Рис. 2. Спотворення типу «сходинка»

Розглянемо ділянку значень КТН (0,8–0,9) – фрагмент графіка КТН на рис. 3:

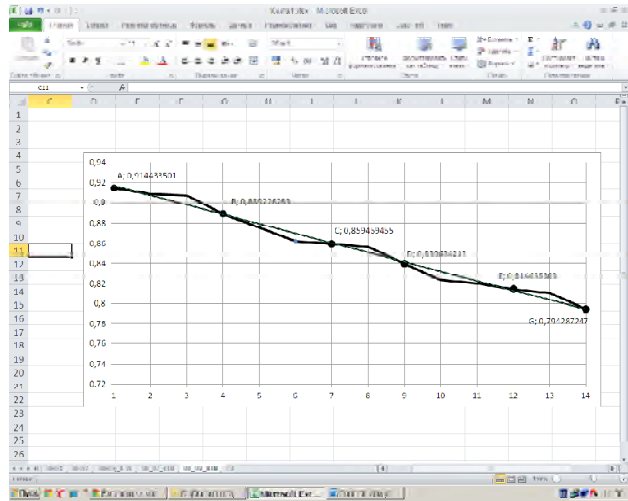


Рис. 3. Розподіл КТН на ділянці 0,8–0,9

Точки перетину кривої розподілу КТН з лінією тренда лінійного розподілу дозволяють знайти найбільш змістовні визначення поняття «контент-аналіз» для подальшого аналізу. Дані занесемо в таблицю (табл. 2):

Таблиця 2

Точка/файл	Значення КТН	Визначення
A 019.txt	0,914433500749649	В том, что следует, я предполагаю, что некоторые аспекты этнографического исследовательского подхода могут быть применены к контент-анализу с тем, чтобы дать этнографический контент-анализ (ЭА), который может быть определен как <i>рефлексивный анализ документов</i>
B 044.txt	0,889226283409789	При проведении эксперимента и опроса люди, выступающие в роли объекта исследования, знают об этом. В данной главе представлены методы, которые позволяют устранить недостатки исследований такого типа. Первый из них - это не собственно метод, а свободное собрание <i>измерений</i> . Они связаны с контент-анализом, возможности которого еще не выявлены во всей полноте. " Контент-анализ - это метод сбора данных и анализа содержания текста... С целью получения количественного описания символического содержания текста исследователь использует в контент-анализе объективный подсчет и систематически фиксирует получаемые данные. На самом деле, по мнению Маркоффа, Шапиро и Вейтмана, контент-анализ лучше было бы назвать " <i>текстуальным кодированием</i> ". Существуют качественная и интерпретативная версия контент-анализа, но все же главное в нем - <i>количественные данные о содержании текста</i>
C 036.txt	0,859459454714177	Очевидно, что в контент-анализе нет никакой необходимости, разве что вопрос, на который хотят получить ответ, является количественным

Продовження табл. 2

Точка/файл	Значення КТН	Визначення
D 097.txt	0,83963421325756	контент-аналіз - це опис й інтерпретація значень спостережуваної соціальної поведінки". Виходячи з того, як мало контент-аналітичних досліджень в Україні, бракує знань у галузі контент-аналізу й українським дослідникам
E 096.txt	0,814635363464398	контент-аналіз є одним з класичних методів <i>вивчення документів</i> і лише, що "стосовно текстів можна робити лише чи головним чином якісний аналіз". Зрозуміло, що таке твердження суперечить дійсності, його розбиває весь історичний досвід контент-аналізу. Не всі, навіть американські (тобто представники батьківщини контент-аналізу), дослідники розібралися у тому, що таке контент-аналіз
G 089.txt	0,79428724678285	"Частина обмежує об'єкт контент-аналізу суто журналістськими текстами, причому завдання дослідження фактично збігається з класичним визначенням комунікаційного ланцюжка Г. Лассвелла. Наприклад, деякі московські дослідники вважають, що <i>кількісно-якісний аналіз</i> (контент-аналіз)... може переслідувати різні завдання: наскільки адекватно відображено у них (журналістських матеріалах. - В. І.) зміст дійсності; якою мірою реалізуються завдання видавця...

Проаналізуємо вибрані визначення: оскільки ділянка 0,8–0,9 обрана ліворуч – високі значення коефіцієнта тематичного спрямування, то й отримаємо найбільш загальне визначення досліджуваного поняття «контент-аналіз»:

$$\text{Контент-аналіз} = \left\{ \begin{array}{l} \text{рефлексивний аналіз документів} \\ \text{вільне зібрання вимірювань} \\ \text{кількісні дані про зміст тексту} \\ \text{інтерпретація значень} \\ \text{кількісно-якісний аналіз} \end{array} \right\} = \text{Кількісний аналіз}$$

Отриманий результат очевидний. Однак він отриманий шляхом обчислення, а не міркування.

Якщо взяти значення з центру графіка КТН (значення 0,6–0,7), отримаємо визначення:

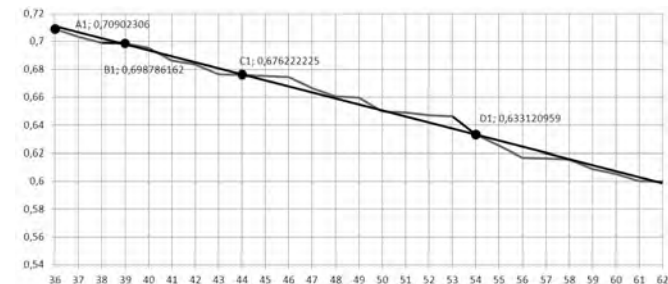


Рис. 4. Розподіл КТН на ділянці 0,6–0,7

Проаналізуємо вибрані визначення: A1, B1, C1, D1, які обрані у центрі кривої розподілу КТН (рис. 4), значення коефіцієнта тематичного спрямування (0,6–0,7) і отримаємо визначення досліджуваного поняття «контент-аналіз» (табл. 3):

Таблиця 3

Точка/файл	Значення КТН	Визначення
A1 101.txt	0,709023059726348	Контент - аналіз - дослідження як наявного, так і прихованого змісту комунікацій. Звичайно, в останньому випадку потрібні особливі процедури для підвищення достовірності й валідності висновків дослідження. Співвідношення якісних і кількісних начал у контент-аналізі дозволяє досягти великого ступеня надійності й валідності
B1 098.txt	0,698786162352785	контент-аналіз - це майже єдиний метод, у якому переважають квантифікаційні процедури. Він вигідно відрізняється від інших методів своєю об'єктивністю й систематичністю
C1 050.txt	0,659889688190607	Данные, полученные из архивных записей и документов, могут быть систематически проанализированы при помощи метода контент-анализа. Можно анализировать содержание писем, дневников, газетных статей, протоколов собраний и тому подобное. Контент-анализ является как методом анализа данных, так и методом наблюдения. Вместо непосредственного наблюдения за поведением людей или вопросов о нем, исследователь берет коммуникации, произведенные людьми, и задает вопросы коммуникации. Содержание коммуникации служит основой для выводов
D1 028.txt	0,633120959270858	Контент-анализ является систематическим методом исследования сообщения или содержания печатных медиа для получения выводов о коммуникационной системе

Як і в попередньому випадку, зробимо аналіз змісту вибраних визначень:

Контент-аналіз = { дослідження змісту
квантифікаційні процедури
методаналізу даних
метод дослідження
дослідження змісту } = Дослідження змісту

Отже, контент-аналіз – це дослідження змісту з метою виявлення явних або неявних аспектів змісту, розподілу за категоріями, отримання висновків тощо.

Вищевикладена методика дозволяє виявити й інші аспекти поняття «контент-аналіз», вибравши ділянку правіше або задати інтервал значень КТН. Слід враховувати, що чим менше КТН, тим специфічніші властивості досліджуваного поняття.

Лінійний характер КТН може бути використаний як критерій надмірності кількості дослідних інформаційних одиниць: наявність на графіку КТН спотворень типу «сходінка».

Висновки.

Отже, ми визначили поняття «контент-аналіз» для різних предметних областей. Поняття контент-аналізу залежить від цілей дослідника. Для більш загальних випадків (великі значення КТН 0,8–0,9) – це кількісний аналіз для загальної систематизації або загальних висновків. Для випадків з середніми значеннями КТН (0,6–0,7) – це дослідження змісту для специфічних цілей. Ця методика інваріантна для будь-яких понять.

Список використаних джерел

- Осин В. В. Процессы конституирования и воспроизводства в науке : исследование вариаций контент-анализа : монография / В. В. Осин. – Д. : Лири ЛТД, 2007. – 636 с.
- Зверев В. М. Социологическое прочтение философских идей России XVIII века : контент-анализ. – СПб : Изд-во Рус. христиан. гуманист. ин-та, 1998. – 159 с.
- Ландэ Д., Литвин А. «Феномены современных информационных потоков» / [Электронный ресурс]. – Режим доступа : <http://www.visti.net/~dwl/art/content/>, Информационный центр «ЭЛВИСТИ».
- Кузнєцов О. Ю. Технологія інформометричних досліджень матеріалів конференцій (на прикладі конференцій «Крим-2005» і «Крим-2010») / [Електронний ресурс] // Міжнародна наук. конф. «Формування і розвиток бібліотечного електронного середовища» : матеріали конф. (Київ, Україна 3–6 жовтня 2011 року). – Режим доступу : archive.nbuv.gov.ua/new/11_Kiev/kuznetsov_o.ppt
- Богомолова Н. Н. Контент-анализ / Н. Н. Богомолова. – М. : Изд-во Моск. ун-та, 1992. – 60 с.
- Миронов Б. Н. История в цифрах: математика в исторических исследованиях / Б. Н. Миронов. – Ленинград : Наука, 1991. – 168 с.

Стаття надійшла до редакції 25.04.2014 р.

UDC 004.774.6CMS:001.82

Oleksandr Kuznetsov

Content Analysis as a Method of Identify of Key Concepts

This article discusses qualitative and quantitative content analysis of text. Through graphs and case studies, the author puts forth examples of content analysis used for search query optimization, website analysis, identifying plagiarism, analysis of the media, delivering content to the right audience, and other uses.

Keywords: content analysis, the range of values, thematic areas, quantitative analysis, research content.

УДК 004.774.6CMS:001.82

Александр Кузнецов

Выявление ключевых понятий из текста с помощью контент-анализа

Создано спектр определенных исследовательских понятий из текстов с помощью контент-анализа в виде графика коэффициента тематического направления (КТН) с дальнейшим анализом.

Ключевые слова: контент-анализ, спектр определений, тематическое направление, количественный анализ, исследование содержания.

УДК 025.5:005.642.2

Ольга Василенко,

канд. іст. наук,

директор Інституту науково-методичного забезпечення
бібліотечно-інформаційної роботи НБУВ

БІБЛІОТЕЧНО-ІНФОРМАЦІЙНА ДІЯЛЬНІСТЬ ЯК ОБ'ЄКТ СТАТИСТИЧНОГО ДОСЛІДЖЕННЯ

Запропоновано формування системи статистичних показників бібліотечно-інформаційної діяльності на основі комплексного, збалансованого та інтегрального підходів. Розглянуто доцільність використання інтегральних показників у бібліотечній статистиці, термінологічні та методичні аспекти їх визначення; обґрунтовано необхідність подальшого розвитку теоретико-методологічних, нормативно-методичних і практичних засад розвитку вітчизняної бібліотечної статистики; представлено проект форми статистичного звіту.

Ключові слова: бібліотечна статистика, формування системи статистичних показників бібліотечно-інформаційної діяльності, інтегральні показники бібліотечної статистики, статистична бібліотечна звітність.

Сучасні інформаційно-комунікаційні технології не просто модифікували діяльність наукових бібліотек, але й поставили перед ними принципово нові завдання, пов'язані з інтеграцією до вітчизняного та світового інформаційного простору, консолідації зусиль з іншими інформаційними, освітніми та культурно-просвітницькими установами щодо задоволення потреб суспільства в інформації та знаннях. В умовах проголошеного розвитку інформаційного суспільства в Україні та формування оновленої інформаційної парадигми, де місце і роль бібліотек визначатиметься їхньою здатністю до системної трансформації та конкурентоспроможністю на ринку інформаційних продуктів і послуг якісно нового значення для всієї системи бібліотечного управління набуває статистична інформація. Саме ці відомості, що формуються в процесі статистичного спостереження, узагальнення та аналітико-синтетичного опрацювання і виражені у статистичних показниках, створюють основу для оцінки роботи бібліотек як суспільнозначущих інституцій. Важливим складником отримання статистичної інформації є методологічна та