

Б. М. Юськів

КОНТЕНТ-АНАЛІЗ
Історія розвитку
і світовий досвід

Рівне
"Перспектива"
2006

ББК 60.5
Ю - 91

*Розповсюдження і тиражування
без офіційного дозволу видавництва заборонено.
Усі права застережені.*

Рецензенти:

д-р політ. наук, проф. **Є.Б.Тихомирова**
(завідувач кафедри міжнародної інформації Рівненського
інституту слов'янознавства Київського славістичного
університету);
д-р істор. наук, проф. **С.С.Троян**
(завідувач кафедри політичних наук Рівненського державного
гуманітарного університету)

**Рекомендовано до друку вченою радою
Рівненського інституту слов'янознавства
Київського славістичного університету
(протокол № 9 від 30 травня 2006 р.)**

Юськів Б. М.

Контент-аналіз. Історія розвитку і світовий досвід: Монографія. –
Рівне.: "Перспектива", 2006. – 203 с.

ISBN 966-7643-77-8

Монографія присвячена проблемам розвитку одного з найпоширеніших методів аналізу масових комунікацій – контент-аналізу. Розглядаються етапи розвитку контент-аналізу, дається характеристика застосування його на кожному етапі, описуються особливості методики та напрями вдосконалення. Особлива увага приділяється комп'ютерному контент-аналізу, який поступово перетворює контент-аналіз з наукового методу в сучасну технологію, яка знаходить повсюдне масове застосування. Однією з технологій, яка має в основі контент-аналіз, є Text Mining. Про її можливості та застосування також ведеться мова в роботі.

Дослідження може прислужитися викладачам, науковцям, політикам, аспірантам, студентам, усім, хто цікавиться проблемами і методами аналізу текстів.

ISBN 966-7643-77-8

© Юськів Б. М., 2006

ЗМІСТ

ВСТУП	5
Розділ 1. ЗАРОДЖЕННЯ І ЛОГІКА РОЗВИТКУ КЛАСИЧНОГО КОНТЕНТ-АНАЛІЗУ	7
1.1. Зародження контент-аналізу	7
1.2. Формування основ класичного контент-аналізу	11
1.3. Контент-аналіз пропаганди під час Другої світової війни і в післявоєнний час	18
1.4. Нові напрями розвитку класичного контент-аналізу	26
Розділ 2. КОНТЕНТ-АНАЛІЗ У 1950-1980-х рр.	34
2.1. Передумови розвитку і використання контент-аналізу на початку 1950 рр.	34
2.2. Практика контент-аналітичних досліджень пропаганди	36
2.3. Диференціація і розширення сфери застосування контент-аналізу	42
2.4. Удосконалення методики контент-аналізу і розвиток якісних досліджень	53
Розділ 3. ТЕНДЕНЦІЇ РОЗВИТКУ КОНТЕНТ-АНАЛІЗУ В ПЕРІОД З 1990 ДО 2000 рр.	69
3.1. Статистика і основні тенденції розвитку контент- аналізу	69
3.2. Основні цілі і напрями використання контент-аналізу... ..	74
3.3. Дискусії довкола переваг і недоліків контент-аналізу.... ..	81
3.4. Особливості методики сучасного кількісного контент- аналізу та її вдосконалення	87
3.5. "Якісний" розвиток контент-аналізу	102
Розділ 4. КОМП'ЮТЕРИЗОВАНИЙ КОНТЕНТ-АНАЛІЗ	114
4.1. Перші спроби механізувати і автоматизувати контент- аналіз	114
4.2. The General Inquirer і особливості другого покоління програм контент-аналізу	119
4.3. Третє покоління програм контент-аналізу	131
4.4. Сучасні технології контент-аналізу і особливості Text Mining	139
ВИСНОВКИ	155
БІБЛІОГРАФІЯ	157

ДОДАТОК А. Основні поняття і алгоритм кількісного контент-аналізу	170
ДОДАТОК Б. Програми контент-аналізу.....	186
ДОДАТОК В. Програми Text Mining.....	195

ВСТУП

На ХХ ст. припадає формування такого потужного феномена, як масова інформація. Саме він, по суті, став основою запровадження нового інформаційного світопорядку і, урешті-решт, призвів до виникнення в масштабах усього світу інформаційного суспільства. Тому зрозуміло, що жодна наука не лише не змогла обійти своєю увагою масову інформацію, але й змушена була ретельно її вивчати. Потреба об'єктивного аналізу змісту інформаційних потоків дала безпосередній поштовх для появи різного роду методів і методик дослідження змісту інформаційних матеріалів.

Одним із методів дослідження масових інформаційних потоків став контент-аналіз, який зародився та розвивався разом із зростанням феномена масової інформації. До кінця ХХ ст. контент-аналіз як методика дослідження документальних комунікацій не лише вийшов з наукових келій, редакцій журналів і газет, а й став одним із загальнодоступних базових аналітичних інструментів, який завдяки комп'ютерній техніці та Інтернету нині потужно прогресує.

Об'єктом даного дослідження став контент-аналіз, а безпосереднім *предметом* вивчення – історія зародження і тенденції його розвитку. *Досліджуваний період* охоплює часи від зародження ідей контент-аналізу до початку 2000 рр.

Як і будь-яка наукова методика, контент-аналіз склався не одразу, а розвивався поступово. Зарубіжні дослідники, зокрема Т.Ф.Карней¹, О.Олсті², К.Кріпендорф³, К.Е.Розенгрєн⁴,

¹ Carney T.F. Content analysis: Techniques for systematic inference from communications. – Winnipeg: University of Manitoba Press, 1972.

² Holsti O.R. Content Analysis for the Social Sciences and Humanities. – MA: Addison-Wesley, 1969. – <http://www.questia.com/PM.qst?a=o&d=54363997>

³ Krippendorff K. Content analysis: An introduction to its methodology / Vol. 5. The Sage Communication Text Series. – Beverly Hills: Sage Publications, 1980.

⁴ Rosengren K.E. Advances in Content Analysis / Vol. 9, Sage annual reviews of communication research. – Beverly Hills: Sage Publications, 1981.

Ф.Мейрінг⁵ виділяють три стадії розвитку контент-аналізу: газетний, класичний і сучасний аналіз. Однак від недавнього часу в методиці контент-аналізу сталися істотні зміни. З огляду на це, є зміст сформувати більш сучасне розуміння етапності розвитку контент-аналізу, а також дати конкретну характеристику кожного етапу. Це і стало *завданням* нашого дослідження.

Однак перед тим, як перейти до викладу основного матеріалу дослідження, варто зробити два зауваження. Перше стосується джерельної бази. У процесі роботи автор намагався користуватися винятково закордонними інформаційними матеріалами. На жаль, через відсутність доступу до багатьох першоджерел довелося послуговуватися вторинними матеріалами, як правило, з Інтернет. Щодо вітчизняного наукового доробку, то автор використовував його лише в тому випадку, коли він містив огляд або аналіз закордонних досліджень. Такий підхід зумовлений винятково намаганням подивитися на контент-аналіз з позиції тих, хто формував його основи і використовував у процесі теоретичних і практичних досліджень.

Друге зауваження стосується можливих закидів щодо певної упередженості до здобутків вітчизняних дослідників. Насправді це абсолютно не відповідає дійсності, оскільки автор вважає, що досвід наших науковців і практиків є по-своєму унікальним і вартий окремого дослідження. Особливо це стосується використання контент-аналізу. Тому таке дослідження – ще попереду.

⁵ Mayring P. Qualitative Content Analysis // Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-line Journal], 2000, June, 1(2). – <http://qualitative-research.net/fqs/fqs-e/2-00inhalt-e.htm>

Розділ 1. ЗАРОДЖЕННЯ І ЛОГІКА РОЗВИТКУ КЛАСИЧНОГО КОНТЕНТ- АНАЛІЗУ

Для більш глибокого розуміння самої методики сучасного контент-аналізу і виявлення тенденцій його розвитку актуальним є вивчення процесів зародження і формування основ класичного контент-аналізу.

1.1. Зародження контент-аналізу

Аналоги найсучасніших методів можна знайти в анналах історії. Не є винятком і контент-аналіз.

Історія наукового пізнання комунікацій почалася ще в Древній Греції та Древньому Римі. Античні мислителі поряд із поняттям "логос-розум" поважали "логос-мову". До певної міри це було зумовлено й тим, що використання письмових текстів лише входило в моду, тоді як ораторське мистецтво було саме тим засобом, через який не лише здійснювалося політичне життя греків, а й розвивалися та розповсюджувалися філософські ідеї. Оратори, які володіли силою усного слова, користувалися особливою довірою народних зборів. Тому риторика розвивається, як важливе мистецтво зі своїми специфічними формами, засобами і стратегіями аргументування до аудиторії. Саме школи риторів і софістів зробили перші кроки до пізнання правил інтерпретації текстів.

Мабуть, однією з перших спроб "моделювання комунікативних процесів" можна вважати опис Арістотелем (384–322 рр. до н.е.) у його "Риториці" структури публічного виступу: мова складається з трьох елементів: самого автора, предмета, про який він говорить, та особи, до якої звертається. Інший видатний представник античності, державний і політичний діяч Древнього Риму, оратор Цицерон (106–43 рр. до н.е.) у своїх роботах з риторики особливу увагу приділяв урахуванню психології, інтересів і смаків публіки. Зрозуміло,

що аналіз, який робили глядачі, слухаючи публічні виступи ораторів, навіть близько не був схожим на контент-аналіз, тим не менше це були перші спроби оцінити в певний спосіб як зміст, так і передачу мови виступаючих.

У Середньовіччя християнська церква не забула уроків античного красномовства, і в програму підготовки служителів церкви було включено риторику, граматику і діалектику, які утворювали "тривіум" – три перших і основних предмети вивчення. Середньовічні схоласти відродили древньогрецьку герменевтику – науку про розуміння і тлумачення поетичних (Гомер) і священних (Біблія) текстів.

На нашу думку, ще одним попередником контент-аналізу можна вважати систему бального оцінювання в школах, яка дозволяла чисельно оцінити рівень знань учнів. Зокрема, уже в єзуїтських школах XVI-XVII ст. учні розподілялися за розрядами, позначеними цифрами. Одиниця мала значення вищої оцінки. Підвищуючи свій розряд, учень здобував цілий ряд привілеїв.

У XVIII ст. в Німеччині, що була тоді "інкубатором" педагогічних інновацій і впливала на розвиток педагогічної думки в Європі, виникла 12-бальна шкала оцінок. Основна роль у її становленні належить, імовірно, Йогану-Бернарду Базедову. У 1774 р. він відкрив у Дессау показову виховну установу *Philanthropinum*. У школі застосовувалася спеціальна система заохочення вихованців: на особливій дошці проти прізвища учнів ставили крапки, за кількістю яких визначалися успіхи і відповідні пільги; вихованця, що одержав визначену кількість крапок, нагороджували знаком відмінника або ласощами⁶.

Ранній контент-аналіз розпочинався, як систематизований підхід до вивчення текстів. Це були ранні інтерпретації біблій, ранній газетний аналіз, аналіз почерків, аналіз мрій.

Багато закордонних дослідників⁷ одностайно погоджуються, що першим задокументованим випадком

⁶ Арапов М.В. Пятибальная шкала оценок: стоит ли менять? – <http://www.courier.com.ru/top/content/score.htm>

⁷ Carney T.F. Content analysis: Techniques for systematic inference from communications...; Holsti O.R. Content Analysis for the Social Sciences and Humanities...; Krippendorff K. Content analysis: An introduction to its

контент-аналізу був проведений у Швеції в 1640 р. аналіз змісту апокрифічного видання 90 церковних гімнів "Пісні Сіону". Їхнє розповсюдження пов'язувалося з поширеними тенденціями недоброзичливого ставлення населення до офіційного лютеранства. Гімни хоч і пройшли державну цензуру і набули великої популярності, однак були звинувачені у невідповідності релігійним канонам. Наявність або відсутність такої відповідності визначалися шляхом підрахунку в текстах гімнів релігійних символів і порівняння їх з іншими релігійними текстами, зокрема гімнами офіційної церкви та забороненими церквою текстами "моравських братів". Ураховувалася також і якість подачі символів – позитивна, негативна, нейтральна, складність стилю і контекст, у якому з'являлися ідеї. Було визнано, що образи Христа, а також його символіка як люблячого спасителя з'являлася частіше за інші християнські цінності ортодоксального лютеранства.

Безпосередньо сам контент-аналіз зародився з потреб масових комунікацій і розпочався як кількісно-орієнтований метод аналізу преси. Наприкінці ХІХ ст. – початку ХХ ст. у США з'явилися перші контент-аналітичні дослідження текстів газет, виробництво яких перетворилося в масове. Мотивація авторів досліджень виглядає на диво знайомою: вони задавалися цілком продемонструвати сумне "пожовтіння" тодішньої нью-йоркської преси. Спостерігався інтерес до громадської думки і таких етичних проблем, як відповідальність тих, хто повідомляє новини. Розроблені аналітичні методи систематичного, об'єктивного та кількісного опису матеріалів преси, які вийшли з американської школи журналістики, отримали назву *кількісного газетного аналізу*⁸.

Однією з перших робіт, яка описує використання газетного аналізу, є дослідження Дж.Спіда "Чи газети сьогодні дають

methodology...; Dovring K. Quantitative Semantics in 18-th Century Sweden // Public Opinion Quarterly. – 1954-1955. – Vol.18. No.4.

⁸ Gallagher C, Marken M., Kim M., Phillipson T., Dodge T. The Use of Content Analysis in Researching Adolescent Issues. – Indiana University, 2000. – 17 p. – www.secretstage.com/other/web/messages/school/y520/monograph.pdf

новини?"⁹, опубліковане в журналі "Форум" за 1893 р. Дослідження мало на меті зафіксувати зміни тематичної структури недільних випусків нью-йоркської преси в період з 1881 по 1893 рр. Саме в цей час одна з найбільших американських газет "Нью-Йорк Таймс" різко підняла свій тираж, знизивши при цьому ціну номера з трьох до двох центів та змінивши на більший газетний формат. Інші видання, як виявилось, також пішли тим же шляхом. Постало питання: що ж дало змогу це зробити? Дж.Спід прокласифікував зміст матеріалів за темами (література, політика, релігія, плітки, скандали тощо) і виміряв довжину газетних колонок (у дюймах) для кожної теми. Порівнявши дані за два роки, він дійшов висновку: після проведеної "реформи" газети стали надавати більше місця повідомленням про чутки і скандали за рахунок зменшення кількості й обсягу матеріалів з проблем літератури, мистецтва, політики, релігії тощо¹⁰.

Дослідженням функцій газет займався і Д.У.Уїлкоккс, який у 1900 р. в книзі "Американська газета у світлі соціальної психології"¹¹ проаналізував зміст номерів 240 газет за один день, використавши таку класифікацію матеріалів:

1) новини, які поділені на воєнні та цивільні; останні, у свою чергу, діляться на міжнародні, політичні, новини про злочини та людські недоліки, інші новини;

2) ілюстрації;

3) література;

4) думки (редакційні статті, листи);

5) реклама.

Фахівці в галузі американської журналістики, які займалися аналізом змісту, звертали увагу не тільки на розробку різних класифікацій, але й на загальні теоретичні положення

⁹ Speed G. Do Newspapers Now Give the News? // The Forum. – 1893. – Vol. XY. – P.705-711.

¹⁰ Методы сбора информации в социологических исследованиях. Кн.2. Организационно-методические проблемы опроса. Анализ документов. Наблюдение. Эксперимент / Отв. ред. В.Г.Андреенков, О.М.Маслова. – М.: Наука, 1990. – С.68.

¹¹ Wilcox D. The American Newspaper. A Study in Social Psychology // Annals of the Academy of Political and Social Science. 1900. Vol. XYI.

процедури аналізу. 1912 р. було сформульоване положення про необхідність вивчати ступінь уваги преси до різних тем. Аналіз цієї характеристики преси став традиційним і активно використовується в сучасних дослідженнях масової комунікації¹².

Праці американських науковців стали відомими і знайшли продовження в Європі. Уже в 1910 р. на першому засіданні Німецького Соціологічного товариства Макс Вебер закликав використовувати контент-аналіз для оцінки охоплення пресою політичних акцій в Німеччині та вивчення громадської думки¹³. Конкретною метою його мало стати вивчення ефектів впливу преси на тогочасну людину та культурні цінності, причому пропонувалося робити це шляхом безпосереднього застосування до вивчення газет таких базових вимірювальних і вибіркового інструментів, як "вирізання й осмислення"¹⁴. На жаль, цьому честолюбному проєкту так і не судилося бути реалізованим.

Контент-аналіз почав використовуватися і для аналізу інших видів документів. Зокрема, ранні використання контент-аналізу були спрямовані на перевірку правдивості історичних документів. Дослідники займалися головним чином підрахунком слів у документах сумнівного походження і порівнювали їхні частоти і приклади використання з відповідними словами в документах, чия вірогідність була безсумнівною¹⁵.

1.2. Формування основ класичного контент-аналізу

На межі XIX–XX ст. мали місце як об'єктивні, так і суб'єктивні причини для становлення контент-аналізу як наукової методології. До об'єктивних причин відносять:

¹² Методы сбора информации в социологических исследованиях... – С.69.

¹³ Content Analysis: An Introduction. –

http://www.coe.uga.edu/~cwise/content_analysis/index.html

¹⁴ Англійський варіант фрази звучить так: with the basic measuring and sampling tools of "scissors and compasses" on "the newspapers themselves"

¹⁵ Department of Journalism Studies, The University of Sheffield, 2002. –

<http://doj.shef.ac.uk/ResearchMethods-Lecture%20Eight.doc>

– стрімке поширення засобів масової комунікації (зміст інформації став розглядатися як об'єкт спеціального дослідження); розмаїття вироблених ЗМІ текстів як нове соціальне явище вимагало вивчення;

– тенденція до використання методів природничих наук у науках соціальних;

– необхідність вивчення текстів ЗМІ з метою виявлення ефектів, які вони справляють на людей (на тлі розвитку пропагандистської діяльності у зв'язку з революціями і Першою світовою війною)¹⁶.

Суб'єктивна необхідність у такого роду методі виникала з уявлень соціологів-теоретиків і практиків журналістики тієї пори про роль і можливості засобів масової інформації. У перших теоріях масової комунікації преса розглядалася як величезний потенціал впливу на маси, а людська маса – абсолютно беззахисною перед рупором пропагандистської машини. Роль установок індивіда, вплив групи, лідера думок у комунікативному процесі з'явилися в теорії пізніше. У цьому випадку на порядку денному актуальним було дослідження впливу пропаганди на масову свідомість, тобто з'ясування через процедуру аналізу змісту політичної поведінки комунікатора або передбачуваного ефекту в аудиторії.

Зауважимо, що в умовах Першої світової війни контент-аналіз був єдиною можливістю вивчати в широких масштабах пропаганду противника, моральний дух населення його країни, події, що відбуваються в ній, оскільки масова інформація була доступною і за межами цієї країни, особливо з розвитком радіо. "Така можливість і донині залишається великою перевагою цього методу в арсеналі способів вивчення інших країн, відгороджених різного роду бар'єрами"¹⁷.

Дослідження преси попередніх років в основному зводилися до підрахунку площі, яку займали ті чи інші теми на

¹⁶ Костюк В.В. Анализ документов и текстов: Лекции по курсу / Белорусский Государственный Университет, 2003. –

http://www.ffsn.bsu.by/infocom/kostuk/document_analysis/lectures.html

¹⁷ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации. – М.: Институт социологии РАН, 2001. – С.17.

сторінках газет, або частоти вживання в літературі різних періодів певних слів, частин мови тощо. Контент-аналіз як науковий метод, розроблений в США в 1920-1930 рр., уже мав принципово іншу основу. Це зокрема відзначає Т.Ф.Карней¹⁸. К.Кріпендорф у своїх висновках іде далі, він убачає істотну різницю між кількісним газетним аналізом, який вимірює пресу в категоріях предмета досліджень, і тим, що він називає "*ранній*" контент-аналіз, який став новою стадією інтелектуального розвитку контент-аналізу¹⁹.

Так, у 1926 р. М.Уїллі в дослідженні "Провінційна газета"²⁰ вже не зупиняється на тематичному дослідженні преси. Насамперед він використав класифікацію газетних повідомлень за трьома параметрами:

- темою або тією сферою дійсності, яку ці матеріали відображають (наприклад, політика, економіка, література);
- формою (новини, передові статті тощо);
- їхньою комбінацією (наприклад, політичні новини).

Ці групування, незалежно від способу їхнього утворення, одержали в американській журналістиці та соціології назву "категорій". М.Уїллі також сформулював критерії, за якими виділяються сукупності категорій. Перший критерій полягає в тому, що система категорій має застосовуватися до всіх газет загального характеру, а не орієнтуватися лише на спеціальні видання. Другий критерій: система категорій повинна будуватися таким чином, щоб дати можливість порівнювати номери різних газет за той самий день, одні й ті ж газети в різний час або різні газети в різний час. Ця вимога універсальної порівнянності означала необхідність високої стандартизації категорій, що дозволяло використовувати статистичні методи для аналізу змісту. Третє правило: система категорій повинна давати максимум точності і зводити елемент суб'єктивності до мінімуму. Виходячи з таких посилок, М.Уїллі розробив систему класифікації, що нараховувала 49 категорій. Серед них –

¹⁸ Carney T.F. Content analysis: Techniques for systematic inference from communications...

¹⁹ Krippendorff K. Content analysis: An introduction to its methodology...

²⁰ Методы сбора информации в социологических исследованиях... – С.69.

внутрішні політичні і зовнішньополітичні новини, дипломатична кореспонденція, міжнародні відносини, ціни тощо. І, нарешті, у гіпотетичному плані М.Уїллі розглядає твердження про те, що чим більше в газетах інформації з різним ступенем локальності, тим ширшою є аудиторія читачів цих газет.

Дослідження М.Уїллі базувалося на таких теоретичних посилках, які могли стверджуватися лише з гіпотетичною долею впевненості. Однак вони стали дуже плідними для пізніших досліджень змісту ЗМК, що будували свої програми, а значить, і принципи розгляду тексту, на тих або інших залежностях між ЗМК і аудиторією²¹.

У дослідженні С.Кінгсбері, Х.Харта і А.Кларка²² спектр новин у газеті розкладається за ступенем соціального в тому інтересі читача, до якого ці новини апелюють. З цієї точки зору всі новини поділяються на три групи:

– новини, що стосуються чисто споживацьких інтересів читача (наприклад, повідомлення про погоду);

– новини, що торкаються читача як члена певної соціальної групи, більш широкої спільноти, нації і т.д. (такими є повідомлення про вибори міської влади, про військово-морську конференцію, а також про нове тарифне законодавство тощо);

– сенсаційні новини²³.

Автори ставили за ціль, зваживши частку кожної групи новин у газеті, визначити соціальну цінність газети.

Дослідження новин Дж.Вудворда²⁴ в американських ранкових газетах мало аналогічну ціль: за обсягом тематичних пластів оцінити суспільну думку, що формується газетою. Це можливе тому, що, за твердженням науковця, преса впливає на зміни в суспільній думці радше за допомогою перекручувань,

²¹ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации... – С.14.

²² Kingsbury S., Hart H., A. Klark. News and the Newspaper // The New Republic. – 1930. – Vol. 64. – October 8. – P.15-22.

²³ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации... – С.15.

²⁴ Woodward J. Foreign News in American Morning Newspapers (A Study in Public Opinion). – N.Y., L, 1930.

пропусків, концентрації або зафарбовування фактів, аніж прямими редакційними проповідями. Саме в такий спосіб новини, що ллються на аудиторію зі шпальт щоденних газет, формують уявлення особистості про світ і події в ньому, а ці стереотипізовані уявлення, у свою чергу, стають базою, яка визначає наступне ставлення людини до цього світу. Таким чином, поділ усїєї газетної площі на визначені тематичні пласти, з'ясування частки, що припадає на кожен пласт, дозволили автору зробити відповідний висновок щодо газетної політики – прокоментувати переваги одних тематичних шарів на шкоду іншим²⁵.

Дослідження, подібні до наведених, поступово сформували підґрунтя, на базі якого починають розвиватися теоретичні засади *класичного контент-аналізу*. Фундатором школи класичного контент-аналізу вважають американського дослідника засобів масових комунікацій Гарольда Лассуелла, який у 1920–1930-і рр. не лише використовував методику кількісного аналізу для вивчення змісту воєнних і політичних пропагандистських матеріалів, але і дав йому теоретичне обґрунтування. Основну мету вивчення текстів пропаганди Г.Лассуелл сформулював так: визначити, що пропагандист ставить у центрі уваги для досягнення очікуваного ефекту в аудиторії. Лассуеллівська школа контент-аналізу базується на ідеях Г.Лассуелла про знаки-стимули і відповіді-реакції на них. Систематичне дослідження – підрахунок і аналіз – знаків-стимулів, репрезентованих у тексті словом, судженням або фрагментом, складає, на думку Лассуелла, сутність даного методу.

У 1927 р. Г.Лассуелл видав свою знамениту книгу "Техніка пропаганди у світовій війні"²⁶. Автор поставив собі завдання проаналізувати, якими соціальними моделями поведінки маніпулює пропаганда воюючих країн, які цілі ставить вона у воєнний час. Аналізу піддалися багато можливих каналів пропаганди США, Англії, Франції та Німеччини: газети,

²⁵ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации... – С.70.

²⁶ Lasswell H. Propaganda Technique in the World War. – N.Y.: Knopf, 1927.

випуски бюлетенів інформаційних агентств, пропагандистські матеріали в журналах, тексти виступів і проповідей тощо. В аналізованих матеріалах пропаганди кожної з воюючих країн автор знайшов переважання таких тверджень: "ми" захищаємося, "ворог" – підступний агресор; "ворог" зруйнував райське благоденство і тому повинен бути знищений; "ми" переможемо, "ворог" буде знищений. Кожна країна до цього додавала свої деталі: для англійської пропаганди були характерні докази гуманності; німецька пропаганда апелювала до великої німецької культури, яка потребує захисту. Після узагальнення цілі пропаганди воюючих країн виглядали так: пробудити ненависть до ворога, зміцнити дружбу із союзниками, налагодити дружні відносини з нейтральними країнами, деморалізувати супротивника²⁷.

Значний вклад у розвиток теоретичних основ контент-аналізу зробили і найближчі співробітники Г.Лассуелла – Н.Лейтес, І.Пул, І.Яніс, Р.Фаднер, Д.Лернер. У рамках їхніх досліджень розроблялися методичні проблеми: вибір одиниці контексту при частотному підрахунку символів, способи перевірки результатів на обґрунтованість і зіставність тощо.

Проведені в ті часи дослідження сьогодні вважаються класикою контент-аналізу, а самого Г.Лассуелла сучасні соціологи називають "патріархом" контент-аналізу.

З появою нових медіа – радіо та кіно – контент-аналіз намагається охопити і їх. Типовими прикладами досліджень радіо тих років можна вважати роботи Дж.Ландберга (1927 р.) і В.Албіга (1937 р.), у яких аналізувалися програми радіопередач. У першому випадку досліджувалися на основі щоденних газетних оголошень усі нью-йоркські радіостанції за місяць, у другому – програми 9 американських і 1 англійської станції мовлення в період з 1925 по 1935 рр. Передачі класифікувалися за характеристиками в програмах (музика, програми для жінок,

²⁷ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации... – С.19.

нариси, освітні передачі, спорт, новини, погода, церковні служби, реклама, політичні програми тощо)²⁸.

Перші спроби досліджувати кіно в термінах і з позицій аналізу змісту відносяться до середини 1930-х рр. Історики контент-аналізу вважають першою роботою, де були застосовані методи кількісного аналізу змісту кіно, дослідження Е.Дейла²⁹.

Піонерською була й робота Д.Джоунс³⁰ з дослідження продукції Голлівуду напередодні 1940-х рр. Автор з'ясовувала економічний, соціальний і родинний стан головних персонажів в аналізованих фільмах, стать, вік, національність, їхні життєві ідеали, засоби досягнення цілей, уключення в соціальні групи. За результатами дослідження головні персонажі фільмів (їх в аналізованому масиві було 188) переважно добре забезпечені американці (на 50%), із середніх верств, їхні надії пов'язані зі спокоєм, доходом і повагою до себе. Для 68% – важливе кохання, для 26% – репутація і престиж, для 16% – безпека, для 14% – спосіб життя, для 10% – матеріальні блага, для 9% – право вести власну справу. Найістотнішим є те, що троє з кожних п'яти героїв мають у підсумку виконання своїх бажань. Саме такою була кількісна оцінка знаменитого "хепі-енд" американського кіно³¹.

Наступний приклад стосується аналізу творів літератури. Так, Б.Берельсон і П.Сальтер³² проаналізували белетристику (переважно короткі оповідання) з восьми найбільш популярних американських видань у період з 1937 по 1943 рр. Авторів цікавила національність дійових осіб, роль, яку вони відігравали у творі, і вид зображення – із симпатією, без симпатії тощо³³. За даними дослідження, у відібраних розповідях представники

²⁸ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации... – С.36.

²⁹ Dale E. Content of Motion Pictures. – N.Y., 1935.

³⁰ Jones D. Quantitative Analysis of Motion Picture Content // Public Opinion Quarterly. – 1942. – Vol. 6. No.3.

³¹ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации... – С.74.

³² Berelson B., Salter P. Majority and Minority Americans // Public Opinion Quarterly. – 1946. – Vol.10. – No.2.

³³ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации... – С.25.

національних меншин порівняно зі "справжніми" американцями значно рідше виступали як "герої", а їхні образи найчастіше носили негативне забарвлення. Вони мали низький соціальний і економічний рівень життя, дуже часто були замішані в незаконних або щонайменше темних аферах. Таким чином, у цілому представники національних меншин характеризувалися гірше, менш повноцінно, ніж американці. Чим ближчою була фігура до стереотипу американця, тим частіше (у відсотках до всіх згадувань кожної національності) вона зображалася як гарна, пристойна, шляхетна, заможна і т.д.

1.3. Контент-аналіз пропаганди під час Другої світової війни і в післявоєнний час

Друга світова війна і післявоєнні події створили своєрідні випробувальні полігони для нового методу. Хоча на той час контент-аналіз починає застосовуватися в соціологічних, історичних і політичних дослідженнях, однак аналіз змісту пропаганди складає майже чверть усіх контент-досліджень в 1940 рр.³⁴ Наведемо приклади кількох контент-досліджень.

Під час Другої світової війни мав місце, мабуть, найзнаменитіший епізод в історії контент-аналізу – прогнозування британськими аналітиками початку використання Німеччиною проти Великобританії крилатих ракет "Фау-1" і балістичних снарядів "Фау-2". Прогноз був здійснений англійським дослідником А.Джорджем³⁵ спільно з американськими фахівцями на основі аналізу внутрішніх пропагандистських кампаній у Німеччині. Підставою для висновків було різке збільшення оптимістичних виступів лідерів третього рейху щодо подальшого ходу війни, а також відомості з інших джерел, у тому числі огляди та прогнози британських і американських військових експертів. Останні дані були екстралінгвістичними (не з газетних сторінок), і тому історики

³⁴ Gallagher C, Marken M., Kim M., Phillipson T., Dodge T. The Use of Content Analysis in Researching Adolescent Issues...

³⁵ George A. Propaganda Analysis: A Study of Inferences from Nazi Propaganda in World War II. – N.Y., 1959.

контент-аналізу відмовляли цьому дослідженню в праві називатися чисто контент-аналітичним дослідженням. Однак, як зауважує Л.М.Федотова³⁶, слід віддати належне можливостям, які продемонстрував контент-аналіз змісту: висновки більш загального плану робилися на основі чисто текстової продукції.

Ще одним прикладом вдалого контент-аналізу є здійснення союзниками моніторингу музичних передач (тривалість музики і репертуар) на європейських радіостанціях. Порівнюючи мелодії, які звучали на радіостанціях у Німеччині, з радіомузикою на території окупованої фашистами Європи, союзники мали змогу з високою ймовірністю оцінювати успіхи противника, зміни в концентрації і розташуванні військ на континенті. Аналогічно чинили американці на тихоокеанському театрі бойових дій, вивчаючи розміщення японських військ на островах³⁷.

Багато прикладів використання контент-аналізу дала діяльність американської військової цензури в роки Другої світової війни. Перед експертами було поставлене завдання: абсолютно вірогідно визначити, чи не є політична лінія певної газети або інформаційної агенції профашистською, що послаблює моральний дух американського народу і посилює його супротивників. Розроблені методи виявлення ЗМІ, які проводять антинаціональну політику, визнані класичними і отримали назву "тестів Лассуелла" – за іменем керівника групи дослідників. Метою тестів було виявлення пропаганди визначеного типу в тих або інших матеріалах. Уся сукупність тестів дістала назву метод "виявлення" (detection). Метод "виявлення" включав такі тести:

- 1) пряме визнання – очевидна ідентифікація з ворожою стороною;
- 2) паралелізм – збіг змісту матеріалів досліджуваного каналу зі змістом матеріалів ворожого каналу;
- 3) зв'язність – відповідність потоку матеріалів проголошеним цілям ворожої пропаганди;

³⁶ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации... – С.80.

³⁷ Department of Journalism Studies...

4) презентація – баланс негативного і позитивного ставлення до символів обох сторін;

5) джерело – переважання якого-небудь одного джерела;

6) приховане джерело – часта публікація матеріалів без зазначення джерела;

7) своєсвідність – використання слів і зворотів, характерних для пропаганди ворожої сторони;

8) перекручування – постійні перекручування тих чи інших висловлювань на загальні і конкретні теми в напрямку, вигідному лише одній стороні³⁸.

Метод "виявлення" став яскравим прикладом виходу на характеристики комунікатора (автора чи замовника текстів) через аналіз створених ним текстів. За результатами подібного аналізу Верховний Суд США неодноразово приймав безпомилкові рішення про закриття видань або зняття з них безпідставних звинувачень. Однак, як зауважує О.Іголкин³⁹, антинаціональну діяльність могла довести лише наявність цих тверджень у сукупності, у системі. Наприклад, якщо газета писала про "любов" В.Черчілля до спиртного, то це ще зовсім не визначало її політичної орієнтації, зрештою, на цьому й не акцентували видання, які займали профашистську позицію.

Прикладом використання тесту "паралелізм" може служити наступний⁴⁰. Фахівцями різного профілю було сформульовано 14 основних положень, які сходилися з тезами німецької пропаганди і були свідомо спрямовані проти США та Великобританії:

- США внутрішньо розкладені;
- зовнішня політика США аморальна;
- президент США гідний осуду;
- Великобританія внутрішньо розкладена;
- зовнішня політика Великобританії аморальна;
- прем'єр-міністр В.Черчілля гідний осуду;

³⁸ Методы сбора информации в социологических исследованиях... – С.80.

³⁹ Иголкин А.А. Масс-медиа в условиях глобализации // Вызовы глобализации: Ответы России и Кореи. Материалы международного научного симпозиума, 5-6 июля 2001 г. – М.: РГГУ, 2001. – С. 60. – <http://liber.rsu.ru/Conf/Globalization/globalization.pdf>

⁴⁰ Иголкин А.А. Масс-медиа в условиях глобализации... – С. 60.

- фашистська Німеччина справедлива і добропорядна;
- зовнішня політика Японії морально виправдана;
- фашистська Німеччина сильна;
- Японія сильна;
- США слабкі;
- Великобританія слабка;
- союзники роз'єднані;
- існує загроза США і світу з боку комуністів, євреїв і плутократів.

Експерти інтерпретували всі тексти в підозрюваній газеті "Галілеянин", виявляючи конкретні твердження, що відповідають чи суперечать наведеним 14 положенням. За підрахунками американських фахівців, у газеті за період з 22 грудня 1941 р. по 2 березня 1942 р. було виявлено 1195 тверджень, які можна інтерпретувати як антиамериканські, і лише 45, що суперечать їм. Найчастіше в найрізноманітніших формах зустрічалися твердження, що США слабкі (317 разів) і внутрішньо розкладені (279 разів), протилежні їм пропозиції зустрічалися лише відповідно 5 і 26 разів. Такий "перекіс" Верховний Суд США визнав переконливим свідченням профашистської позиції газети і прийняв рішення про її закриття.

Ще одним прикладом на цю ж тему є результати дослідження Г.Лассуелла і Н.Лейтеса⁴¹ публікацій газети "Ширий американець" за період 3 по 31 грудня 1943 р. Окремі підсумки аналізу можна побачити в табл. 1.1. Загалом виявилось, що профашистські заяви більше, ніж в 11 разів перевищували частоту проамериканських висловлювань, і Верховний Суд США на цій підставі також прийняв рішення про закриття газети⁴².

⁴¹ Lasswell H., Leites N. Language of Politics. – New York, 1949; Lasswell H. Describing Contents of Communications // Smith B., Lasswell H., Casey R. Propaganda, Communication and Public Opinion. – Pr., 1946, – P. 75.

⁴² Федотова Л.Н. Контент-аналитические исследования средств массовой информации и пропаганды. Учебно-методическое пособие. – М.: МГУ, 1988. – С.23.

Таблиця 1.1

Результати контент-аналізу газети "Щирий американець"

Твердження	Газета (кількість випадків)	
	стверджувала	спростовувала
Держапарат США просочений корупцією	301	34
Зовнішня політика США – антихристиянська	41	0
Президент Рузвельт – небажана особа	150	17
Німеччина справедлива і мужня	56	2
Політика Японії морально виправдана	22	6
Японія – могутня країна	30	2
Німеччина – могутня	41	2
Америка – слабка	345	41
США та їхні союзники знаходяться під контролем комуністів	55	0
США та їхні союзники знаходяться під контролем євреїв	99	0
В урядовій адміністрації США багато шпигунів	45	0

Наведемо приклад тесту "зв'язаності"⁴³. Головна ідея цього тесту полягала в зіставленні того, як окремі події висвітлювалися різними виданнями, на чому робився акцент, які конкретно факти і твердження потрапляли або не потрапляли на сторінки. Саме за цими параметрами – такі, що не сходяться, а також присутні або відсутні твердження і факти – порівнювалися два видання: "еталонне" американське і підозрюване. Відтак усі положення, які не сходилися,

⁴³ Иголкин А.А. Масс-медиа в условиях глобализации... – С. 62.

піддавалися перевірці на можливу відповідність одній з основних стратегічних цілей фашистської пропаганди:

- перекласти провину за розв'язану війну на противників Німеччини і показати "миролюбство" Німеччини;
- довести облудність пропаганди противників Німеччини і "щирість" фашистів;
- показати слабкість і занепад супротивника, силу і "високий моральний дух" Німеччини;
- продемонструвати світу загрози з боку євреїв і плутократів.

При порівнянні текстів з описом однакових подій інформаційної агенції "Трансоушн" і газети-еталона "Нью-Йорк Таймс" виявилося, що серед "додаткових" фактів і тверджень у "Трансоушн" 58% сходилися з цілями фашистської пропаганди і лише 3% – суперечили їм. Водночас серед тих, які потрапили на шпальти "Нью-Йорк Таймс", але були відсутніми в "Трансоушн", 48% матеріалів суперечили пропагандистським установкам Німеччини. Таким чином, симпатії підозрюваного агентства виявилися цілком очевидними.

Тест "презентація" вивчає баланс символів, використовуваних протилежними сторонами. Цей тест гарно демонструє такий факт⁴⁴. Під час Другої світової війни група американських соціологів проаналізувала дванадцять програм різних радіостанцій США. Дослідниками було визначено близько тисячі символів, які ідентифікувалися з "об'єднаними націями" (країнами антигітлерівської коаліції) і "державами осі" (країнами фашистського блоку). Виявилося, що в усіх програмах, за винятком однієї, частота символів позитивно корелювала з символами, що представляють "держави осі". Збільшення або зменшення одних символів викликало відповідні коливання кількості символів іншої групи. Отже, у передачах була закладена єдина пропагандистська схема.

У тесті "джерело" визначається співвідношення в інформаційному потоці матеріалів, наданих кожною з двох протидіючих сторін. Скажімо⁴⁵, під час Першої світової війни,

⁴⁴ Методы сбора информации в социологических исследованиях...— С.68.

⁴⁵ Иголкин А.А.Масс-медиа в условиях глобализации...— С. 62.

ще до початку вступу в неї США, інформація про бойові дії подавалася в "Нью-Йорк Таймс" на 70% з посиланнями на джерела в країнах Антанти і лише на 5% – у Німеччині й Австро-Угорщині.

Контент-аналіз у воєнний період використовувався не лише в цензурних цілях. Зокрема, аналіз змісту повідомлень дозволяв пояснити свідомо приховані або неусвідомлені причини провалів або низьку ефективність пропагандистських кампаній, окремих матеріалів. В американській літературі з проблем контент-аналізу наводиться досить багато таких прикладів. Розглянемо один з них⁴⁶.

У США антигітлерівська пропаганда серед негритянського населення базувалася на такій ідеї: негри, хоча і страждають від расової дискримінації, усе ж чогось домоглися, живучи в США, а Гітлер, відомий своєю ненавистю до кольорових, у випадку перемоги позбавить їх усього, що вони мають. Ці теми були порушені у великій статті відомого негритянського письменника, яка повинна була стати початком усієї пропагандистської кампанії. Після проведення спеціального опитування з'ясувалося, що цей матеріал не дав очікуваного ефекту. Тому пропагандистам довелося звернутися за допомогою до фахівців з контент-аналізу. Кількісне зіставлення тематичних елементів матеріалу дало досить красномовні дані: перша тема (опис здобутків негрів завдяки американській демократії) складала 84% матеріалу, тоді як друга (лихо, яке принесе Гітлер) – лише 16%.

Після війни подібні дослідження в США продовжувалися. Так⁴⁷, наприкінці 1940-х рр. Американський єврейський конгрес звинуватив у недобррозичливому ставленні до національних меншин одну з радіостанцій і вимагав позбавити її урядової ліцензії. Спираючись на техніку контент-аналізу, було вивчено радіопередачі цієї компанії, і звинувачення зняли.

⁴⁶ Методы сбора информации в социологических исследованиях...– С.73.

⁴⁷ Иголкин А.А.Масс-медиа в условиях глобализации...– С. 63.

З нечисленних досліджень радіопередач варто згадати аналіз Л.Сасменн⁴⁸ новин 33-х радіостанцій, які стосувалися робітничого і профспілкового руху. Це дослідження було розпочате за позовом лідерів робітничого і профспілкового руху до радіостанцій, які, на думку позивачів, вирізнялися недоброзичливістю до робітників і профспілок.

У повоєнний період увага пропагандистів було повернута на Радянський Союз, а контент-аналіз став активно використовуватися для аналізу радянської преси. Одне з таких досліджень⁴⁹ мало на меті вивчення статей в газеті "Московские новости", присвячених іншим країнам. Було встановлено, що співвідношення позитивних і негативних матеріалів про закордонні країни в 1938–1939 рр. було 69 проти 328. Тоді як 1903 позитивним статтям про СРСР протистояли лише 3 несприятливих.

Відомі також численні дослідження лексики радянської преси⁵⁰.

Американські фахівці на основі вивчення газети "Московські новини" за 1938–1939 рр. за частотою появи розмістили політичні терміни в такій послідовності: народ, робітники, колгоспи і радгоспи, Сталін (сталінський), Ради, Ленін (ленінський), революція, п'ятирічний план, комісаріат, партія, стаханівський рух, Червона Армія, класова боротьба, клас, соціалістичне будівництво, задача, буржуазія, колективізм, експлуатація, пролетаріат, трудівники, соціалістична революція, соціалістичне змагання, диверсія, соціалістична праця.

У 1940 рр. за допомогою контент-аналізу з'ясувалося, що в радянській пропаганді виділяється декілька пластів, відносна значущість яких постійно змінюється. Дослідники виділили стару "ліберальну" лексику, революційну, троцькістську лексику "світової революції" та національну. Вивчалось, зокрема, співвідношення між словами – символами "світової революції" і "позареволюційною" національною символікою за

⁴⁸ Sussmann L. Labor in the Radio News an Analysis of Content // Journalism Quarterly. – 1945. – Vol. 22. No.3.

⁴⁹ Иголкин А.А. Масс-медиа в условиях глобализации... – С. 63.

⁵⁰ Иголкин А.А. Масс-медиа в условиях глобализации... – С. 63.

період з 1918 по 1943 р. Виявилося, що якщо в досліджених гаслах 1920 р. кількість символів "світової революції" більше ніж у 10 разів перевершувала число національних символів, то з 1929 р. чітко простежується зростання питомої ваги національної символіки, а з 1931 р. – неперервне падіння інтересу до "світової революції". До 1943 р. в радянській пропаганді все, пов'язане зі світовою революцією, практично зникло.

Таким чином, контент-аналіз переконливо демонструє, як початкова більшовицька ідеологія змінювалася, причому рух йшов у напрямі відкидання ультрареволюційного і зміцнення національного ядра.

Більш того, у 1940-х рр. став швидко підсилюватися "значеннєвий" пласт, який походив з дореволюційної Росії. Колишня символіка поверталася не лише в "словах", але і реальних побутових деталях (погони, "гімназійна" форма школярів тощо). Надзвичайно важливою є постійна присутність у ЗМІ (і суспільній свідомості) великої кількості імен як символів – історичних і міфологічних, починаючи від Олександра Невського, Іллі Муромця та Давида Сасунського аж до Валерія Чкалова й Олександра Матросова. Важливо підкреслити, що в довгий символічний ряд входили імена не лише росіян. Преса була насичена ними, частота їхньої появи в звичайних статтях була дуже високою. Очевидним стає факт, що відбувається процес активізації історичної пам'яті.

1.4. Нові напрями розвитку класичного контент-аналізу

У першій пол. ХХ ст., окрім чисто прагматичного використання контент-аналізу для військових потреб і пропаганди, дослідники починають усе більше уваги приділяти розвитку самої методики аналізу та виробленню нових підходів до розв'язання проблем. Показовим у цьому плані є вихід у 1952 р. книги Б.Берельсона "Контент-аналіз у комунікаційних

дослідженнях"⁵¹, яка ввібрала досвід 1920-1940 рр. і практично стала першим підручником з контент-аналізу. Саме тут Б.Берельсон наводить означення, яке багато років служило своєрідною точкою відліку для багатьох дослідників у їхніх намаганнях подати власне розуміння контент-аналізу: кількісний контент-аналіз являє собою "метод дослідження для систематичного, об'єктивного і кількісного опису очевидного змісту комунікацій"⁵².

У його основі лежить процедура сегментації змісту комунікації на окремі одиниці, прив'язка одиниць до деякої категорії і чисельне вимірювання кожної категорії. Іншими словами, мета класичного контент-аналізу: узяти вербальний (нечисловий) текст і перетворити його в кількісні дані. Під час цього первинний неопрацьований зміст перетворюється в деяку стандартизовану форму. Значно пізніше за Б.Берельсона, оцінюючи властивості класичного контент-аналізу, С.В.Робертс зауважує: "Контент-аналіз являє собою метод для перетворення символічних даних у дані матричної форми, що створює можливість для опрацювання їх з використанням статистичного аналізу"⁵³. Як можна зауважити, тут основний акцент робиться не стільки на слові "аналіз", скільки на певному способі "виміряти" (знайти деяку міру) первинні символічні дані.

Класичний контент-аналіз – це строго кількісний метод, основним і центральним інструментом якого виступає система категорій. Найпростішим видом оцінювання категорій є підрахунок числа появи категорії (припускається, що між частотою появи змісту і значенням існує взаємозв'язок)⁵⁴.

Окрім того, на думку Б.Берельсона, контент-аналіз повинен фокусуватися винятково на очевидних характеристиках тексту.

⁵¹ Berelson B. Content analysis in communication research. – Glencoe: Free Press, 1952.

⁵² Berelson B. Content analysis in communication research...– P.18.

⁵³ Roberts C.W. Content Analysis // International Encyclopedia of the Social & Behavioral Sciences / Editors-in-Chief Neil J. Smelser, Paul B. Baltes.– Amsterdam: Elsevier Science Ltd, 2001.– P.2697.

⁵⁴ Kohlbacher F. The Use of Qualitative Content Analysis in Case Study Research // Forum: Qualitative Social Research [On-line Journal].– 2005.– December, 7(1), Art. 21.– [http:// www.qualitative-research.net/fqs-texte/1-06/06-1-21-e.htm](http://www.qualitative-research.net/fqs-texte/1-06/06-1-21-e.htm)

Текстові аналітики не повинні бути інтерпретаторами, а концентрувати свою увагу на зовнішніх характеристиках змісту тексту, так званих "об'єктивних даних", наприклад, словах, ідіоматичних виразах, метафорах, які вони можуть безпосередньо ідентифікувати і підрахувати їхню абсолютну і відносну частоту згідно зі строго визначеними правилами. Будучи стурбованим саме підрахунком кількості слів або фраз, контент-аналіз лише іноді пов'язує їх з деякими іншими оцінками, наприклад, ваговими коефіцієнтами.

Таким виглядає класичний контент-аналіз та його особливості. Проте саме вони стали надто жорсткими та обтяжливими обмеженнями, і метод почав поступово виходити за ці рамки. У ньому виразно вимальовувалися два напрями: спочатку інструментальний аналіз, а з 1940-х рр. – репрезентативний.

Щодо першого, Ч.Осгуд⁵⁵, можливо, інтуїтивно, але надзвичайно слушно, зауважив: "Фактично, ми можемо визначити метод контент-аналізу як такий, що уможливорює *інструментальний аналіз*, якщо він виловлює (фіксує) елементи повідомлення, про які явно не сказано в ньому, що дозволяє відтворити справжню інформацію, незважаючи на стратегію самого джерела повідомлення" (цит. за ⁵⁶). Іншими словами, контент-аналіз використовує слова як симптоматичний інструмент, на основі якого дослідник намагається діагностувати предмет дослідження, визначаючи неусвідомлені або невідомі його характеристики.

Яскраві приклади інструментального аналізу представлені в роботі вже згаданого англійського дослідника А.Джорджа "Аналіз пропаганди"⁵⁷. У ній використовується підхід, який отримав назву не кількісного методу аналізу змісту. При цьому для розгляду відношень між елементами ланцюжка "зміст – пропагандист – керівні органи" використовувалися екстралінгвістичні фактори.

⁵⁵ Osgood Ch. The Representational Model and Relevant Research Methods // Trends in Content Analysis / I. de Sola Pool (ed.). – Urbana, Ill.: University of Illinois Press, 1959.

⁵⁶ Roberts C.W. Content Analysis... – P.2697.

⁵⁷ George A. Propaganda Analysis...

На відміну від інструментального, *репрезентативний аналіз* допомагає досліднику класифікувати, помітити або зрозуміти той зміст, який укладає в повідомлення його автор або джерело інформації (і яке може істотно відрізнятись від справжнього змісту досліджуваного об'єкта). Ця різниця між двома підходами призводить до відповідної інтерпретації тексту: або з позиції предмета дослідження, як в інструментальному аналізі, або з позиції джерела інформації. Якщо дослідник підходить до тексту інструментально, то він інтерпретується в термінах теорії, яку сформулював дослідник на основі свого розуміння сутності предмета дослідження. Якщо дослідник розуміє текст репрезентативно, то тексти ідентифікують зміст, закладений у повідомлення автором тексту. Відповідно методи інструментального текстового аналізу використовуються для ідентифікації об'єкта дослідження, наприклад, індивідуальних або соціальних характеристик членів суспільства, про які вони, можливо, навіть не підозрюють; репрезентативний аналіз використовується, щоб охарактеризувати тексти з того боку, як його розуміє джерело інформації або як хоче джерело інформації, щоб його зрозуміли ті, кому тексти призначалися⁵⁸.

Для кращого розуміння відмінності між ними можна навести такий приклад. Нехай вивчається невелике коло письменників-чоловіків, які як головні героїні своїх творів вибирають винятково старших жінок. Дослідник репрезентаційної орієнтації, насамперед, у передмовах і примітках намагається віднайти пояснення необхідності такого вибору головних героїнь, зокрема висуваючи версію відновлення в такий спосіб справедливості щодо старших жінок, оскільки більшість авторів зловживає вибором головних героїв на користь молодих і сильних чоловіків. На противагу цьому, дослідник інструментальної орієнтації може тлумачити вибір головних героїнь як симптом комплексу Едіпа в письменників. Зауважимо, що в першому випадку думки авторів безпосередньо виражені ними в текстах, тоді як у другому варіанті – докази на

⁵⁸ Roberts C.W. Content Analysis...– P.2698.

користь своєї гіпотези дослідник має не просто віднайти, а видобути з тексту.

Репрезентативний підхід у розглядуваний період насамперед пов'язують з іменем представника школи Г.Лассуелла та одного з його найближчих співробітників Абрахама Каплана, який запропонував розширити сферу використання контент-аналізу від статистичної семантики (значення текстів) політичних дискусій до аналізу значень символів – семіотики⁵⁹. Іншими словами, ним було запропоновано здійснювати не лише кількісний, але й якісний аналіз символічних значень політичної бесіди. У післявоєнний період популярність семіотики як окремої дисципліни інтенсивно зростає в напрямі якісно орієнтованого контент-аналізу для вивчення "ідеологічних" повідомлень у таких "аполітичних" жанрах, як телевізійні розваги та комерційна реклама.

Зауважимо, що класичний контент-аналіз сформувався в рамках кількісного (квантитативного) підходу на протиположності підходу якісному (квалітативному), який на початку ХХ ст. був підданий різкій критиці через його суб'єктивізм. Починаючи з кінця 1930-х рр. і особливо після Другої світової війни, "дослідники кількісного напрямку зайшли далеко вперед як у виробленні чітких доказів, так і в перекладі теоретичних концептів у дослідницький інструментарій. У результаті з'явилася можливість вимагати більш строгого підтвердження теорії. Тому переваги кількісних методів сприяли тестуванню непідтверджених теорій за допомогою фактів"⁶⁰. Більше того, переваги кількісних дослідницьких технік призвели до того, що "риторика верифікації", властива кількісним дослідженням, стала поширюватися і на якісні. Тому цілком логічно, що з позицій верифікації якісним методам відводилася лише другорядна, допоміжна роль.

⁵⁹ Kaplan A. Content Analysis and the Theory of Signs // Philosophy of Science. – 1943. – No 10. – P. 230-247.

⁶⁰ Ключкина О.Б. Построение теории на основе качественных данных (из истории развития методологического направления) // Социс. – 2000. – №10. – С.99.

Величезним поштовхом до розвитку контент-аналізу мали нові ідеї Г.Лассуелла. У 1948 р. в роботі “Структура і функції комунікації в суспільстві”⁶¹ він сформулював нову системоутворюючу конфігурацію масової комунікації, запропонувавши тепер класичне означення акта комунікації: "ХТО – повідомляє ЩО – яким КАНАЛОМ – КОМУ – з яким ЕФЕКТОМ". Відповідно до цієї структури виділяються такі основні розділи дослідження комунікацій:

- аналіз управління процесами масової комунікації;
- аналіз змісту переданих мас-медіа повідомлень (контент-аналіз);
- аналіз роботи самих мас-медіа;
- аналіз їхніх аудиторій;
- аналіз результатів ("ефекту") масово-комунікаційного впливу.

При цьому Г.Лассуелл відзначає три досить значущі й очевидні функції процесу комунікації:

- спостереження за навколишнім середовищем для виявлення загроз суспільству і визначення можливостей здійснення впливу зовні на цінності як суспільства в цілому, так і його складових частин;
- кореляція співвідношення складових частин суспільства при його реагуванні на "поведінку" навколишнього середовища;
- передача соціальної спадщини від покоління до покоління⁶².

Саме під впливом цих ідей аналітики масових комунікацій, і в тому числі контент-аналізу, перебували протягом наступних кількох десятиліть.

* * *

⁶¹ Lasswell H.D. The Structure and Function of Communication in society. The Communication of Ideas / Ed. by Bryson. – N. Y., 1948.

⁶² Терин В.П. Основные направления исследований массовой коммуникации // Социологические исследования. – 1997. – №11. – С.28.

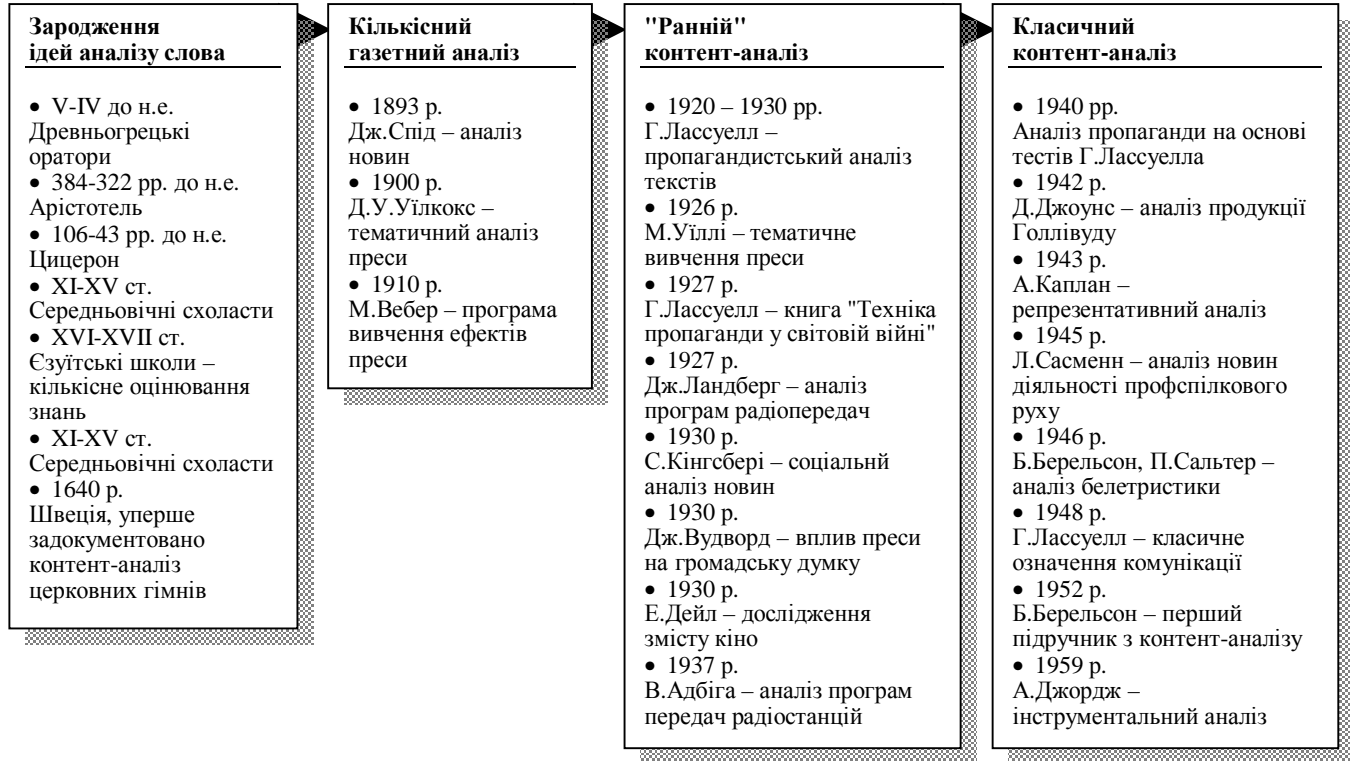


Рис.1.1. Зародження і розвиток класичного контент-аналізу

Таким чином, зароджуючись як логічне продовження раннього газетного аналізу XIX ст., класичний контент-аналіз до середини XX ст. став одним із найпотужніших методик досліджень не лише масових комунікацій, а й вивчення пропаганди у військовій і повоєнній час, проблем психології, лінгвістики, історії та інших наук (рис. 1.1).

Розділ 2. КОНТЕНТ-АНАЛІЗ У 1950-1980-х рр.

Після 1940 рр. контент-аналіз, розвинений методично і збагачений практикою досліджень, широко використовується не лише як самостійна техніка вивчення матеріалів ЗМІ. Його намагаються активно поєднувати з іншими методиками – спостереженнями, оглядами, опитуваннями, експериментами. Цей розділ має на меті відслідкувати логіку розвитку контент-аналізу в надзвичайно бурхливий для нього період – 1950-1980 рр.

2.1. Передумови розвитку і використання контент-аналізу на початку 1950 рр.

На початку 1950 рр. проведення контент-аналізу вимагало чимало матеріальних і людських ресурсів. Саме це, а також людські помилки і часові обмеження робили його непрактичним для вивчення великих текстів. Однак, попри вказані проблеми, потреба в контент-аналізі зростала.

Роботи американських учених у галузі контент-аналізу продовжували займати провідну роль у розвитку і поширенні методу. Вони стали загальноприйнятими і протягом усього періоду 1950-1980 рр. залишалися своєрідним стандартом для науковців усього світу.

Ще однією особливістю було те, що на 1950-1960 рр. припадає прихід у сферу контент-аналізу свіжих сил. Молоде покоління аналітиків активно зайнялося проблемами і намагалося побороти обмеженість класичного контент-аналізу.

На початок періоду основним об'єктом досліджень контент-аналізу й далі залишалася масова комунікація та пропаганда. Уряд США добре зрозумів серйозність використання ЗМІ. У 1960 рр. на його замовлення дослідження проводилися в найсильніших і найпрестижніших дослідницьких університетах і лабораторіях – Rand Corporation, Гарварда, Принстона, Йеля, Стенфорда, у Массачусетському технологічному інституті,

причому відкриті публікації стосувалися тільки дуже загальних та історичних проблем⁶³.

Контент-аналіз розвивався в руслі теорій масових комунікацій. Науковці, перебуваючи під впливом ідей Г.Лассуелла, намагалися скористатися практичним і теоретичним досвідом попередніх років. Модель комунікацій Г.Лассуелла стала теоретично і практично затребуваною не лише для опису пропагандистських, а й інших видів масових комунікацій – рекламних, маркетингових, політичного PR, PR-кампаній, розрахованих на масову аудиторію; вона досить широко використовувалася у практиці ЗМІ та соціології. З'явилася нова тенденція: дослідницькі центри намагаються залучити до контент-досліджень висококваліфікованих фахівців з інших наукових сфер – психолінгвістики, семіотики, соціології, політології, історії, обчислювальної техніки.

Тим часом сам Г.Лассуелл зосередив увагу на дослідженні політичних комунікацій, вважаючи, що "всі науки є політичними, оскільки вони дозволяють зрозуміти процес і механізми здійснення політики або надають конкретні теоретичні підходи й емпіричні дані, необхідні для прийняття політичних рішень. На його думку, політологія і комунікативні науки повинні використовувати не тільки методи емпіричної соціології, соціальної психології, але й психоаналізу і психіатрії у вивченні масових комунікацій, політичної комунікації, політичної поведінки та політичної пропаганди"⁶⁴.

Г.Лассуелл розглядав комунікацію як складний процес, що має свої соціальні функції, внутрішню структуру і загальну спрямованість. Він виділяє три найважливіших універсальних соціальних функції масової комунікації:

– інформаційну (пізнання навколишнього світу: "медіа розширюють обрії пізнання індивіда");

– кореляційну (корелює із соціальною структурою та "відповідальністю" суспільства, впливає на нього і його

⁶³ Иголкин А.А. Масс-медиа в условиях глобализации...– С. 60.

⁶⁴ Гнатюк О.Л. Из истории американской коммуникологии и коммуникативистики: Гарольд Лассуелл (1902-1978) // Актуальные проблемы теории коммуникации: Сборник научных трудов.– СПб: Изд-во СПб ГПУ, 2004.– С. 11-20.– http://russcomm.ru/rca_biblio/g/gnatuk.shtml

пізнання через зворотний зв'язок, пояснює та інтерпретує інформаційні повідомлення, забезпечує підтримку існуючої влади та пануючих норм);

– пізнавально-культурологічну ("трансмсія" культурної спадщини, наступність культури).

Починаючи з 1960-х рр., аналіз основних соціальних функцій масової комунікації був доповнений К.Райтом (розважальна функція), Д.Мак-Квейлом (мобілізаційна), Р.Клоссом (раціональна – інформування, навчання; ірраціональна – тиск, маніпулювання, реклама, пропаганда; соціальна психотерапія – релаксація)⁶⁵.

У 1967 р. на Міжнародній конференції в Індії у виступі, присвяченому порівняльному аналізу соціальних процесів, Г.Лассуелл доповнив свою знамениту формулу-модель новими елементами: "хто, з якими намірами, у якій ситуації, застосовуючи яку стратегію, досягає якої аудиторії, з яким результатом"⁶⁶.

2.2. Практика контент-аналітичних досліджень пропаганди

Практика зарубіжних пропагандистських досліджень насамперед зосереджувалася на вивченні матеріалів з Радянського Союзу. Так, у 1950 рр. західні аналітики на основі кількісного аналізу статей газети "Правда" відзначили різке зниження кількості посилань на Й.Сталіна. Звідси закономірний висновок: послідовники колишнього керівника держави намагаються дистанціюватися від нього⁶⁷.

Американські експерти також ретельно досліджують тексти офіційних виступів вищих посадових осіб СРСР. Аналіз формально-структурних особливостей цих текстів дозволив

⁶⁵ Гнатюк О.Л. Из истории американской коммуникологии и коммуникативистик...

⁶⁶ Lasswell H. The Uses of Content Analysis Data in Studying Social Change // Science and Culture.– 1967.– Vol. 33.– No 4.

⁶⁷ Воронько В., Костинский А. Компьютерный анализ текстов / Радио Свобода, 26-06-2001.– <http://www.svoboda.org/hotnews/ticker.asp>

розкрити протиріччя в радянському керівництві. Прикладом є дослідження полеміки, що розгорнулася в політбюро ЦК КПРС у 1969-1974 рр. довкола питання про сутність радянської військової доктрини. У полеміку втягнулися члени політбюро Ю.Андропов, Л.Брежнєв, Г.Гречко, А.Громико та інші. Позиції сторін були реконструйовані американськими дослідниками шляхом контент-аналізу публічних виступів радянських керівників за матеріалами газет "Правда" і "Красная звезда". Аналогічним чином аналіз політичної інформації, опублікованої в загальнодоступній пресі, дозволив фахівцям США виробити вірогідні прогнози щодо зміни генерального секретаря ЦК КПРС у період 1982-1985 рр.⁶⁸

У 1970 рр. аналітики американської розвідки, вивчаючи китайські ЗМІ, відзначили зміну одного з термінів і припустили перехід Китаю до більш агресивних дій. Незабаром дійсно стався в'єтнам-китайський прикордонний збройний конфлікт⁶⁹.

Дослідження ідеологічної та політичної лексики стали популярними в 1950 рр. На думку науковців, зокрема О.Іголкіна⁷⁰, "якість" мови визначає рівень демократичності суспільства: чим простішою є мова і менш багатим лексичний запас, тим легше керувати людьми. Тому в тоталітарних суспільствах мова характеризується максимальною простотою і бідністю. Відтак ставиться проблема кількісного оцінювання міри цієї простоти.

Одним із досліджень, у якому зроблена спроба розв'язати цю проблему, став проєкт групи американських науковців Інституту Гувера Стенфордського університету на чолі з Г.Лассуеллом під назвою RADIR (Revolution and the Development of International Relations)⁷¹, виконаний у 1952 р.

⁶⁸ Исследование анонимных текстов на предмет выявления их авторов.– <http://bond007.h1.ru/new/glava12.htm>

⁶⁹ Исследование анонимных текстов на предмет выявления их авторов...

⁷⁰ Иголкин А.А. Масс-медиа в условиях глобализации...– С. 67.

⁷¹ Lerner D., Pool I., Lasswell H. Comparative Analysis of Political Ideologies // Public Opinion Quarterly.– Winter 1951-1952.– P.716-723; Lasswell H.D., Lerner D., Pool I.D. The Comparative Study of Symbols // Hoover Institute Studies.– Series C: Symbols.– No 1.– Stanford, CA: Stanford University Press, 1952.

Дослідження містить аналіз "ключових символів сучасної політики": основних тенденцій їх трансформації і характеру функціонування, а також власне методологічні проблеми вивчення символічної сфери⁷².

Соціологи пояснили свій підхід до аналізу символічного змісту так: "Що об'єднує більшість дослідників, які говорять про "символи" (як технічний термін для слів), так це інтерес до потоку слів як виразника уявлень. Слова є "символами", оскільки вони визначають (символізують) уявлення тих, хто використовує їх. Символьний аналітик має справу зі словами, відбираючи лише ті, що найточніше відображають уявлення, наявність або відсутність яких треба виявити та описати. Символи в такому розумінні служать "операціональними індексами" уявлень"⁷³.

Акцент у дослідженні робився на "зміні словникового запасу сучасних еліт". При цьому як аналітичний матеріал (досліджувані символи) були вибрані ключові слова, безпосередньо пов'язані зі стабільністю еліти того часу. Загальна схема проекту RADIR передбачала проведення порівняльного аналізу світових тенденцій стосовно п'яти цінностей-цілей: демократії, безпеки, братерства, достатку й освіти.

Схема аналізу включала такі категорії: обсяг уваги, яка приділяється відповідним виданням об'єкту дослідження (тій чи іншій цілі-цінності); оцінюваний контекст; інтенсивність (частота згадування); трансформація в часі і просторі, характер взаємодії символа із символом, символа з подією, події з символом.

Вивчалися передові та редакційні статті за 1 і 15 число кожного місяця, що з'явилися з 1890 р. по 1949 р. у провідних ("престижних") газетах п'яти країн: США, Великобританії, Росії, Франції, Німеччини (усього 19553 статті). Назви видань і роки, упродовж яких вони досліджувалися, наведені в табл. 2.1.

⁷² Кармадонов О.А. "Символ" в эмпирических социальных исследованиях // Социс.– 2004.– №6.– http://socis.isras.ru/SocIsArticles/2004_06/karmadonov.doc

⁷³ Lasswell H.D., Lerner D., Pool I.D. The Comparative Study of Symbols...– P.29.

Таблиця 2.1

Досліджувана преса та періоди

Країна	Газета	Період
США	The New York Times	1900–1949
Великобританія	The Times	1890–1949
Росія	Новое время	1892–1917
	Известия	1918–1949
Франція	Le Temps	1900–1942
	Le Monde	1945–1949
Німеччина	Nordeutsche Allgemeine Zeitung	1910–1920
	Frankfurter Zeitung	1920–1932
	Volkisihier Beobachter	1933–1945

Оскільки для аналізу були відібрані найбільш "престижні" газети, які читалися політичною елітою всього світу, то їх можна вважати досить надійним індикатором ідеологічних трансформацій у країнах, де вони друкуються⁷⁴.

За одиницю аналізу було взято символ. Далі складено словник мови політиків – набір із 416 символів, з яких: 206 – назви політичних одиниць; 210 – ключові ідеологічні символи. До першої групи включено імена політичних діячів, назви політичних інститутів (конгрес, парламент, прем'єр, фіюрер і т.д.), назви національних або подібного виду утворень (країн, національних меншин, континентів) тощо. Група ідеологічних символів головних ідеологічних систем, які змагалися півстоліття на арені світової політики, уключала назви політичних сил (маса, пролетаріат), позначення політичних актів і станів (світ, війна, революція), назви ідеологій (комунізм, лібералізм, демократія) тощо.

Методологічна посилка авторів⁷⁵ полягала в тому, що з точки зору комунікатора символи словника являють собою

⁷⁴ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации... – С.52.

⁷⁵ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации... – С.53.

ключові (ударні) моменти в будь-якій політичній інформації, водночас – це і ті вузли, на яких застрягає увага аудиторії. Таким чином, у політичному словнику конкретизована ця двостороння увага.

Необхідно підкреслити, що Г.Лассуелл з колегами обмежили список символів іменниками, вважаючи, що більш широкий список утруднить аналіз диференціації розподілу частот різних елементів стосовно елемента аналізу.

Дослідження полягало у фіксації частоти вживання кожного символу та оцінки (сприятлива, несприятлива і нейтральна), якою вони супроводжувалися в тексті. Для порівняльного аналізу політичного змісту газет було використано операції: порівняння частоти вживання одного і того ж символу як для різних газет, так і в різні періоди; визначення варіації вживання різних оцінок при символах; визначення варіації вживання однакових оцінок при однакових символах у різних газетах і в різні періоди.

Якщо звернутися до кількісних результатів⁷⁶, то виявилось, що в усіх газетах за весь період 416 символів були присутні 105004 рази. Як і очікувалося, символи з'являлися вкрай нерівномірно, проте дивувала міра цієї нерівномірності. З'ясувалося, що на будь-якому проміжку часу ядро символічної системи утворюють лише 18 слів – на них припадало від 52 до 86 відсотків усіх появ виділених символів. Відповідно, на інші 398 символів – від 14 до 52 відсотків.

Щодо оцінювання "якості" мови, то, як і очікувалося, найбідніша ідеологічна мова виявилася у США, найбагатша – у Франції (див. табл. 2.2). До 1917 р. Росія за багатством мови йшла слідом за Францією. У всіх країнах під час Першої та Другої світових воєн газетна мова істотно спрощувалася. Загальна тенденція ХХ ст. – збідніння мови.

⁷⁶ Иголкин А.А. Масс-медиа в условиях глобализации...– С.68.

Таблиця 2.2

Коефіцієнти одноманітності ідеологічної лексики передових і редакційних статей провідних газет п'яти країн у різні періоди ХХ ст.

Країна	До Першої світової війни	Перша світ. війна	20-і роки	Поч. 30-х років	Кін. 30-х років	Друга світова війна		Після Другої світ. війни	У цілому по країні
						1-й період	2-й період		
США	358	543	334	424	287	696	837	379	328
Великобританія	288	878	330	284	267	536	654	383	299
Франція	221	307	218	204	212	268	287	263	189
Німеччина	–	577	239	196	509	493	562	–	235
Росія	257 (1892-1904 рр.) 261 (1905-1913 рр.)	400	306	408	454	493	530	487	261

Примітка: Для розрахунків ступеня одноманітності ідеологічної лексики була використана формула англійського статистика У.Юла:

$$K=10000 \times \frac{S_2 - S_1}{S_1^2} ,$$

де S_1 – загальна кількість ідеологічних символів, що з'явилися в газетах за досліджуваний період (наприклад, 2000 за 20-і роки); S_2 – сума квадратів числа появ окремих символів (наприклад, слово "демократія" з'явилося 40 разів, "комунізм" – 35, "воля" – 30 і т.д.; тоді S_2 дорівнює $40^2 + 35^2 + 30^2 + \dots$)⁷⁷.

⁷⁷ Иголкин А.А. Масс-медиа в условиях глобализации... – С. 69.

2.3. Диференціація і розширення сфери застосування контент-аналізу

Як зазначено вище, у 1950-х рр. до досліджень проблем пропаганди і масових комунікацій починають активно залучати фахівців з різноманітних галузей знань. Однак незабаром виявилось, що контент-аналіз є чудовим інструментом і для вивчення широкого кола проблем гуманітарних і соціальних наук. Відтак він знайшов ефективне застосування в дослідженнях етнокультурних, політичних явищ, історії думки. Контент-аналіз почав інтенсивно розвиватися в літературних і лінгвістичних розрахунках.

Деякі ілюстративні приклади використання контент-аналізу в політичних дослідженнях 1950-1980 рр. маємо на сайті "Контент-аналіз у політиці"⁷⁸. Таким прикладом є дослідження Френка Клінгберга циклічності зовнішньої політики США⁷⁹.

Багато уваги науковці приділяли вивченню різних аспектів промов президентів країн, зокрема інаугураційних виступів президентів США. Наприклад, якщо Родерік Харт вивчав стилістику виступів Президентів США, то Метью Морен досліджував зміст промов Рональда Рейгана⁸⁰. Лаура Ольсон досліджувала релігійні мотиви (частоту і характер посилань) у всіх президентських промовах, починаючи від Гаррі Трумена і завершуючи Рональдом Рейганом⁸¹.

Науковці намагалися оцінити політичні платформи різних партій. Так, використовуючи комп'ютерні програми, Цві Нейменвірс проаналізував основні положення партійних

⁷⁸ Content Analysis in Political.–

<http://www.janda.org/workshop/content%20analysis/seminalstudies.htm>

⁷⁹ Klingberg F.L. The Historical Alternation of Moods in American Foreign Policy // World Politics.– 1952.– January.– PP. 239-273; Klingberg F.L. Cyclical Moods in American Foreign Policy.– Lanham, MD: University Press of American, 1983.

⁸⁰ Hart R.P. Verbal Style and the Presidency: A Computer-Based Analysis.– Orlando, Florida: Academic Press, 1984; Moen M.C. The Political Agenda of Ronald Reagan: A Content Analysis of the State of the Union Messages // Presidential Studies Quarterly.– 1988.– No 18. – Fall, (1988).– PP.775-785.

⁸¹ Olson L. Ronald Reagan and the New Fundamentalist/Evangelical Christian Right / Paper prepared for Political Science C27.– March.– 1989.

платформ за період з 1844 по 1964 рр. Пізніше ним було проведено схоже дослідження спільно з Г.Лассуеллом⁸². Джеральд Помпер також вивчав платформи політичних партій, широко застосовуючи статистичні методи, а британець Річард Роуз здійснив порівняльне дослідження "електоральних маніфестів" ("election manifestos") британських партій⁸³.

У 1980-х рр. група європейських учених зробила спробу реалізувати "Проект порівняння маніфестів" ("Comparative Manifesto Project"), який передбачав спільними зусиллями провести контент-аналіз передвиборних програм політичних партій 20 країн, починаючи з кінця Другої світової війни і завершуючи 1985 р.⁸⁴

Одним із напрямів застосування контент-аналізу стала історія. Необхідність урахування інформації масових джерел підштовхнули істориків до застосування кількісних методів, ЕОМ і в тому числі контент-аналізу. У 1960 рр. майбутнє такого підходу уявлялося надзвичайно перспективним. Ним захоплювалися у великих наукових центрах Великобританії, Німеччини, США, Франції. Навіть сформувався новий науковий напрям – квантитативна історія або кліометрика⁸⁵. Добре відомим є вислів її класика француза Емануеля Ле Руа Ладюрі (1973 р.): "...історія, яка не є квантифікованою, не може претендувати на те, щоб уважатися науковою"⁸⁶.

⁸² Namenwirth J.Z., Some Long- and Short-Term Trends in One American Political Value: A Computer Analysis of Concern with Wealth in 62 Party Platforms // *The Analysis of Communication Content* / George Gerbner et al. (eds.).– New York: Wiley, 1969.– PP. 223-241; Namenwirth J.Z., Lasswell H.D. *The Changing Language of American Values: A Computer Study of Selected Party Platforms.*– Beverly Hills, CA: Sage Publications, 1970.

⁸³ Pomper G., Lederman S.S. *Elections in America: Control and Influence in Democratic Politics.*– New York: Longman, 1980; Rose R. *Do Parties Make a Difference?*– NJ: Chatham, 1984.

⁸⁴ *Ideology, Strategy and Party Change: Spatial Analyses of Post-War Election Programmers in 19 Democracies* / Ian B., Robertson D., Hearl D. (eds.).– Cambridge: Cambridge University Press, 1987.

⁸⁵ Ключин А.И., Кончаков р.Б., Сельцер Д.Г. Компьютерные методы в исследовании номенклатуры КПСС.– <http://elis.pstu.ac.ru/kljukin-selts.htm>

⁸⁶ Le Roy Ladurie E. *The Territory of the Historian.*– Chicago, 1979.– P.15.

Прикладом застосування контент-аналізу в дослідженнях з історії США може служити такий⁸⁷. Визвольній війні 13 англійських колоній у Північній Америці проти англійського панування в 1775-1783 рр. передувало формування північноамериканської нації, що супроводжувалося зростанням національної самосвідомості. Питання про те, коли відбулася психологічна, духовна і моральна емансипація мешканців колоній від метрополії, тривалий час залишалося дискусійним. Згідно з однією точкою зору, "американізація" колоністів проходила повільно і поступово протягом усього колоніального періоду. Інші історики схилилися до думки, що відчуття належності до американської нації реально виникло лише після здобуття колоніями незалежності. Для перевірки цих гіпотез був проведений контент-аналіз семи американських газет періоду 1735-1775 рр. з п'ятох найбільших міст країни – Нью-Йорка, Бостона, Філадельфії, Вілліамсбурга і Чарлістона.

Як значеннєві одиниці було взято п'ять груп слів-символів, які ототожнювали (ідентифікували) населення колоній з британським або північноамериканським співтовариством:

– символи прямої британської ідентифікації ("Британська Північна Америка", "англійські колонії", "англійські провінціали", "британські колоністи");

– символи ідентифікації з британською короною ("колонії його величності", "королівські колонії", "королівські американці");

– символи неявної британської ідентифікації ("колонії", "провінції", "наші колонії в Америці", "колоністи", "провінціали");

– символи неявної американської ідентифікації ("північноамериканський континент", "американські колонії", "країна", "з'єднані колонії", "американські колоністи");

– символи прямої американської ідентифікації ("американці", "північноамериканці", "Америка", "Північна Америка").

⁸⁷ Миронов Б. История в цифрах.– Л., 1991.– С.24-26.

Три перші групи символів утворюють кластер, що ототожнює колонії і колоністів з Британією, інші – ідентифікували колонії та їхнє населення з Америкою.

Було проаналізовано 1435 випусків газет за всі роки, прослідковано появу кожного слова-символу або відповідного займенника. Результати показали, що використання символів національної належності за 41 рік зросло в 320 разів – з 0,66 символу на один газетний номер у 1735–1740 рр. до 22 символів у 1771–1775 рр. Це свідчило про актуальність проблеми національної ідентифікації. Аналіз також дозволив простежити, як змінювалося співвідношення символів американської і британської ідентифікації. У 1730 рр. переважали символи Британії – на них припадало 57 відсотків. До кінця ж колоніального періоду кількість символів, що ідентифікують населення з північноамериканською нацією, досягло 63 відсотків. Можна констатувати, що зростаюче почуття належності до американської нації психологічно готувало північноамериканців до війни за незалежність.

Ще одним напрямом використання контент-аналізу стала лінгвістика. Тут уже традиційною проблемою є визначення авторства художніх і наукових творів. Зокрема, коли перед Нобелівським комітетом з присудження премій постала проблема ідентифікації "Тихого Дону" М.Шолохова, шведські вчені провели скрупульозний аналіз ряду текстів М.Шолохова, написаних ним у різні роки: зіставлялися довжина речень, особливості прикметників, лексичний запас різних творів тощо⁸⁸. З появою ЕОМ застосування контент-аналізу в лінгвістиці істотно активізувалося.

Значний внесок у розвиток контент-аналізу внесли дослідження з психології та психотерапевтики. Яскравим прикладом творчого застосування контент-аналізу в психології є роботи одного з найвідоміших у 1930-1970 рр. американських фізіологів і психологів Келвіна С.Хелла. Він розпочав вивчення мрій ще в 1940 рр., і вже перше контент-дослідження, зміст

⁸⁸ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации... – С.10.

якого викладений у статті "Про що люди мріють"⁸⁹, дало К.С.Хеллу підґрунтя до створення власних теорій мрій.

Ранні дослідження К.С.Хелла⁹⁰ базувалися на анонімних матеріалах, написаних студентами, які, як здавалося, були єдиними ідеальними об'єктами через свою готовність та інтерес взяти участь у дослідях, а також готовність відверто відповідати на запитання. Однак незабаром учений почав збирати повідомлення від дітей, людей похилого і працездатного віку з різних континентів, а також тих, хто вів щоденники мрії. К.С.Хелл почав роботу з тематичних досліджень 15-25 мрій студентів, а до кінця свого життя зібрав колекцію яка містила аж 50000 мрій. Намагаючись їх узагальнити, К.С.Хелл незабаром розробив кількісну систему кодування, яка класифікувала зміст мрій за призначенням, ціллю, особливостями, взаємодією, емоціями, невдачами тощо.

Саме контент-аналіз дав ідеї для прекрасних статей про когнітивну теорію мрій і метафоричну теорію символів мрій. У своїй когнітивній теорії К.С.Хелл стверджував, що мрії виражають "концепції" особистості, членів її сім'ї та друзів, соціального середовища. Наприклад, вони демонструють такі особистісні концепції, як "слабкий", "наполегливий", "нелюбий", "владний", "ворожий" тощо. К.С.Хелл також розвинув метафоричну теорію символів мрій, яку демонстрував через відповідні метафоричні вирази, що мали місце в слензі та поезії.

Емпіричні дані К.С.Хелла підтверджують, що мрії груп людей з усіх континентів більше схожі між собою, ніж відмінні. Щодо відмінностей, то вони переважно пов'язані з використанням різної культурної термінології. У той же час, науковець відзначав істотні індивідуальні розходження в частоті елементів мрій; ці розходження передають занепокоєність, емоційну стурбованість, інтереси, пропозиції, що, за К.С.Хеллом, виражає "нерозривність" між процесами

⁸⁹ Hall C.S. What People Dream About // Scientific American.– 1951.– No 184.– PP. 60-63.

⁹⁰ A Brief Biography of Calvin S.Hall.–
<http://mind.ucsc.edu/dreams/About/calvin.html>

задоволення мрій і пробудження думки. Його робота зі щоденниками мрій кількох людей протягом багатьох років і навіть десятиліть показала дивовижну стабільність змісту мрій, хоча й спостерігалися деякі зміни, зокрема в тих, кого називають "вічними" мрійниками.

У період 1961–1965 рр., К.С.Хелл вивчав зміст нічних мрій, зібраних у лабораторії сну в Інституті досліджень мрій у Майямі. Саме тоді він і Роберт Ван де Кастрл намагалися розробити загальну всебічну систему кодування для відтворення об'єктивного змісту мрій. Ці результати були покладені в основу їхньої спільної книги "Контент-аналіз мрій" (1966 р.), а також ряду його, тепер класичних, статей⁹¹.

Уявлення про спектр контент-аналітичних досліджень у психотерапії і суміжних науках можна отримати з невеликого переліку робіт з цього напрямку за різні роки, наведеного в⁹².

Методика контент-аналізу почала застосовуватися не лише науковцями і політиками. Починають усвідомлювати його важливість представники бізнесу, громадських, міжнародних, релігійних організацій. Зокрема, відомі факти контент-

⁹¹ Hall C.S., Van de Castle R. The Content Analysis of Dreams.– New York: Appleton-Century-Crofts, 1966; Hall C.S. Content Analysis of Dreams: Categories, Units, and Norms // The analysis of communication content / G.Gerbner (Ed.).– New York: Wiley, 1969; Hall C.S. Normative Dream Content Studies // Dream Psychology and the New Biology of Dreaming / M.Kramer (Ed.).– Springfield, IL: Charles C.Thomas, 1969.– PP. 175-184.

⁹² Harway N.I., Iker H.P. Computer analysis of content in psychotherapy // Psychological Reports.– 1964.– №14.– pp.720-722; Laffal J. An approach to the total content analysis of speech in psychotherapy // Research in psychotherapy / J.M.Shlien (Ed.).– Washington: American Psychological Association, 1968.– pp.277-294; Dahl H. A quantitative study of a psychoanalysis // Psychoanalysis and contemporary science / Eds. Holt R.R., Peterfreund E.– New York: Macmillan, 1972.– PP. 237-257; Martindale C. The therapist-as-fixed-effect fallacy in psychotherapy research // Journal of Consulting and Clinical Psychology.– 1978.– №6(46).– pp.1526-1530; Mergenthaler E., Kachele H. Changes of latent meaning structures in psychoanalysis // Sprache und Datenverarbeitung.– 1985.– №9.– PP.21-28; Mergenthaler E. Textbank Systems. Computer science applied in the field of psychoanalysis.– Heidelberg: Springer, 1985; Luborsky L., Crits-Christoph P., Mintz J., Auerbach A. Who will benefit from psychotherapy? Predicting therapeutic outcomes.– New York: Basic Books, 1988.

досліджень музики, творів мистецтва, телевізійних шоу, комерційної реклами⁹³.

Одними з перших почали використовувати контент-аналіз теологічні науки. Перше комп'ютеризоване дослідження в гуманітарній галузі було проведене священиком Робертом Бузою для складання частотного словника робіт італійського теолога Фоми Аквінського. Католицькі теологи створили самостійну школу комп'ютеризованого аналізу сакральної літератури з центром у французькому католицькому університеті м.Лувен. У 1968 р. при ньому було створено Центр електронної обробки документів CETEDOC⁹⁴.

Активно використовувало і використовує контент-аналіз ЮНЕСКО. Коли восени 1976 р. конференція ЮНЕСКО в Найробі зіткнулася з проблемами, пов'язаними з міжнародними комунікаціями і новим світовим інформаційним порядком, то одним із її рішень стало включення до бюджету на 1977-1978 рр. витрат для вивчення на основі ЗМІ різних країн світу образів іноземних держав різних соціальних систем. До запропонованого рішення Індія, Югославія та скандинавські

⁹³ Paisley W.J. Identifying the unknown communicator in painting, literature and music: the significance of minor encoding habits // *Journal of Communication*.– 1964.– No14.– PP.219-237; Simonton D.K. Thematic fame and melodic originality in classical music: a multivariate computer-content analysis // *J. Pers.*– 1980.– No 48. – PP.206-219; Marshall R., Lu Z. A Content Analysis of Chinese Magazine Ads // *Journal of Advertising*.– 1988.– No 17(4).– PP. 43-48; Suzuki N. The Changing Pattern of Advertising Strategy by Japanese firms in the U.S. market: Content Analysis // *Journal of International Business Studies*.– 1980.– No 12.– PP.63-72; Lord J.B., Eastlack J.O., Stanton J.L. Health Claims in Food Advertising: Is There a Bandwagon Effect? // *Journal of Advertising Research*.– 1987.– 27 April/May; Holbrook M.B. More on Content Analysis in Consumer Research // *Journal of Consumer Research*.– 1977.– Vol. 4.– PP.176 -177; Kassarijan H. Content Analysis in Consumer Research // *Journal of Consumer Research*.– 1977.– Vol. 4.– June 8-18.

⁹⁴ Hockey S. An Historical Perspective in Rahts. Information technology in the Humanities. P.22; Tombeur P. Research carried out at the Centre de Traitement Electronique des Documents of the Catholic University of Louvain // A.J.Aitken et al. (eds.) *The Computer and Literary Studies*. Edinburgh. 1973. P.335-340; Hamesse J. Automatic Processing of Philosophical Works at the Catholic University of Louvain in *Ibid*. P.341-344; Bradley J., Rockwell G. TACTweb: workbook.– <http://tactweb.humanities.mcmaster.ca/tactweb/doc/twintro.htm>

країни внесли доповнення: зробити особливий акцент на образі країн, що розвиваються, у ЗМІ промислово розвинених країн⁹⁵.

На осінь 1978 р. було розроблено масштабний проект, і в лютому 1979 р. розіслано учасникам проекту заключні інструкції. Міжнародні новини досліджувалися в країнах Північної Америки (США), Латинської Америки (3 країни), Африки (6 країн), Середнього Сходу і Єгипті (Єгипет), Азії (2 країни), Східної Європи (Радянський Союз), Західної Європи (3 країни). Загалом дослідження охопило 29 країн за період, еквівалентний двом тижням. У кожній країні вивчалися 3-4 тижневих друкованих видання, основні новинні програми радіо і телебачення. Наприклад, у США розглядалися *New York Times*, *Washington Post*, *Los Angeles Times*, *New York Daily News*, *Minneapolis Tribune*, *Charlotte Observer*, *CBS Evening News*; у Радянському Союзі – *Правда*, *Известия*, *Комсомольская правда*, національне телебачення; у Єгипті – *Al Ahram*, *Al Akhbar*, *Al Gumhuriyah*, національне радіо і телебачення.

По кожній новині проект передбачав фіксувати такі параметри, як локальність (місце, де вони відбувалися), джерело (де народжуються новини), авторство, характеристика і національність учасників, тема (зміст новин) і тематика (контекст новин). У свою чергу теми можна було віднести до одного з 17 класів:

- дипломатичні і міждержавні проблеми;
- політичні проблеми в середині країни;
- проблеми безпеки та оборони;
- економічні проблеми;
- проблеми міжнародної допомоги;
- проблеми соціального забезпечення;
- кримінальні події;
- культурно-мистецькі проблеми;
- релігійні проблеми;
- науково-технічні і медичні проблеми;
- спортивні проблеми;

⁹⁵ Stevenson R.L. *Research Methodology // Foreign news and the new world information order* / Robert L.Stevenson, Donald Lewis Shaw (Editors).– Ames, Iowa: The Iowa State University Press, 1986.– P.21.

- розваги;
- особистості;
- хобі, тварини, секс;
- студентські проблеми;
- екологічні проблеми;
- природні катастрофи і катаклізми;
- інше.

Оскільки інформаційні новини можуть охоплювати декілька тем, то передбачалося виділяти головну тему і не більше трьох додаткових.

Фрагмент результатів висвітлення іноземних новин у мексиканських медіа, наведених Робертом Стівенсоном в ⁹⁶, маємо в табл. 2.3 і 2.4.

Навіть побіжний аналіз результатів показує, що найбільше уваги мексиканські медіа приділяють проблемам міжнародної економіки, культурно-освітній тематиці, зовнішній і внутрішній політиці зарубіжних країн, а щодо інформаційного впливу, то безсумнівними лідерами тут є сусідні американські країни, зокрема США, а також Західна Європа. Радянський Союз і Азія займають останнє місце у сфері інтересів мексиканських медіа.

Проведене ЮНЕСКО дослідження не концентрувалося на якихось конкретних проблемах, а являло собою досить загальний макрорівневий аналіз міждержавних інформаційних обмінів. Будучи нескладним у методичному і методологічному плані, але складним у реалізації через міжнародний, різнокультурний, різносоціальний і багатомовний характер, це дослідження однак чітко продемонструвало, що одні держави говорять про інші, які будують стосунки, як виражають власні та поважають чужі інтереси. Це вивчення могло стати основою для розширення і розбудови нового інформаційного порядку і діалогу між державами.

⁹⁶ Stevenson R.L. Research Methodology... – P.34.

Таблиця 2.3

Теми іноземних новин у мексиканських медіа

Новини	Усі згадування			Лише головні теми		
	%	Z	ранг	%	Z	ранг
Загальна кількість новин (n=2436)						
Міжнародна політика	22	0.84	4	11	0.39	6
Внутрішня політика	23	0.95	3	12	0.53	4.5
Оборона, безпека	10	-0.38	7	3	-0.77	8
Економіка	30	1.66	1	21	1.83	1
Міжнародна допомога	2	-1.20	11	1	-1.06	10
Соціальні проблеми	1	-1.30	12	0	-1.20	12
Кримінальні події	9	-0.49	8	5	-0.48	7
Освіта, наука, культура	24	1.05	2	14	0.82	3
Спортивні новини	19	0.54	5.5	18	1.39	2
Хобі, тварини, секс	19	0.54	5.5	12	0.53	4.5
Природні катастрофи і катаклізми	3	-1.10	9.5	2	-0.91	9
Інші	3	-1.10	9.5	1	-1.06	10.5

Таблиця 2.4

Географічні акценти в новинах мексиканських медіа

Новини	Усі згадування			Лише головні теми		
	%	Z	ранг	%	Z	ранг
Загальна кількість колонко-сантиметрів (n=136070)						
Північна Америка	36	1.06	2	20	0.66	2.5
Латинська Америка	41	1.44	1	46	1.91	1
Африка	12	-0.76	5	3	-0.87	7
Середній Схід / Єгипет	10	-0.91	6	6	-0.60	5
Азія	7	-1.13	8	4	-0.77	6
Західна Європа	28	0.45	4	20	0.66	2.5
Східна Європа / Радянський Союз	8	-1.06	7	9	-0.83	6
Світ у цілому	34	0.91	3	10	0.24	4
Загальна кількість новин (n=2436)						
Північна Америка	22	0.13	4	23	1.00	2
Латинська Америка	46	1.91	1	32	1.87	1
Африка	6	-1.05	8	3	-0.93	7.5
Середній Схід / Єгипет	9	-0.83	6	6	-0.63	5
Азія	9	-0.83	6	5	-0.73	6
Західна Європа	30	0.11	3	21	0.81	3
Східна Європа / Радянський Союз	9	-0.83	6	3	-0.93	7.5
Світ у цілому	31	0.80	2	8	-0.45	4

Примітка: Для поліпшення розуміння результатів розподіли частот у таблицях приведені до стандартизованої величини Z і проранжовані.

2.4. Удосконалення методики контент-аналізу і розвиток якісних досліджень

Як зазначає Л.Нейман⁹⁷, цілі, для досягнення яких застосовували контент-аналіз у 1950-1980 рр., були найрізноманітнішими. Вивчали: особливості і тематику популярних пісень і релігійної символіки в гімнах; тенденції, що відображалися в газетних статтях, та ідеологічну тональність редакційних передовиць; гендерні стереотипи в підручниках і навчальних посібниках; частоту появи людей різних рас у телевізійних рекламних роликах і програмах; ворожу пропаганду в роки війни; обкладинки популярних журналів; психіку самогубців через передсмертні записки; тематику оголошень і гендерні розбіжності під час бесід.

Однак попри цільові відмінності були спільні риси в контент-аналізі цих років. Як стверджує О.Ольсті⁹⁸, цей період характеризується переходом від досліджень, які були головним чином зайняті підрахунком, до досліджень з вивчення причин і ефектів змісту зв'язків. Це твердження стосується як інструментального, так і репрезентативного напрямку в контент-аналізі.

У рамках інструментального напрямку далі виділяються декілька течій: одні дослідники прагнули модернізувати методику Лассуелла на базі символів, виражених у мовній формі, інші – звертаються до аналізу немовних факторів.

Перша течія тісно пов'язана з проблемою стандартизації категорій аналізу, тобто відшукуванням таких "універсальних" категорій аналізу, які можна застосувати практично в будь-якому дослідженні. У вже згадуваному проекті RADIR Г.Лассуелла стратегія дослідження безпосередньо і неоднозначно мала на меті використання універсального поняття "символ" як засобу вимірювання (інструмент аналізу), категорії (теоретичний базис) й одиниці аналізу. Тим не менше

⁹⁷ Ньюман Л. Неопросные методы исследования // Социологические исследования.– 1998.– №6.– С.121.

⁹⁸ Цитата за: Gallagher C., Marken M., Kim M., Phillipson T., Dodge T. The Use of Content Analysis in Researching Adolescent Issues...

навіть самі автори проекту відверто визнавали проблему свого підходу і такого способу редукції: як "символи" використовувалися винятково іменники. Значною мірою це було наслідком недостатнього теоретичного аналізу символічного. "У міру того, як ми просувалися вперед, стало зрозумілим, що з багатьох причин краще було б вибрати якийсь інший підхід. Унаслідок ігнорування частин мови, які не були іменниками, залишилися необстеженими широкі території змісту. Вирішення цієї проблеми ми залишаємо для контент-аналізу майбутнього"⁹⁹.

Крім того, запровадження універсальної одиниці аналізу складає проблему і з більш загальної точки зору. Ця проблема і не могла мати однозначного вирішення, оскільки вибір категорій аналізу, у першу чергу, визначається предметом конкретного дослідження, його цілями та завданнями.

Хоча в деяких випадках такий підхід виявився корисним. Показовим у цьому плані є дослідження кількості різних слів, які людина вживає в мові текстових масивів з 100, 200, 500 і 1000 слів. У шизофреніків, зокрема, ця кількість є набагато меншою. Тексти, написані ними, ідентифікуються такими показниками: усе подається в негативному тоні; тексти орієнтовані на минуле; багато місця приділено власним переживанням і висловленню власних думок про все. Проведений науковцями аналіз лінгвістичних розходжень між шизофреніком і нормальною людиною підтвердив гіпотезу, що кількість прикметників на 100 дієслів у нормальної людини є значно більшою, ніж у шизофреніка¹⁰⁰.

Як уже зазначалося, екстралінгвістичні фактори використовував А.Джордж, який вивчав залежності між елементами системи "зміст – пропагандист – керівні органи".

У рамках інструментальної течії мав місце й інший підхід – соціально-критичний, пов'язаний з іменами Х.Хорта, Л.Лоуентала, Дж.Гербнера, які використовували як одиниці аналізу визначений набір тем. Тематичний аналіз допомагав виявити явні і приховані теми текстів. Його намагалися

⁹⁹ Lasswell H.D., Lerner D., Pool I.D. The Comparative Study of Symbols... – P.59.

¹⁰⁰ Исследование анонимных текстов на предмет выявления их авторов...

застосувати тоді, коли потрібно визначити сукупне значення ряду частин певної програми або пропагандистського матеріалу.

Прикладом тематичного контент-аналізу може служити цікаве дослідження, проведене Торренсом і Медоу. Вони спробували з'ясувати, що, на думку іноземців, є характерним для американців. Були відібрані статті про Америку 11 іноземних авторів, опубліковані в 1908-1948 рр. У цих матеріалах дослідники зафіксували 307 тем, згрупованих у 12 класів. Ураховавши частоту появи в текстах тем за класами, автори отримали таблицю, що характеризує найпопулярніші уявлення іноземців про Америку та американців. Такі уявлення важливі для пропагандиста, оскільки він може варіювати теми своїх пропагандистських матеріалів так, щоб змінити або закріпити бажані уявлення про свою країну¹⁰¹.

Однак найреволюційніші зрушення мали місце в рамках репрезентативного контент-аналізу, і пов'язані вони були насамперед з іменем американського психолога Чарльза Осгуда. Будучи фахівцем у галузі експериментальної психології, він став одним з ініціаторів лінгво-психологічного вивчення мовної діяльності. Базовою теоретичною посилкою при цьому стала ідея про зв'язок породження такого лінгвістичного феномена, як текст із комплексом усєї соціальної і психологічної діяльності автора тексту¹⁰².

Узявши в 1955 р. участь у конференції з проблем контент-аналізу, скликаній сектором лінгвістики і психології Ради з досліджень у сфері соціальних наук університету Іллінойс, Ч.Осгуд у своїй доповіді дав пропозиції щодо прагматичних висновків із семантичних звичок джерела. Методика Лассуелла при цьому отримала таку інтерпретацію: індикатором уваги комунікатора до певної проблеми може служити частота зустрічання в тексті логічних одиниць, які стосуються цієї проблеми.

¹⁰¹ Методы сбора информации в социологических исследованиях...– С.79; Коробейников В.С. Редакция и аудитория.– М., 1983.– С.140-141.

¹⁰² Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации...– С.42.

Ч.Осгуд запропонував методика оцінювального аналізу тверджень, розроблену ним разом із Дж.Нанеллі і С.Сапортою¹⁰³. Основна ціль методики: статистично виміряти ступінь (інтенсивність) відношення комунікатора до певних об'єктів завдяки тому, що це відношення лінгвістично зафіксоване в тексті. Методика полягає у фіксуванні цього відношення під час кодування текстів шляхом використання семантичного диференціалу, тобто сукупності певних шкал на просторі діаметрально протилежних якісних прикметників (наприклад, корумпований/чесний, близький/далекий, активний/пасивний). Л.Н.Федорова дає оцінку: "Семантичний диференціал демонструє свої можливості в тих випадках, коли необхідно дати узагальнену, але достатньо багатопланову оцінку складного за своїми компонентами тексту"¹⁰⁴.

Узагальнення методики дістало назву "репрезентативної моделі Осгуда"¹⁰⁵. Ця модель виводила зовсім на інше розуміння контент-аналізу: тексти тісно корелюють зі своїм контекстом, вони представляють його; тому з самого тексту можна вийти на розуміння контексту. Інше твердження про контент-аналіз виводиться з того, що тема буде тим виразнішою, чим більше посилань на неї існує в тексті. У результаті маємо два підходи: кількісні дослідження, які завершуються якісними висновками; а також якісні дослідження, що потребують визначення кількості.

Збагативши контент-аналіз методикою "зв'язаності символів", Ч.Осгуд зі своїми колегами дозволив віднайти не випадково зв'язані між собою елементи змісту, одержати уявлення про "асоціативну структуру" свідомості комунікатора. Ця методика стала початком застосування в контент-аналізі кореляційної техніки і факторного аналізу, а з появою ЕОМ її практична реалізація стала цілком реальною.

Як можна побачити з усього сказаного вище, обидва напрями – інструментальний і репрезентативний – переходять

¹⁰³ Osgood Ch., Saporta S., Nunnally J. Evaluative Assertion Analysis // Litera.– 1956.– Vol.3.

¹⁰⁴ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации...– С.43.

¹⁰⁵ Osgood Ch. The Representational Model and Relevant Research Methods...

до більш тонких методів аналізу текстів. Від середини 1950 рр. дослідники все більше приділяли увагу не лише простій наявності або відсутності категорій у тексті, а й зв'язкам між категоріями. Для цього, у першу чергу, ураховують спільне зустрічання ("concordances") слів різних категорій. "Може виявитися, що для деяких категорій спостерігається тенденція їхнього спільного вживання, а для інших – навпаки. Зрозуміло, що вивчення зв'язків між категоріями значно розширило коло проблем, які може вирішувати контент-аналіз"¹⁰⁶. Наприклад, дозволило точно визначити значення конкретних слів і фраз у контексті або передбачати поєднання кількох слів-символів в одній і тій же ідеологічній системі.

Зауважимо, що інструментарій спільного зустрічання бере свій початок із середніх віків. Середньовічні дослідники Біблії намагалися шукати паралелі між Старим і Новим Заповітом, відшукуючи місця, де слова з тексту Старого Заповіту присутні в Новому. Здійснювалося це шляхом створення для різних значущих слів так званих індексів, які містили всі фрагменти текстів книг Біблії, де виникали вказані слова. Саме в такий спосіб було розпочато вивчення тематичної відповідності головних осіб, місць, речей та ідей, присутніх у Біблії¹⁰⁷.

Тут варто знову повернутися до конференції 1955 р. в Іллінойському університеті, у якій взяло участь чимало науковців – піонерів контент-аналізу. Науковим доробком конференції стало видання відомого всім дослідникам з контент-аналізу збірника статей¹⁰⁸, у якому представлені основні тенденції розвитку методу. Серед учасників найвпливовішою і найчисельнішою була група дослідників Гарвардського університету, яка представляла методіку, яку вони назвали "ситуаційний контент-аналіз"¹⁰⁹. Методику ситуаційного аналізу можна наочно представити таким чином. На першому етапі формується тематична матриця (табл. 2.5),

¹⁰⁶ Контент-анализ в информационном поле.–

<http://taimer.wormhole.ru/discussion/ContentAnalysis.doc>

¹⁰⁷ Bradley J., Rockwell G. TACTweb: workbook...

¹⁰⁸ Trends in Content Analysis / I. de Sola Pool (ed.).– Urbana, Ill.: University of Illinois Press, 1959.

¹⁰⁹ Roberts C.W. Content Analysis...– P.2697.

кожен рядок якої представляє окремі блоки (фрагменти) тексту, а колонки відповідають тематичним змінним або категоріям – словам і фразам тексту. Відтак, кожна клітина таблиці, яка лежить на перетині певного рядка і певної колонки, містить частоту появи відповідної категорії у відповідному блоці тексту. На другому етапі обчислюється матриця, що оцінює зв'язки між темами: якщо деякі теми з'являються в текстових блоках разом, то вони позитивно пов'язані; якщо вони не зустрічаються в одних і тих же текстових блоках, то зв'язок між темами – негативний. Завершуючи процедуру ситуаційного аналізу, дослідник шукає пояснення зв'язків між темами. Узагальнюючи, можна сказати, що в особі ситуаційного аналізу маємо тематичний аналіз зв'язків у рамках класичного контент-аналізу. Зауважимо, що аналіз зв'язків рівноцінно може мати місце у випадку як інструментального, так і репрезентативного аналізу.

Таблиця 2.5

Матриця даних класичного контент-аналізу

Номер текстового блоку	Категорія аналізу (тема)			
	Тема 1	Тема 2	...	Тема n
1	14	5	–	4
2	–	2	–	–
3	5	–	–	2
4	2	–	6	9
5	–	2	5	–

Тим не менше, навіть доповнюючи класичний аналіз аналізом зв'язків, неможливо позбутися обмежень першого: підрахунок частоти категорій дозволяє робити висновки лише щодо переважання категорій (тем) у тексті. Наприклад, якщо в процесі дослідження виявилось, що такі поняття, як "маніпуляції новинами" і "обурення населення" в багатьох текстових блоках зустрічаються разом, то це лише означає, що вони певним чином пов'язані між собою. На основі такого аналізу неможливо робити правдиві висновки про реальні причини і наслідки: маніпуляції новинами викликають обурення людей чи приховування обурення населення здійснюється

шляхом маніпулювання новинами. Для цього не вистачає інформації про семантичні зв'язки між категоріями.

Подібного роду проблеми призвели до появи нових видів контент-аналізу: семантичного та мережного, які активно розвиваються від 1970 рр. Семантичний аналіз дозволяє не лише виявляти зв'язки такими елементарними категоріями, як слова і фрази, а виявляти і класифікувати зв'язки між іншими різними лінгвістичними одиницями (не обов'язково слово або фраза) на основі а рїогі сконструйованих шаблонів (семантичних граматик). У семантичній граматиці категорії виступають як її синтаксичні компоненти, між якими будуються взаємовідносини. Наприклад, у¹¹⁰ розвивається двомісна семантична граMATика поняття "образа", яка містить одну синтаксичну компоненту для об'єкта образи та іншу – для позначення дії, яка визначає, в який спосіб образили об'єкт. Табл. 2.6 унаочнює результат першого етапу перетворення початкового тексту. Зауважимо, що цифри в клітинах таблиці засвідчують не число появи елементів, а код (номер) відповідно суб'єкта, дії та об'єкта. У таблиці текстовий блок кодується деякою послідовністю трійок "суб'єкт – дія – об'єкт". Відтак, наприклад, змінна (колонка) "суб'єкт" показує, що саме виступає в ролі суб'єкта в тексті, а, аналізуючи всі три змінні, можна визначити, у який спосіб відбувається взаємодія між суб'єктами і об'єктами.

Таблиця 2.6

Матриця даних семантичного аналізу тексту

Номер текстового блоку	Суб'єкт	Дія	Об'єкт
1	10	34	53
2	8	21	78
3	13	34	61
4	10	35	61
5	16	29	44

¹¹⁰ Markoff J., Shapiro G., Weitman S. Toward the integration of content analysis and general methodology // Sociological Methodology / Heise D.R. (ed.).– San Francisco: Jossey-Bass, 1975.– PP.1-58.

Зауважимо, що семантичний аналіз використовується більшою мірою дослідниками репрезентативного напрямку, ніж інструментального. Приклад першого виду дослідження маємо в роботі С.В.Робертса¹¹¹, інструментальне дослідження представляють Л.А.Готтшалк і С.М.Каплан в¹¹².

Для дослідження змінних (категорій), між якими існують складні взаємозв'язки, насамперед причинно-наслідкового характеру, використовують мережний аналіз. Він дозволяє не лише визначати місце змінних у мережі, але й дозволяє в певний спосіб виміряти цю взаємодію. Наведемо приклад. Нехай у мережі маємо три категорії А, В, С, які пов'язані між собою причинно-наслідковими зв'язками: $A \rightarrow B$, $A \rightarrow C$, $C \rightarrow B$, $A \rightarrow C \rightarrow B$, представленими на рис. 2.1.

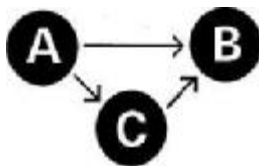


Рис. 2.1. Приклад причинно-наслідкових зв'язків у мережі

Тоді зв'язок $A \rightarrow B$, у якому А є причиною, а В – наслідком, характеризується числом 0,5. Що означає: з усіх чотирьох зв'язків між елементами мережі половина припадає на зв'язок між $A \rightarrow B$. Зауважимо, що взаємодія зазначених категорій стосується лише окремого блоку тексту. Що стосується всього тексту, то причинно-наслідкові зв'язки в ньому можна подати, наприклад, у вигляді табл. 2.7.

¹¹¹ Roberts C.W. Other than counting words: A linguistic approach to content analysis // Social Forces.– 1989.– No 68.– PP.147-177.

¹¹² Gottschalk L.A., Kaplan S.M. A quantitative method of estimating variations in intensity of a psychologic conflict or state // Archives of Neurology and Psychiatry.– 1958.– No 79.– PP.688-696.

Таблиця 2.7

Матриця даних мережного аналізу тексту

Номер блока тексту	A→B	A→C	C→B	B→A	C→A	B→A
1	0,50	0,25	0,25	0,00	0,00	0,00
2	0,25	0,00	0,50	0,00	0,25	0,00
3	0,00	0,00	0,25	0,25	0,50	0,00
4	0,00	0,00	0,00	0,50	0,25	0,25
5	0,00	0,25	0,00	0,50	0,00	0,25
...

Зауважимо, що мережний аналіз використовується для реалізації представницького напрямку. Прикладами таких досліджень можуть служити¹¹³.

Зібрані разом, усі зазначені особливості складають одну тенденцію зміни особливостей контент-аналітичних досліджень, яка намітилася в період 1950-1980 рр. і стала наслідком тривалого розвитку методології і теорії контент-аналізу. "Із середини ХХ ст. були підняті заперечення проти поверхового аналізу, який здійснювався без урахування прихованого змісту і контексту, спрощено та з кількісними спотвореннями"¹¹⁴. Контент-аналіз ставав усе більше і більше якісним у міру нагромадження підтверджень того, що одні лише кількісні інтерпретації не в змозі ідентифікувати неявний контекст¹¹⁵.

Більше того, знову на щит підіймається проблема співвідношення між кількісними та якісними дослідженнями. Адже, як заявляють прихильники останнього, узагальнення, що робить дослідник на основі контент-аналізу, зумовлені та обмежені особливостями культурної комунікації. Зокрема, "контент-аналіз не може претендувати на істину у твердженнях

¹¹³ Osgood Ch., Saporta S., Nunnally J. Evaluative Assertion Analysis...– PP.47-102; Carley K. An approach for relating social structure to cognitive structure // Journal of Mathematical Sociology.– 1986.– No 12.– PP. 137-89.

¹¹⁴ Mayring Ph. Qualitative Content Analysis...

¹¹⁵ Content Analysis: An Introduction...

про естетичні якості літературних текстів"¹¹⁶. Він розкриває зміст тексту, але не здатний інтерпретувати значення цього змісту, і тому дослідник повинен вивчати текст безпосередньо. Отже, як стверджує О.Олсті, "контент-аналіз можна розглядати тільки як доповнення, а не заміщення до суб'єктивного дослідження документів"¹¹⁷.

Тенденція повернення через півстоліття до якісних методів після тривалого захоплення кількісними підходами відчувалася не лише в контент-аналізі, а й у соціології загалом. О.Б.Клюшкіна називає три причини того, чому якісні методи знову привернули увагу дослідників:

- теоретико-методологічна передумова ґрунтувалася на тому, що до моменту написання перших робіт з "обґрунтованої теорії" в соціології був актуальним феноменологічний підхід;

- методико-методологічна передумова була пов'язана з тим, що, з одного боку, відбулося нагромадження великого масиву кількісних даних, і на цій основі були усвідомлені межі пізнавальних можливостей опитувальних технік і контент-аналізу текстів; з іншого боку, дослідження з використанням якісної стратегії також сприяли нагромадженню матеріалу для методичних узагальнень;

- процедурна передумова безпосередньо впливала з попередньої, адже пізнавальна обмеженість масових опитувань і контент-аналізу провокувала дослідників на створення нових методик, які б дозволяли досліджувати такі соціальні феномени, які не піддаються кількісному обстеженню¹¹⁸.

Чітке теоретичне обґрунтування якісних підходів дала так звана "обґрунтована теорія" ("grounded theory"), уперше представлена в 1967 р. соціологами Ансельмом Страусом і Барні Глейзером¹¹⁹. Її зміст: виведення теорії з інтерпретації даних на противагу традиційному підходу, коли спочатку висуюються

¹¹⁶ Ньюман Л. Неопросные методы исследования...– С.121.

¹¹⁷ Holsti R. Content analysis // Handbook of Social Psychology, 2-nd ed. Vol. 2 / Ed. by Zindzey Y., Aronson E.– 1968, – P. 596-692.

¹¹⁸ Клюшкіна О.Б. Построение теории на основе качественных данных...– С.92-101.

¹¹⁹ Glaser B.G., Strauss A.L. The discovery of grounded theory.– Chicago: Aldine, 1967.

гіпотези, а потім вони перевіряються. Тобто, побудова теорії в самому процесі досліджень, "розгортання з даних".

Поступово мова починає йти не лише про якісні моменти в методиці контент-аналізу, а про власне якісний контент-аналіз. Ф.Мейрінг сформулював головний напрям розвитку самої процедури якісного контент-аналізу так: "зберігати переваги кількісного контент-аналізу, розвинуті в рамках вивчення комунікацій, використовувати і далі розвивати його якісно-інтерпретаційні кроки"¹²⁰. Особливості якісного контент-аналізу розглянемо в наступному розділі.

У розглядуваний період чимало досліджень присвячуються розробці конкретних методик розв'язання широких класів проблем. До них, зокрема, належить уже згадувана методика К.С.Хелла контент-аналізу мрії.

Ще одним прикладом конкретної методики є кількісно-якісний контент-аналіз преси¹²¹, біля джерел якої на початку 1960 рр. знаходився відомий французький учений, журналіст і політичний діяч Жак Кайзер. Він розробив оригінальну методику статистичного аналізу періодики. У її основі лежить підхід до тестового масиву як до інформаційної системи. Тим самим Ж.Кайзер підготував теоретичну базу наступного поширення соціологічних методів у сферу вивчення всіх нарративних джерел, включаючи епіграфічний та епістолярний матеріал.

У роботі Ж.Кайзера акцентувалася увага на зовнішній формі організації матеріалу: його розміщенні, змісті, оформленні і т.д. Науковець розробив цілий комплекс дослідницьких процедур, які забезпечують повну формалізацію як окремого газетного номера, так і сукупності однотипних періодичних видань. Тим самим Ж.Кайзер сформулював систему, що дозволяє фіксувати розвиток тенденцій у публікаціях ЗМІ.

¹²⁰ Mayring Ph. Qualitative Content Analysis...

¹²¹ Манекин Р.В. Контент-анализ как метод исторического исследования.– <http://manekin.narod.ru/contentM.htm> ; Алексеев А.Н. Метод Жака Кайзера (Из опыта исследований французской ежедневной прессы) // Проблемы современной печати: Сб. ст. / Отв. ред. В.С.Соколов.– Л.: Изд-во Ленинградского ун-та, 1969.– С.76.

Опишемо елементи його методики, використовуючи матеріали¹²².

Результати дослідження Ж.Кайзера були підсумовані в книзі, виданій друком у рік його смерті "Le quotidien francais"¹²³. Автором було досліджено всі 12 щоденних паризьких газет і 88 провінційних. Одиницею дослідження вибрано газету, а головним принципом називався принцип об'єктивності. На кожну газету була заведена окрема картка, у якій дослідник записував таке:

- назва газети і підпис, що йде за назвою;
- розміщення адміністрації і розміщення редакції (якщо воно не збігається з першим);
- періодичність;
- час виходу (ранок, вечір);
- дата виходу першого номера;
- територія розповсюдження;
- тираж;
- ціна;
- розмір;
- максимальна і мінімальна кількість сторінок, кількість колонок на сторінці;
- назва і адреса друкарні;
- кількість публікацій і географія розподілу кожної публікації;
- деякі особливості газети;
- місце зберігання газет (архів);
- редакційні каталоги і картотеки.

Не вдовольняючись цим, Ж.Кайзер складав досьє, яке відображало юридичну і фінансову структуру газети, особливості випуску і розповсюдження, організацію роботи редакції та політичну лінію газети. Усю площу газети науковці поділяють на дві частини: площа, придбана рекламодавцями для своєї реклами, і площа, відведена для матеріалів редакції

¹²² Алексеев А.Н. Метод Жака Кайзера...– С.76; Kostenko N.V., Ivanov V.F. Experience of Content Analysis. Models and Practices: Monograph.– К.: Centre for Free Press, 2005.– 234 p.

¹²³ Kayser J. Le quotidien francais.– Paris: A. Colin, 1963.

(редакційна площа). Співвідношення між ними Ж.Кайзер вважає однією з найважливіших характеристик видання, яке служить показником ступеня залежності газети від рекламодавців.

Редакційна площа поділяється на текст, ілюстрації і заголовки. Вони також визначаються певною пропорцією. Ж.Кайзер запроваджує як одиницю аналізу спеціальну величину, яку називає "редакційна одиниця" (unite redactionnelle) і яка визначає будь-який матеріал, розміщений у газеті. Аналітик класифікує редакційні одиниці за жанрами (інформація, інформація-стаття, стаття, фейлетон, листи читачів, додаткові рубрики), за джерелами (хто автор матеріалу?), за географією, за темами тощо. Поділяючи газету на редакційні одиниці і поєднуючи їх з певними категоріями, можна отримати в підсумку "незаперечні свідчення реальної дійсності".

Наприклад, оцінювання політичної лінії газети здійснюється за формулою:

Важливість матеріалу = Місце + Заголовок + Оформлення

Усі три наведені категорії у свою чергу поділяються на дрібніші. Місце матеріалу, тобто його розміщення, по-перше, визначається сторінкою, на якій він знаходиться, а по-друге, де саме на сторінці він подається. Аналізуючи заголовок, враховуються: – реальні розміри (довжина, ширина, розмір і тип); – відносні розміри (порівняно з іншими заголовками); – зміст (означення, пояснення, вказівка, сенсаційність). Оформлення матеріалу, тобто його дизайн, визначає ілюстрації, тип і структуру (довколишній фон, композиція, розмір, посилання на джерело інформації).

Ж.Кайзер запропонував оцінити кожен показник. Максимум матеріал може набрати 100 балів, у тому числі: за розміщення – 40, заголовок – 40, оформлення – 20 балів. Наприклад, 30 балів дається за розміщення на титульній сторінці, а 10 балів – за анотацію на ній. Розміщення в лівому верхньому кутку – 10 балів, в інших місцях на першій сторінці – по 2 бали, на інших сторінках – по 1 балу. 40 балів, які можна набрати за рахунок заголовка, діляться на: 30 – за реальний розмір і 10 – за відносний розмір.

Свій подальший розвиток "кайзерівський напрям" методології контент-аналізу отримав у роботах Е.Морен. Е.Морен ввела в науковий обіг термін "одиниця інформації" – семантичний блок, зміст якого відповідає на запитання: "Про що йде мова?" Остання обставина уможливила вивчення будь-яких форм організації текстового матеріалу, причому як на термінологічному рівні, так і на рівні фраз, абзаців, статей і навіть цілих книг. Тим самим, Е.Морен відкинула критерій однорідності, який застосовувався раніше під час статистичної обробки нарративів. Замість цього, вона запропонувала ідеологію "семантичних груп", які, на її думку, повинні враховуватися за тематичною ознакою. Крім того, Е.Морен розробила концепцію "тону" матеріалу, яка визначається соціометрично: "позитивна", "негативна", "нейтральна" інформація.

Розроблений Ж.Кайзером метод аналізу преси отримав широке визнання, а деякі способи статистичного опису преси, запропоновані ним, стандартизувала і використала у своїх проєктах ЮНЕСКО.

Саме в період, який розглядається, зароджується комп'ютерний контент-аналіз, який узяв свій відлік з Гарварда, де в 1961 р. група науковців Гарвардського університету в США під керівництвом Ф.Стоуна¹²⁴ розробила основні принципи машинного контент-аналізу. Розвиток машинного контент-аналізу був настільки інтенсивним, що період 1960-1970 рр. стали називати "золотим віком" ("golden age") машинного контент-аналізу. Журнал *Data Based Advisor*¹²⁵ на кінець 1980 рр. нарахував 28 різних програмних продуктів – "менеджерів текстів", які забезпечували широкий діапазон швидкості обробки тексту, індексації текстів і здатності виправляти його. Хоча не всі вказані продукти можна

¹²⁴ Stone Ph., Dunphy D., Smith M., Ogilvie D. *The General Inquirer: A Computer Approach to Content Analysis.*– Cambridge, 1966.

¹²⁵ Perez E. *Managing text // Data Based Advisor.*– 1990.– No 8(6).– PP. 83-93; Morris R. *Computerized content analysis in management research: a demonstration of advantages & limitations // Journal of Management.*– Winter.– 1994.– http://www.findarticles.com/p/articles/mi_m4256/is_n4_v20/ai_16549030

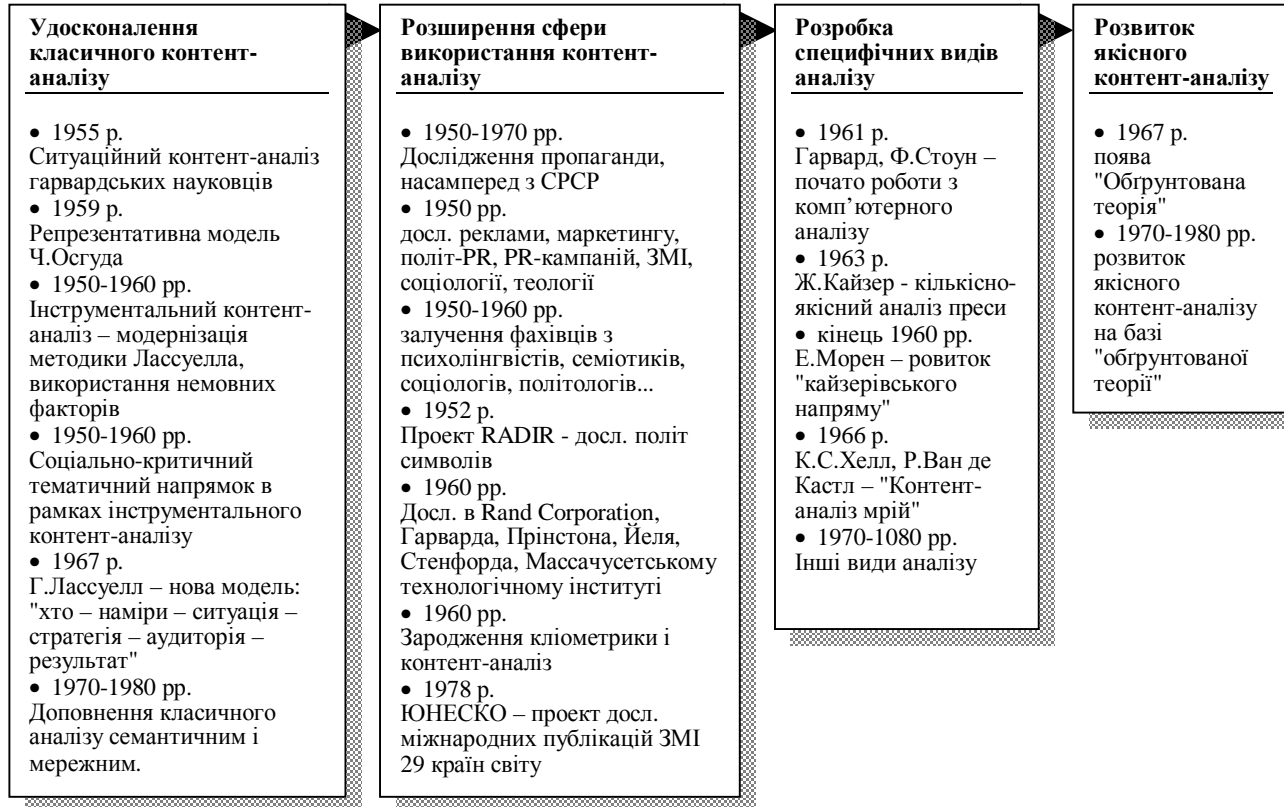


Рис.2.2. Розвиток контент-аналізу в 1950-1980 рр.

однозначно класифікувати як програми контент-аналізу, проте багато з них давали користувачу хороший інструментарій для проведення контент-досліджень.

На завершення відзначимо, що саме в цей період було надруковано чимало навчальних посібників і підручників з контент-аналізу, що сприяло його розумінню та розповсюдженню.

* * *

Таким чином, у 1950-1980 рр. проявляються такі тенденції в розвитку контент-аналізу (рис. 2.2). По-перше, розширюється сфера використання контент-аналізу. По-друге, істотно вдосконалилася його методика, відбувався зворотний процес до того, що мало місце на початку ХХ ст. – повернення до якісних методик. Тепер кількісний аналіз сам поступово набуває рис якісної методики. І, нарешті, у 1960 рр. з появою обчислювальної техніки починаються роботи з машинного (комп'ютерного) контент-аналізу.

Розділ 3. ТЕНДЕНЦІЇ РОЗВИТКУ КОНТЕНТ-АНАЛІЗУ В ПЕРІОД З 1990 ДО 2000 рр.

У період 1990-2000 рр. контент-аналіз продовжував розвиватися в рамках тих же тенденцій, які намітилися в попередні десятиліття. Однак, якщо врахувати масштаби застосування та зміни, що мали місце в методиці аналізу, то період з 1990-2000 рр. заслуговує на особливо уважний розгляд, оскільки саме в цей час контент-аналіз перетворився в технологію загального користування.

3.1. Статистика і основні тенденції розвитку контент-аналізу

За підрахунками О.Ольсті, у перші два десятиліття ХХ ст. в середньому проводилося 2,5 контент-досліджень за рік. У 1920 рр. їхня кількість зросла до 13,3, в 1930 рр. – до 22,8, у 1940 рр. – до 43,4 на рік, а в кінці 1950 рр. проводилося щорічно вже понад 100 контент-аналітичних досліджень¹²⁶. Г.Бернард і Г.Раян у середині 1997 р. здійснили пошук і виявили понад 500 книг, у тому числі 200, написаних до 1990 р., у заголовку яких містилися фрази "якісний аналіз даних" і "аналіз тексту", а також майже тисячу журнальних статей, присвячених аналізу якісних даних¹²⁷. Ось назви журналів, які спеціалізуються винятково на якісних дослідженнях: "Qualitative Sociology", "Qualitative Inquiry", "Journal of Contemporary Ethnography", "Symbolic Interaction", "Qualitative Health Research", "Quality and Quantity", "Studies in Qualitative Methodology", and "International

¹²⁶ Holsti O.R. Content analysis for the social sciences and humanities...

¹²⁷ Bernard H.R., Ryan G. Text Analysis: Qualitative and Quantitative Methods // Handbook of Methods in Cultural Anthropology / H. Russell Bernard (ed).– Walnut Creek: Altamira Press, 1998.– P.627.–
nersp.nerdc.ufl.edu/~ufruss/documents/bernard_ryan.pdf

Journal of Qualitative Studies in Education", "Cultural Anthropology Methods Journal" та інші.

Для вивчення динаміки розвитку контент-аналізу за період 1990-2000 рр. ми пішли аналогічним шляхом і спробували проаналізувати тематику публікацій з контент-аналізу. Для цього було використано досить об'ємну (понад 900 одиниць) бібліографію, складену К.Нойєндоф станом на липень 2000 р.¹²⁸ Даний список робіт, зрозуміло, не може претендувати на цілковиту повноту (насамперед це стосується досліджень періоду зародження і класичного контент-аналізу, а також, з огляду на час складання бібліографії, кінця 1990 рр. і 2000 р.), однак уявлення про загальні тенденції в розвитку контент-аналізу на його основі отримати можна.

Аналіз кількості публікацій у зазначеній бібліографії в розрізі років показує (рис. 3.1), що дослідження, присвячені контент-аналізу та його використанню, лавиноподібно зростають. Особливо це помітно від середини 1990 рр.

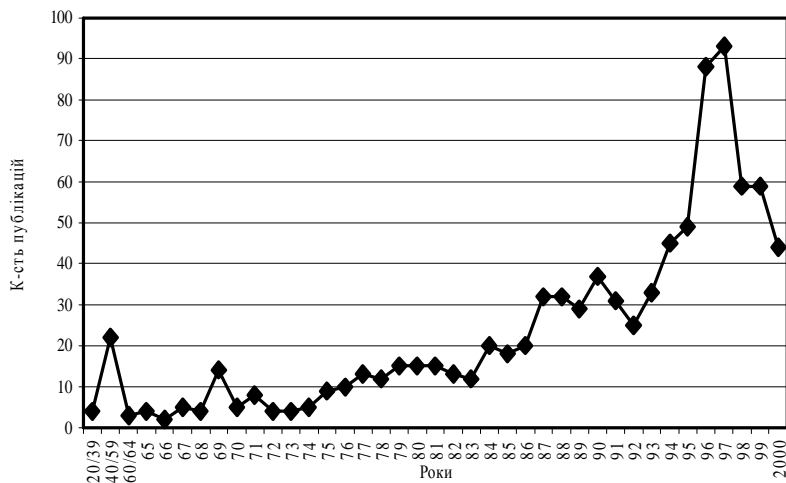


Рис. 3.1. Динаміка кількості публікацій з контент-аналізу

¹²⁸ Neuendorf K.A. Content Analysis Bibliography, 7.7.2000 // Neuendorf K.A. The Content Analysis Guidebook Online . – <http://academic.csuohio.edu/kneuendorf/content/bibs/bookbib.htm>

Тематичний аналіз динаміки публікацій дозволяє виявити якісні тенденції в розвитку контент-досліджень (табл. 3.1). Найпомітнішою з них те, що частка публікацій, присвячених методиці контент-аналізу, зменшилася більше, ніж 2 рази, тоді як частка публікацій щодо практики його використання і комп'ютерного аналізу зросла більше, ніж удвічі, а досліджень особливостей різноманітних носіїв змісту – понад 7 разів.

Однак, детальніше розглядаючи публікації, присвячені методиці, можна помітити, що різке зниження їх кількості мало місце лише до кінця 1980 рр., а починаючи з початку 1990 рр. увага до методики побільшала знову, що, на нашу думку, немалою мірою пов'язано з розвитком якісного контент-аналізу, який у 1980 рр. отримав додаткові стимули у вигляді відповідних комп'ютерних програм.

Характеризуючи сфери застосування, слід одразу зауважити, що майже протягом усього періоду, починаючи від початку 1970 рр., більше половини публікацій були присвячені конкретним тематичним дослідженням. Так, завжди увагу контент-аналізу привертала психологічна тематика, міжособистісні та публічні комунікації. Від середини 1970 рр. контент-аналіз активно використовувався в рекламі та маркетингу, а, починаючи з другої половини 1980 рр., після деякого спаду, знову зростає число досліджень гуманітарної спрямованості та на освітянську тематику.

Постійно увага контент-аналізу була повернута до вивчення особливостей різноманітних носіїв змісту. Від 1960 рр. найбільшу цікавість у дослідників викликали електронні медіа – телебачення і радіо. Структурна динаміка і зростання масштабів їхнього розвитку, а також зростання впливу на протікання суспільних процесів, виливалося у відповідні контент-дослідження. Як не парадоксально, а проявляється ще одна закономірність: не зменшується, а навпаки – зростає увага до аналізу друкованих і рукописних текстів. Після 1990 рр. Інтернет і його змістовне наповнення також стало об'єктом численних досліджень контент-аналітиків.

Поява комп'ютерного контент-аналізу вимагала розробки його власної методики. Тому, починаючи з 1960 рр., ця

Таблиця 3.1

Тематична структура зарубіжних публікацій з методики і використання контент-аналізу (в %)

Тематика публікацій	Роки						
	1965-1969	1970-1974	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999
1. Методика контент-аналізу у тому числі:	71	57	45	43	21	33	34
– якісного контент-аналізу	0	0	0	0	1	3	4
2. Сфера досліджень у тому числі:	21	57	55	37	54	58	57
– міжособистісні / публічні комунікації	0	7	13	8	10	6	9
– у гуманітарних науках	0	29	18	10	5	12	13
– в освіті	0	7	3	0	1	3	8
– у рекламі / маркетингу	0	0	3	8	17	21	10
– психологія	21	14	20	12	21	16	18
3. Джерела досліджуваних текстів у тому числі:	7	21	20	41	47	44	52

Тематика публікацій	Роки						
	1965-1969	1970-1974	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999
– друківані видання	0	0	3	8	14	15	14
– кінофільми	0	7	0	2	2	3	5
– аудіо / відео	7	14	18	29	31	20	24
– Інтернет	0	0	0	2	0	5	10
4. Комп'ютерний контент-аналіз у тому числі:	7	7	10	6	6	16	18
– комп'ютерний аудіо / відео контент-аналіз	0	0	3	0	0	1	4

Джерело: Обчислення проведені на основі бібліографії, представленої К.Нойєндорф в ¹²⁹.

Примітка: Сума відсотків за окремими підрубриками не обов'язково дорівнює загальному результату по рубриці, оскільки одна і та ж публікація може представляти різні підрубрики. Наприклад, в одній публікації може розглядатися питання загальної методики контент-аналізу, методики комп'ютерного аналізу, а також наводитися приклади вивчення матеріалів з різних носіїв інформації, наприклад, преси і телебачення.

¹²⁹ Neuendorf K.A. Content Analysis Bibliography...

проблематика завжди знаходила висвітлення в публікаціях, і в останні роки їхнє число істотно зросло.

Такою є кількісна характеристика розвитку контент-аналізу в період 1990 рр. Розглянемо його на якісному рівні.

3.2. Основні цілі і напрями використання контент-аналізу

Ще від часу класичного контент-аналізу його застосування вийшло далеко за межі звичайного опису тексту. Б.Берельсон¹³⁰ сформулював сімнадцять цілей, які можуть бути досягнуті з допомогою контент-аналізу. Серед них:

- опис тенденцій при зміні змісту комунікативних процесів;
- опис розходжень у змісті комунікативних процесів для різних країн;
- порівняння різних ЗМІ;
- виявлення використовуваних пропагандистських прийомів;
- визначення намірів і інших характеристик учасників комунікації;
- визначення психологічного стану індивідів і/або груп;
- виявлення настанов, інтересів і цінностей (ширше – систем переконань і "моделей світу") різних груп населення і суспільних інститутів;
- виявлення фокусів уваги індивідів, груп і соціальних інститутів.

Дещо пізніше О.Ольсті виділяє сім запитань, на які намагається відповісти кількісний контент-аналіз:

- опис тенденцій (на основі змісту тексту);
- виявлення взаємозв'язків між характеристиками і повідомленням тексту;
- оцінка відповідності стандартам комунікацій;
- аналіз методів переконань (пропаганди);
- аналіз стилю;

¹³⁰ Berelson B. Content Analysis in Communication Research...

- виявлення взаємозв'язків між відомими характеристиками аудиторії і повідомленнями, підготовленими для неї;
- опис зразків (шаблонів) комунікації¹³¹.

К.Нойендорф указує на чотири головні ролі контент-аналізу:

- описова;
- формування висновків;
- психометрична;
- прогнозна¹³².

Якщо психометрична роль тісно пов'язана з медициною та психоаналізом і використовується для інтерпретації мови пацієнтів та інтерв'ю з ними, то три інші цілком підходять для використання в суспільних науках і сфері культури. Описова роль, зрозуміло, є основною і забезпечує розуміння повідомлень і зображень, представлених у ЗМІ. Формування висновків і прогнозна роль служать "помічниками" для остаточних висновків з досліджуваної проблеми завдяки тому, що дозволяють виробляти і оцінювати вірогідні ефекти текстів на певну аудиторію і суспільство в цілому.

Наведемо приклад ролі контент-аналізу у випадку аналізу ЗМІ. Як відомо, ЗМІ мають самі здатність впливати на громадську думку. Водночас вони володіють здатністю відображати й інші точки зору завдяки інформаційним матеріалам, показуючи, що люди, компанії чи організації говорять і роблять. Крім того, публікації ЗМІ можуть у такий спосіб подавати повідомлення, що часто абсолютно змінюється зміст подій та їхній порядок денний. У цьому світлі контент-аналіз ЗМІ забезпечує дві важливі функції (рис. 3.2):

1) оцінювання ступеня ефективності комунікацій організації (PR) з цільовою аудиторією через використання медіа, публікацій повідомлень, позиціонування щодо конкурентів або в сегменті конкурентів;

¹³¹ Holsti O.R. Content analysis for the social sciences and humanities...

¹³² Neuendorf K.A. The Content Analysis Guidebook.– Thousand Oaks, CA: Sage Publications, 2002.– P.53.

2) формування стратегічного бачення й аналітичного осмислення, здійснюючи моніторинг випусків видання або конкурентний аналіз.

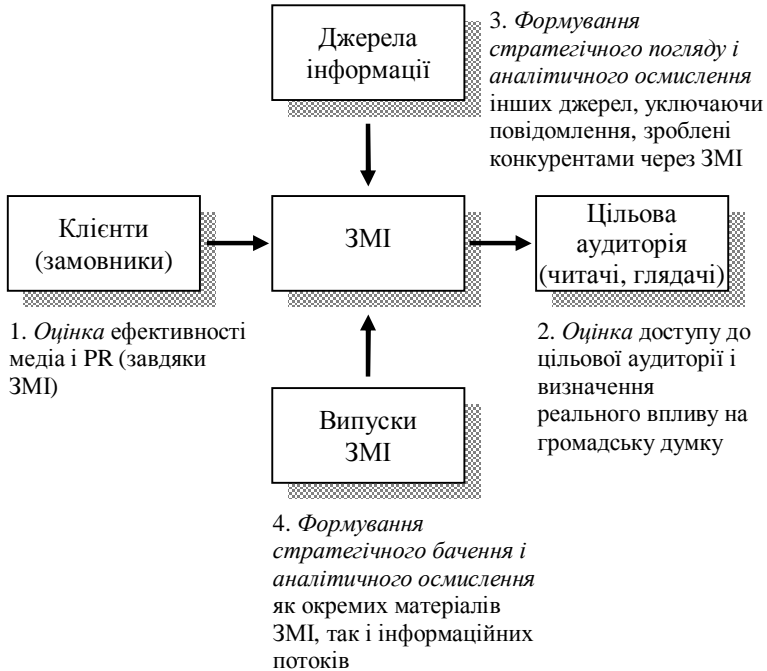


Рис. 3.2. Схема формування нових знань у Text Mining

Нині, не зважаючи на дебати про ефективність винятково описового контент-аналізу за змістом, контент-аналіз широко використовується як дослідницька техніка з широким полем застосувань. Виходячи з аналізу конкретних прикладів, Р.Уіммер і Дж.Домінік¹³³ ідентифікують п'ять основних цілей контент-аналізу:

- опис зразків або тенденцій у медіавідображеннях;
- перевірка гіпотез щодо політики або цілей медіа-виробників;

¹³³ Wimmer R.D., Dominick J.R. Mass Media Research: An Introduction. – Belmont, CA: Wadsworth Publishing Company, 1994.

- порівняння змісту медіа з реальним світом;
- оцінка представлення специфічних груп населення в суспільстві;
- формування висновків щодо ефектів медіа¹³⁴.

З огляду на вказані цілі, проаналізуємо дослідження, що проводилися в період 1990-2000 рр.

Опис зразків і тенденцій стосуються найрізноманітніших процесів і явищ. Однак виділилися напрями, до яких привертається найбільша увага дослідників. Серед них описи: медіа-насильства та психології його сприйняття; сексуального медіа-змісту і зв'язку із сексуальними злочинами; жахів і тривоги, викликаних медіа-продукцією; зображення національних меншин, міжрасових і гендерних стосунків; інформаційних кампаній і політичної пропаганди. Останнім часом набувають важливості теми, пов'язані з використанням нових інформаційних технологій, зокрема Інтернет.

Тема, до якої, мабуть, найуважніше приглядаються за останні півстоліття, – це насильство, показане на телеекранах. Перші дослідження припадають на 1950 рр. (див. ¹³⁵). Контент-аналіз насильства на телебаченні з кінця 1960 рр. до середини 1980 рр., який проводився Пенсільванському університеті, вивчав відверті прояви використання фізичної сили проти себе або інших¹³⁶. Подальше вивчення, окрім зазначеного, бере до уваги вже спеціально підготовлені масштабні або жахливі нещасні випадки та природні катастрофи¹³⁷. Якщо канадське дослідження¹³⁸ кінця 1980 рр. дещо ліберальніше ставиться до поняття насильства, включаючи до розгляду лише

¹³⁴ Department of Journalism Studies...

¹³⁵ Smythe D. Three Years of New York Television: 1951-1953. – Urbana, IL: National Association of Educational Broadcasters, 1954.

¹³⁶ Gerbner G. Violence in television drama: trends and symbolic functions // Television and Social Behavior. Vol. 1: Media Content and Control / Comstock G.A., Rubinstein E.A. (eds.). – Washington, DC: US Government printing office, 1972. – P.31.

¹³⁷ Gerbner G., Gross, L. Living with television: the violence profile // Journal of Communication.. –1976. – No 26. – PP.173-199.

¹³⁸ Williams T.M., Zabrack M.L., Joy L.A. The portrayal of aggression on North American television // Journal of Applied Social Psychology. – 1982. – No 12 (5). – PP.360-380.

правдоподібні і вірогідні дії та виключаючи пустопорожні погрози, лихослів'я, комічні жести без фактичних наслідків, то фінське вивчення¹³⁹ теленасильства 1997 р. навпаки: посилює вимоги до нього, беручи до уваги неявні погрози, вибухи гніву та інші види загрозливої поведінки, які призводять до негативних наслідків не лише щодо людей, але й тварин і неживих об'єктів.

У масштабах держави Національною асоціацією кабельного телебачення США проводилося дослідження телевізійного насильства¹⁴⁰ протягом трьох років 1994-1997 рр. Його методика передбачала не лише отримання оцінки чисельності насильств, але дослідження їхньої природи та контексту здійснення. Поняття насильство втілювалося в трьох категоріях: вірогідна загроза насильства, випадки відверто насильницької поведінки і шкідливі наслідки прихованого насильства. Було також представлено три рівні вимірювань:

– РАТ (взаємодія між такими елементами, як порушник "Р", дія "А" і мета "Т");

– місце дії;

– план або програма дій.

Розглянуті нами приклади складають лише незначну частину матеріалів з даної тематики, детальніше з ними та іншими дослідженнями телевізійного насильства можна ознайомитися в¹⁴¹.

Чимало контент-досліджень виходять за межі простого опису ситуації і намагаються за допомогою аналізу змісту робити висновки щодо прихованих матеріальних характеристик,

¹³⁹ Mustonen A., Pulkkinen L. Television violence: A development of a coding scheme // *Journal of Broadcasting and Electronic Media*. – 1997. – No 41. – PP.168-189.

¹⁴⁰ Potter W.J., Linz D., Wilson B.J., Kunkel D., Donnerstein E., Smith S.L., Blumenthal E., Gray T. Content Analysis of Entertainment Television: New Methodological Developments // *Duke University Conference on Media Violence and Public Policy in the Media*, Durham, NC 27-29, June, 1996; Wilson B.J., Smith S.L., Linz D., Potter J., Donnerstein E., Kunkel D., Blumenthal E., Gray T. Content Analysis of Entertainment Television: The 1994-95 Results // *Duke University Conference on Media Violence and Public Policy in the Media*, Durham, NC, 27-29 June, 1996.

¹⁴¹ Department of Journalism Studies...

мотивів, ідеології продукції або організації¹⁴², перевіряючи різноманітні гіпотези. Наприклад, що стосується медіавпливів, часто контент-аналіз фокусується на класифікації та визначенні домінуючих особливостей медіановин, формуючи такий порядок денний, який забезпечує потрібні інформаційні впливи. Так, вивчаючи порядок денний газет, С.Лейсі, Ф.Фіко і Т.Саймон¹⁴³ з допомогою контент-аналізу підтвердили гіпотезу, що престижні газети частіше, ніж масова преса, подають різні точки зору на одну проблему.

Ще один приклад: з допомогою контент-аналізу Б.Ітон і Дж.Домінік¹⁴⁴ намагалися ідентифікувати приховані стратегії маркетингових кампаній організацій. Виявилось, що анімаційні фільми, у яких використовуються іграшкові символи та персонажі, якими торгують на ринку, включають більше жбурлянь, кидків і жорсткості, ніж ті мультфільми, персонажі яких не є іграшками.

Контент-аналіз проводиться також і для з'ясування, наскільки медіа-описи подій, об'єктів і явищ відповідають реальному життю. Прикладом може служити дослідження, проведене у США (штат Невада) в рамках проекту PROMISE (Projects for Multicultural and Interdisciplinary Study and Education – Проекти мультикультурних та міждисциплінарних досліджень та освіти)¹⁴⁵. Аналізувалися підручники з геології на предмет виявлення закладеного в межах тексту очевидного і прихованого змісту, що стосувався таких тем:

- гендерні особливості тексту;
- особливості відображення расової й етнічної належності;
- мультикультурність;
- ставлення до військово-промислового комплексу;
- соціальна і наукова інтеграція.

¹⁴² Department of Journalism Studies...

¹⁴³ Lacy S., Fico F., Simon T.F. Fairness and balance in the prestige press // Journalism Quarterly. – 1991. – No 68(3). – PP.363-370.

¹⁴⁴ Eaton B.C., Dominick J.R. Product related programming and children's TV // Journalism Quarterly. – 1991. – No 18. – PP.67-75.

¹⁴⁵ Content Analysis of Introductory Geology Textbooks / Projects for Multicultural and Interdisciplinary Study and Education.–

<http://www.scsv.nevada.edu/~promise/content.html>

Виявилось, що 94,12% підручників написані чоловіками, чоловіки на 94,57% внесені до списку засновників і новаторів геології. У підручниках чоловіки представлені на 72,09% фотографій і 94,74% діаграм і малюнків. Із 528 фотографій лише 8,25% містили неєвропейські форми облич. Десять з дванадцяти чорношкірих людей у всіх текстах були на одній фотографії, яка ілюструвала проблеми перенаселення. Інше відкриття: візуальне представлення показує, що неєвропейці винятково безпорадні в стосунках з природою. Такий же висновок випливає з того, що науковці на фотографіях майже завжди білі чоловіки, а небілі люди потребують допомоги "великої ділової надії науки". Наведені висновки свідчать про свідоме переключення реального стану речей.

Дуже тісно з попереднім переплітаються оцінки зображення певних груп населення в ЗМІ. Тут частота появи в світі ЗМІ людей та їхній образ порівнюється з тим, що має місце насправді. Л.Нейман в ¹⁴⁶ наводить приклад дослідження, у якому вивчалася частота негативного зображення літніх людей у телевізійних виставах. Результатом концептуалізації став перелік стереотипних характеристик або негативних уявлень про літніх людей: фізична безпомічність, забудькуватість, дивакуватість, глухота, уповільненість дій, нездатність доглядати за собою, неактивність, консервативність. Однак цей список не зовсім правдиво описує цю категорію людей. Так, хоча серед осіб віком понад 65 років лише 5% потребують догляду, у телепередачах 50% з них зображаються нездатними доглядати за собою, що свідчить про наявність негативної стереотипізації.

¹⁴⁶ Ньюман Л. Неопросные методы исследования... – С.122.

3.3. Дискусії довкола переваг і недоліків контент-аналізу

Сучасний контент-аналіз піддав перегляду властивості, приписані класичному контент-аналізу як особливі переваги ще в 1952 р. Б.Берельсоном: об'єктивність, систематичність, кількісний характер, дослідження змісту винятково явних комунікацій¹⁴⁷. Ці властивості в різних варіаціях відтворювалися науковцями протягом кількох наступних десятиліть (динаміку зміни означень можна прослідкувати, наприклад, за роботами¹⁴⁸. Разом з тим усі вони стали об'єктом критики і перегляду з боку як опонентів, так і прихильників контент-аналізу.

Основна претензія до контент-аналізу стосувалася того, що він досліджує явно виражені ознаки/параметри в тексті, а це істотно звужує пізнавальні можливості методу і фактично зводить його до описового рівня. На думку багатьох дослідників, завдання контент-аналізу має полягати в спробах розкрити латентний (прихований) зміст комунікації через вивчення проявів цього змісту в структурі тексту.

Наступне, що закидають контент-аналізу, – це втрата істинного контексту і, як наслідок, принципова неможливість кількісно виміряти роль комунікації в ньому. Цей недолік, на думку критиків, лежить у самих методологічних засадах методу (наприклад, див.¹⁴⁹). Найчастіше наводиться такий аргумент: "класичний контент-аналіз зіпсований його довірою до проблемної моделі мови, яка розглядається як сукупність

¹⁴⁷ Berelson B. Content Analysis in Communication Research...– P.18.

¹⁴⁸ Yahya H. Content analysis in human communication / JURNAL IAIN Sunan Ampel. Edisi XVII, Januari-Maret 2000. – <http://www.geocities.com/HotSprings/6774/j-41.html> ; Department of Journalism Studies...

¹⁴⁹ Yahya H. Content analysis in human communication...; Jackson J. Content analysis and objectivity, language and metaphor.– 1998. – <http://www.spinworks.demon.co.uk/pub/content2.htm> ; Mustchin E. Explore the strengths and weaknesses of classical content analysis.– 1998. – <http://www.spinworks.demon.co.uk/pub/content1.htm>

конкретних квантифікованих змінних"¹⁵⁰. З точки зору системного підходу, в результаті контент-аналізу текст як система ("лінійність" тексту) руйнується і перетворюється в сукупність окремих складових, у ролі яких виступають категорії та підкатегорії аналізу. Подальша логіка методики зводиться до виділення тих понять, "які домінують у словнику, також домінують і в контексті"¹⁵¹. Статистичний аналіз отриманих частотних таблиць ще більше спрощує ситуацію і віддаляє від реальності. Іншими словами, метод розуміє мову надто спрощено: вона уподібнюється словнику, у якому представлені лише слова, ізольовані від контексту. Витягнуті з цілого і проаналізовані окремо, елементи тексту не дають повноцінної картини. Навіть якщо взаємозв'язок представлений, то він виражає лише зв'язок між словами даного "словника". Відтак спроби зрозуміти сутність комунікацій фактично зводяться до класифікації слів/понять, не прив'язаних до контексту. Така ізоляція контент-аналізу "обмежує його здатність служити соціальним індикатором"¹⁵². На думку Дж.Джексона, навіть створення індексів слів або ключових слів у контексті (KWIC) не допомагає у випадку складного аналізу¹⁵³.

Третій аргумент висувається проти наступної максими контент-аналізу: частота появи понять фіксує певний феномен, ступінь його важливості. Хоча, як правило, вагомі елементи або зразки є "алергічними" на кількісний аналіз очевидного змісту, проте частота категорії та важливість теми для автора не завжди пов'язані лінійно. Наприклад, "при розгляді болючої теми лише її одноразове згадування може дати більше для роздумів, ніж згадування шістнадцять разів порівняно з п'ятнадцятьма. З іншого боку, незгадування теми взагалі може бути таким же істотним, як і згадування десятки разів"¹⁵⁴. Одна-єдина фраза може вплинути на думку набагато сильніше, ніж багатослівне обговорення: хоча заголовки є сингулярними, проте саме вони

¹⁵⁰ Mustchin E. Explore the strengths and weaknesses of classical content analysis...

¹⁵¹ Mustchin E. Explore the strengths and weaknesses of classical content analysis...

¹⁵² Mustchin E. Explore the strengths and weaknesses of classical content analysis...

¹⁵³ Jackson J. Content analysis and objectivity, language and metaphor...

¹⁵⁴ Yahya H. Content analysis in human communication...

найбільше привертають увагу¹⁵⁵. Тому частотний аналіз треба доповнювати іншими прийомами, які б дозволяли виявити важливі, але приховані мотиви, впливи, елементи соціальної дійсності.

Ще одна проблема, яка підіймається на щит противниками кількісного контент-аналізу, – це заперечення об'єктивного характеру результатів. Чимало дослідників відносять контент-аналіз до категорії "нереактивних" методів, зокрема Л.Нейман в¹⁵⁶, або до методів з обмеженою "реактивністю", наприклад, в¹⁵⁷. Тлумачний словник англійської мови під словом "reactive" розуміє таку властивість, яка являє собою відповідь або реакцію на щось, наприклад, позитивна/негативна відповідь на пропозиції інших людей. Зміст поняття "реактивний" стосовно методів досліджень зводиться до того, що люди, які виступають у ролі об'єкта дослідження, знають про це і тому можуть змінити свою поведінку і свідомо чи несвідомо спотворити результати. Тобто "реактивність" стає одним із джерел необ'єктивності.

Зрештою, уже й самі прихильники кількісного контент-аналізу зрозуміли марність зазіхань на абсолютну об'єктивність. Їхній підхід до оцінки ступеня об'єктивності результатів став більш зваженим, і відтак саме звідси йде чимало вдосконалень методики контент-аналізу. Наведемо логіку міркувань прихильників контент-аналізу щодо механізму внесення суб'єктивізму в методику контент-аналізу й оцінимо, наскільки він "забруднений реактивністю"¹⁵⁸.

Досліджувані контент-аналізом тексти, як правило, виникають у процесі вивчення соціальної дійсності, наприклад, через спостереження, інтерв'ю тощо. Фактично відбувається трансформація соціального середовища в його символічне або

¹⁵⁵ Mustchin E. Explore the strengths and weaknesses of classical content analysis...

¹⁵⁶ Ньюман Л. Неопросные методы исследования...– С.119-130; в цьому перекладі розділу книги автора термін "nonreactive" замінено не зовсім вдалим терміном "неопросный", який не лише не відображає задуму автора, а швидше спотворює його.

¹⁵⁷ Gallagher C, Marken M., Kim M., Phillipson T., Dodge T. The Use of Content Analysis in Researching Adolescent Issues...

¹⁵⁸ Department of Journalism Studies...

текстове представлення. Таке представлення зберігає доречні й істотні елементи соціальної дійсності, щоправда, в іншій формі, яка, однак, не є довершеною. Іншими словами, соціальна дійсність переноситься в альтернативну символічну дійсність, представлену текстом. На наступному етапі подальше її опрацювання полягає в перетворенні тексту в дані, для чого, зокрема, застосовується контент-аналіз. А вже дані служать основою для вироблення висновків.

Реактивні впливи (суб'єктивізм) на першому етапі, викликані присутністю дослідників та інтерв'юєрів, є загальновідомими, і науковцями розроблено чимало прийомів для боротьби з ними.

Щодо контент-аналізу, то реактивність властива і йому: текст також є соціальною дійсністю, до якої також прикладається суб'єктивна "рука" людини. Проблема реактивності в процесі такої трансформації ідентифікується на трьох рівнях¹⁵⁹:

– синтаксичному, де символ (S) визначається через встановлення співвідношення між ним і засобами його представлення (M);

– семантичному, де визначення символу (S) здійснюється через засоби його представлення (M) і посилання на об'єкт (O);

– прагматичному, що визначає трифакторне співвідношення, у якому до вказаних двох елементів долучається людина – інтерпретатор символу (I): $S = (M, O, I)$.

Відтак, текст як соціальна дійсність може бути реактивним саме завдяки впливу фактора (I), наприклад, якщо параметри тексту заважають тому, хто інтерпретує його. На кожному рівні аналізу по-своєму відбувається взаємодія між змістом тексту і людиною, зокрема, кодувальником. Схематичне уявлення цієї взаємодії, зроблене К.Мертеном¹⁶⁰, представлено в табл. 3.2.

¹⁵⁹ Department of Journalism Studies...

¹⁶⁰ Merten K. Reactivity in content analysis // Communications: The European Journal of Communication Research.– No 21(1).– 1996.– P.69.

Таблиця 3.2

Схематичне представлення перспектив аналізу тексту

Рівень тексту	Рівень розгляду тексту людиною (кодувальником, читачем)		
	синтаксичний	семантичний	прагматичний
синтаксичний	1. Синтаксис	2. Синтаксична семантика	4. Синтаксична прагматика
семантичний		3. Денотативна семантика	5. Конотативна прагматика
прагматичний			6. Прагматична прагматика

У клітині 1 подані синтаксичні методи контент-аналізу, наприклад, частотний аналіз літер, слів або речень. Клітина 2 містить усі види контент-аналізу на синтаксичній основі, які вимагають урахування семантичних правил, тобто аналізу слів. Клітина 3 представляє класичний рівень. Тут контент-аналіз винятково семантичний: правила кодування вимагають лише семантичного ототожнення тексту (теми) або частини тексту (ключові слова) із семантичним терміном або категорією. Найзагальнішим видом аналізу в цьому випадку виступає аналіз тем. З перерахованих видів аналізу жоден не реактивний, якщо не враховувати випадкових, чисто механічних помилок кодування.

Цілком інша ситуація має місце на прагматичному рівні. У клітині 4 представлено аналіз використання синтаксичних структур читачами, тобто тими, на кого спрямований текст, наприклад, аналіз читабельності, що базується на обчисленні розміру речень. Результат істотно залежить від читача, тому подібного сорту процедури реактивні. Клітина 5 уключає контент-аналіз, який вимагає від кодувальників комунікативної компетентності для ідентифікації конотативних значень, наприклад, аналіз із використанням "семантичного диференціалу" Ч.Осгуда. Така ідентифікація можлива лише за умови активної ролі кодувальника, який використовує свої

компетентність і світогляд до розуміння контексту, що стає джерелом реактивності.

Клітина 6 включає винятково прагматичні типи аналізу, пов'язані з інтенційними намірами комунікатора тексту: створити ефект на реципієнтів. Ці типи аналізу вимагають від кодувальника в певному розумінні стати учасником комунікаційного процесу, каталізуючи отримання інформації про наміри та ефекти. Будь-які аналізи ефектів комунікації належать до даного типу, тому саме вони і є найбільш реактивними.

Урахування зазначених і інших недоліків відобразилося, насамперед, на еволюції означення контент-аналізу. Одне з недавніх тлумачень кількісного контент-аналізу у формулюванні Д.Ріффе, С.Лейсі і Ф.Фіко подає його як "систематичне і відтворюване дослідження символів комунікації, яким призначаються певні числові величини відповідно до обґрунтованих правил вимірювань, і аналіз із допомогою статистичних методів взаємозв'язків між цими величинами з метою опису комунікацій та отримання висновків щодо значень або виявлення взаємозв'язків з контекстом, як у процесі їх створення, так і використання"¹⁶¹.

Як бачимо, при такому означенні вимоги систематичності і квантифікації залишилися. Систематичність і далі визначає необхідність чітко, крок за кроком дотримуватися наперед визначених правил. При цьому передбачається, що прозорість і зрозумілість правил сприяє можливості відтворення в такий же спосіб усієї процедури аналізу іншими дослідниками, науковій вірогідності результатів і відсутності впливу суб'єктивних факторів, які можуть бути привнесені дослідниками. Квантифікація пов'язана з вимірюванням комунікацій і використанням отриманих числових характеристик не лише для автономного оцінювання змісту комунікацій, а й з урахуванням контексту. Більше того, мова йде як про очевидні, так і приховані символи комунікації. В означенні також

¹⁶¹ Riffe D., Lacy S., Fico F.G. Analyzing Media Messages: Using Quantitative Content Analysis in Research.– Mahwah, NJ: Lawrence Erlbaum Associates, 1998.– P.20.

безпосередньо вказується ціль дослідження: формування висновків щодо комунікацій у рамках визначеного контексту, тобто певної оцінки елементів соціуму. Підставою для таких висновків служать результати квантування символів комунікації та взаємозв'язку з контекстом, отримані при використанні, насамперед статистичних методів. Саме завдяки таким властивостям контент-аналіз, як стверджує Д.Ліст, "обслуговує дві цілі: видаляє наліт суб'єктивності з даних і спрощує виявлення тенденцій"¹⁶². Ця ж думка підтверджується Е.Мастчіном: "Головною силою контент-аналізу є його здатність моделювати довгострокові та великомасштабні дані задля вивчення змін, які мають місце в них, і специфіки доміант дискурсу"¹⁶³.

Як ми побачили вище, уже за своїм змістом не всі види контент-аналізу є об'єктивними, і суб'єктивність з'являється як у процесі кодування та здійснення вибірки, так і читання текстів. Запровадження жорстких правил проведення аналізу може зменшити (і то лише до певної міри) реактивність контент-аналізу. Тому на порядок денний виноситься необхідність розробки системи заходів, які б забезпечували максимальну надійність результатів і валідність контент-аналізу.

3.4. Особливості методики сучасного кількісного контент-аналізу та її вдосконалення

Логіку кількісного контент-аналізу наочно демонструє модель, запропонована Е.Мергенталером в¹⁶⁴. В основі цієї логіки лежить статистична перевірка наперед визначених дослідником гіпотез.

¹⁶² List D. Know Your Audience. Chapter 16. Content analysis. – August 2003.– <http://www.audienceialogue.org/kya.html>

¹⁶³ Mustchin E. Explore the strengths and weaknesses of classical content analysis...

¹⁶⁴ Mergenthaler E. Computer-Assisted Content Analysis // Text Analysis and Computers / Cornelia Zuell, Janet Harkness, Juergen H.P. Hoffmeyer-Zlotnik (Eds.). Zentrum für Umfragen Methoden und Analysen (ZUMA).– Mannheim (Germany): ZUMA, 1996.– P.5.

Е.Мергенталер розділяє предметну область на дві системи – реальну і формальну. У реальній системі функціонує природна мова, формальна система використовує формальну мову. У свою чергу в реальній системі виділяються об'єктно-лінгвістичні та мета-лінгвістичні компоненти. Будь-який текст можна інтерпретувати як об'єктно-лінгвістичну реалізацію мови в реальній системі. Провідна теорія (або гіпотеза), застосовувана для аналізу тексту, виступає мета-лінгвістичною компонентою. Формальна система охоплює систему категорій без будь-якої подальшої диференціації.

Процедура звичайного ручного контент-аналізу тексту за такої інтерпретації може виглядати як виконання трьох наступних кроків (рис. 3.3):

- 1) перетворення (переклад) шляхом кодування тексту реальної системи в систему категорій формальної системи;
- 2) інтерпретація системи формальних категорій у рамках теорії;
- 3) оцінка і перевірка отриманих результатів співвідносно досліджуваного тексту.

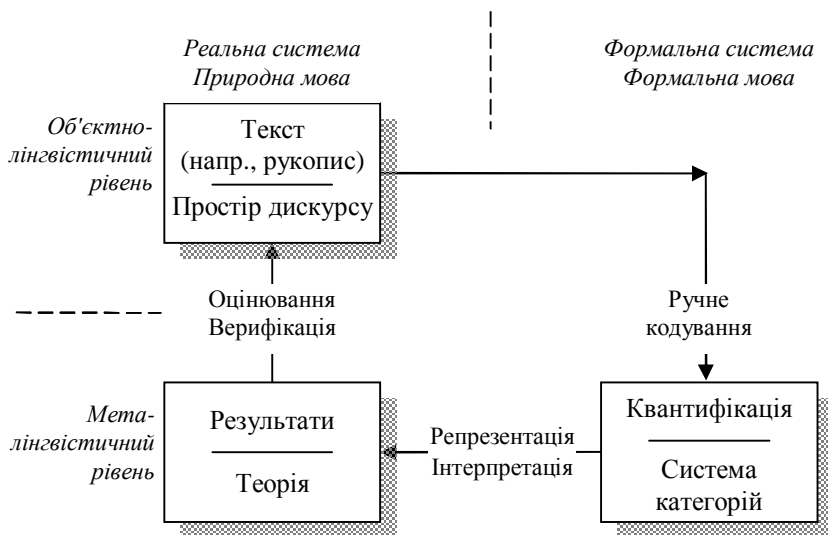


Рис. 3.3. Модель логіки ручного контент-аналізу

Для комп'ютеризованого контент-аналізу критичним є перший крок, який виконується за правилами, реалізованими програмним забезпеченням. Другий крок передбачає процедуру кодування. У рамках моделі це приводить до подальшого диференціювання формальної системи на об'єктно-лінгвістичну та мета-лінгвістичну компоненти. Тим самим установлюється відповідність між текстом і видобутою інформацією (наприклад, сукупність слів), а також теорією і системою категорій відповідно. Відтак процедура аналізу перетворюється в чотирикроковий процес (рис. 3.4):

- 1) скорочення повного тексту до певної вибірки;
- 2) перетворення інформації в систему категорій;
- 3) інтерпретація системи категорій у рамках гіпотези;
- 4) оцінка отриманих результатів співвідносно до тексту, який аналізується, тобто перевірка гіпотези.

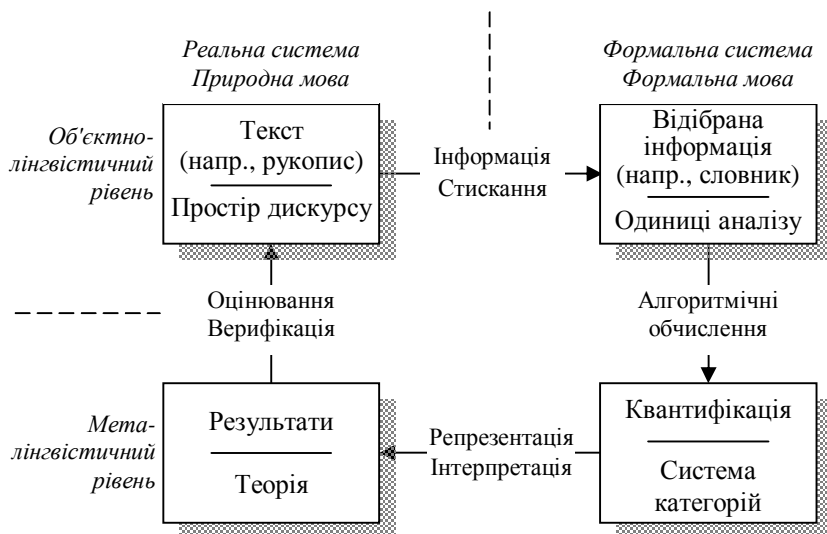


Рис. 3.4. Модель логіки комп'ютерного контент-аналізу

1. ТЕОРІЯ І З'ЯСУВАННЯ СУТНОСТІ ПРОБЛЕМИ

Який зміст перевіряється і чому? Чи ви маєте якісь теорії або практичні міркування щодо актуальності вивчення змісту тексту (наприклад, дослідження насильства на телебаченні показало, що це може мати вплив на дітей, тому потрібно проаналізувати ступінь показу агресії на телебаченні)? Чи потрібно попрацювати в бібліотеці? Чи будете ви доповнювати результати контент-аналізу іншими даними, щоб показати взаємозв'язок між джерелом і характеристиками отримувача інформації? Які є питання по дослідженню? Гіпотези?

2. КОНЦЕПТУАЛЬНІ РІШЕННЯ

Які зміни будуть використані при дослідженні та як ви визначите їх концептуально? Варто навести приклади змісту, щоб переконатися, що вами охоплено все, що планувалося зробити.

3. ОПЕРАЦІОНАЛІЗАЦІЯ

Ваше розуміння має відповідати концепції (це є внутрішня валідність). Якою буде одиниця масиву даних? Чи будете ви використовувати більше однієї одиниці (наприклад, окремо схема кодування промови і окремо схема кодування висловлювань)? Якою є міра для оцінювання змінних (категорії вищого рівня вимірюються вичерпною і взаємовиключною системою підкатегорій)? Має бути "а ргіогу" створена схема, яка описує всі категорії. При цьому має бути забезпечена зовнішня і внутрішня валідність?

Ручне
кодування

Комп'ютерне
кодування

4а. СХЕМА КОДУВАННЯ

Необхідно виробити такі інструменти:

- Інструкцію для кодування (з повним поясненням для всіх змінних)
- Бланки (форми) для кодування

4б. СХЕМА КОДУВАННЯ

У випадку комп'ютерного контент-аналізу також необхідна інструкція для кодування слів – чітке пояснення вашого словника і метод їх використання. Ви можете використати стандартні словники або створювати свої. Створюючи власний словник, спочатку на основі деякого зрзка вашого тексту складіть частотний список слів і визначте ключові слова і фрази.



Рис. 3.5. Типова схема реалізації процесу контент-аналізу

Якщо перші два кроки виконуються програмою здебільшого автоматично, то останні два – потребують утручання людини, оскільки потребують розуміння контексту.

Блок-схема типового процесу здійснення контент-аналітичних досліджень представлена на рис. 3.5¹⁶⁵, у додатку А подається один з варіантів детального алгоритму кількісного контент-аналізу.

Відмітимо особливості контент-аналізу як методики кількісного дослідження. Процес наукового пізнання йде шляхом більшого узагальнення фактів і знань. Однак стратегії якісного і кількісного досліджень є різними. Як бачимо на моделі, використання кількісних методів передбачає значне скорочення розмаїття інформації, початково присутньої в тексті. На відміну від якісних підходів, у яких ідея і додаткова інтерпретаційна інформація, навпаки, навіть додаються, кількісний контент-аналіз буде компактний, але більш загальний феномен. З інтерпретаційної позиції, якщо об'єкт і використовувана методика надто складні, щоб бути зрозумілими і прийнятими іншими, дослідник шукає спрощення (іншої інтерпретації) і більш прийнятної методики. У випадку кількісного аналізу дослідницька стратегія є протилежною: якщо згоди не досягнуто, то покрокове узагальнення триває до тих пір, поки цього не станеться.

У сучасному кількісному контент-аналізі сталися істотні зміни. Намагання усунути недоліки класичного аналізу попередніх років визначило основні напрями його поліпшення. Ці вдосконалення стосуються розвитку всіх концепцій, покладених у його основу. К.Кріппендорф¹⁶⁶, а слідом за ним Г.Ягія¹⁶⁷ виділяють шість базових концепцій:

- вихідні дані аналізу;
- контекст даних;
- поділ аналітиком контексту на сфери відповідно до рівня своїх знань;

¹⁶⁵ Neuendorf K.A. The Content Analysis Guidebook.–

<http://academic.csuohio.edu/kneuendorf/content/resources/flowc.htm>

¹⁶⁶ Krippendorf K. Content Analysis: An Introduction to Its Methodology.– Beverly Hills, CA: Sage Publications, 1980.

¹⁶⁷ Yahya H. Content analysis in human communication...

- цілі контент-аналізу;
- висновки як основне завдання дослідження;
- надійність як оптимальний критерій успіху.

Узявши їх за основу, розглянемо зміни методики контент-аналізу.

Істотне розширення сфери застосування контент-аналізу внесло зміни в самі джерела даних. Якщо на ранніх стадіях контент-аналіз був "стурбований" винятково вербальним змістом, то тепер об'єктом досліджень усе більше стають невербальні дані: поштові марки, музичні твори, твори мистецтва, глиняний посуд, фольклор, тексти і об'єкти Інтернету тощо. З появою комп'ютерної техніки носії даних усе більше стають електронними. Спеціально для комп'ютерних досліджень створюються бібліотеки електронних текстів. Особливого поширення отримали електронні тексти з появою і широким використанням Інтернет.

Ще від 1950 рр. контент-аналіз був підданий критиці за поверхневий аналіз, "неповагу" до прихованого (латентного) змісту та контексту. Саме ці аспекти виявилися головними в контент-аналізі. "Контент-аналіз і далі пов'язують із процесами категоризації та квантифікації, проте особливий наголос робиться на якісних процесах, пов'язаних із вивченням прихованого змісту і значень, а також із вивченням і конструюванням схем кодування для відверто явних повідомлень"¹⁶⁸.

Структуризація контексту все частіше стає об'єктом досліджень через труднощі інтерпретації результатів. Складність інтерпретації полягає в тому, що будь-який текст набуває змісту лише в контексті, під яким розуміють оточення, у яке "занурений" текст. Розрізняють мікро- і макроконтекст. Мікроконтекст – це найближче оточення тексту. Так, речення отримує зміст у контексті абзацу, абзац у контексті параграфу і т.д. Макроконтекст – це вся система знань, пов'язана з предметною областю, тобто знання про особливості і властивості, явно не вказані в тексті. Зрозуміти текст – означає сформулювати "інший текст", що являє собою його семантичну

¹⁶⁸ Jackson J. Content analysis and objectivity, language and metaphor...

(понятійну) структуру, складовими якої виступають "опорні" (ключові) слова або "змістовні віхи" тексту. У процесі аналізу важливо виявити як внутрішні зв'язки між елементами тексту і поняттями, так і зовнішні – з макроконтекстом. Внутрішні зв'язки традиційно поділяють на експліцитні (явні) й імпліцитні (приховані, латентні). Експліцитні зв'язки, які виражаються в зовнішньому подрібнюванні тексту, ділять текст на параграфи з допомогою зазначених вище складових, вставних слів або коннекторів типу "по-перше..., по-друге..., проте і т.д.". Імпліцитні зв'язки між окремими "змістовними віхами" викликають основні труднощі для розуміння.

З позицій макроконтексту, вирізняють такі компоненти, як:

- протекст (початковий задум автора);
- власне контекст (зокрема тематичний, мета якого збільшити активізацію тих знань у пам'яті, які стосуються даної теми; фактори часу і простору, які враховують темпоральну і просторову близькість описуваних подій; фактор асоціативних зв'язків тощо);

- затекст (атмосфера формування тексту);
- підтекст (неявні наміри автора);
- месідж (subliminal message – інтенція або послання, яке діє на підсвідомість, глибинний зміст), який задає комунікаційну спрямованість тексту, тобто вербальне втілення конкретної ситуації.

Така структуризація контексту і її врахування дозволяє в процесі аналізу розглядати досліджувану проблему, виявляючи, що мало місце до появи документа, що знайшло в ньому відображення і що матиме місце після нього, тобто як він сприйматиметься читачем.

Намагання вловити контекст, насамперед проявилось на таких ключових етапах методики, як визначення категорій і одиниць аналізу, формування процедури кодування, представлення та інтерпретація результатів аналізу. Свого часу робилися спроби формального виведення категорій аналізу, створення універсальних категорій, однак ці намагання виявилися або зовсім безрезультативними, або не давали практичного ефекту. Тому основні зусилля були спрямовані на здійснення категоризації з урахуванням різноманітних аспектів

контексту. Від категорій і одиниць аналізу вимагався такий суб'єктивізм, який був би цікавим для наступної політичної, культурної чи соціологічної інтерпретації, тобто те, як дослідник задасть значущість контексту вихідних понять, такий результат він отримає. Унаслідок цього об'єктом аналізу стають не лише елементи тексту, а й контекст.

Яскравим продуктивним зразком у рамках репрезентативного соціопсихологічного напрямку виявилася методика семантичного диференціалу, у якій контексти, що містять ключові слова поточної ситуації, одночасно оцінюються за кількома шкалами семантичного диференціалу, утворюючи своєрідний семантичний простір.

Активно, особливо з появою комп'ютерного контент-аналізу, застосовуються KWIC-індекси (ключові слова або сукупності ключових слів у контексті). Відтепер практично кожна комп'ютерна система кількісного контент-аналізу має засоби для створення таких індексів і представлення їх на вимогу користувача. Однак в¹⁶⁹ доводиться, що у випадку складного аналізу навіть це не допомагає.

Як говорилося раніше, ще від середини 1950 рр. дослідники багато уваги приділяють не просто перевірці наявності/відсутності або підрахунку частоти появи категорій у тексті, а зв'язкам між категоріями, що значно розширює коло задач, які може розв'язувати контент-аналіз. Відповідно до цього з'являються й терміни: концептуальний контент-аналіз, який досліджує самі категорії, і контент-аналіз зв'язків, увага якого прикута до вивчення зв'язків між категоріями (див.¹⁷⁰).

Один із варіантів контент-аналізу зв'язків полягає в тому, що звертають увагу на спільну появу в одних і тих же місцях слів різних категорій. Наприклад, для кожного речення з'ясовують, слова яких категорій у ньому зустрічаються. Можна також задавати ширину "контексту", наприклад, у такий спосіб: критерієм контекстного пошуку списків можуть бути 10 слів до

¹⁶⁹ Jackson J. Content analysis and objectivity, language and metaphor...

¹⁷⁰ Overview: Content Analysis. –

<http://writing.colostate.edu/references/research/content/index.cfm> ; Content Analysis: An Introduction...

появи терміну і 20 слів після появи терміну. Після цього підраховують коефіцієнт кореляції для оцінки сили зв'язку між категоріями і визначення знаку цього зв'язку. Може виявитися, що для деяких категорій спостерігається тенденція їхнього спільного вживання, а для інших – навпаки. В одних випадках це є відображенням свідомої позиції автора тексту, в інших – зв'язком на рівні підсвідомості.

Таке виявлення контекстів уживання слів – колокація (collocation)¹⁷¹ – отримало подальший розвиток. Як приклад, наведемо один із варіантів його практичної реалізації. Вибравши пропозиції, у яких зустрічається конкретне слово або категорія, одержують деяку підвибірку тексту, до якої також застосовують контент-аналіз. Контексти вживання слів і категорій, у свою чергу, можуть бути піддані контент-аналізу і підрахунку абсолютних і відносних частот категорій, оцінок категорій щодо норми тощо. Очевидним є величезний потенціал контекстного аналізу при моніторингу великих текстових обсягів, оскільки він дозволяє повністю автоматизувати процес збору інформації. Так зародилася ідея контент-моніторингу.

Чітке розуміння контексту потрібне і для правильного кодування, стаючи запорукою надійності і валідності результатів контент-аналізу. Для цього виробляються різноманітні інструментарії (інструкції кодувальнику, бланки контент-аналізу тощо) і стратегії взаємодії кодувальників у процесі спільного багаторазового кодування тексту, що дозволяє враховувати контекст безпосередньо на рівні процедури кодування. Так, у роботі¹⁷² розглядається "контекстуальний контент-аналіз" (авторська назва), який пропонує з багатьох контекстуальних соціальних вимірів враховувати в процедурі

¹⁷¹ Під *колокацією* розуміють систематичну (не одноразову) спільну появу слів у досліджуваних текстах. Спільна поява означає, що одне слово знаходиться на деякій відстані (задана кількість слів) справа або зліва від іншого слова.

Слова, які становлять предмет дослідження, називаються вузловими, наприклад, терміни, за якими створюється конкорданс, тоді як слова, які з'являються спільно з вузловими, – колокаційними. Детальніше див.:

Collocations, Colligations, and Phrases.–

http://www.ling.lancs.ac.uk/staff/andrew/aw-col_files/frame.htm

¹⁷² Mctavish D.G., Pirro E.B. Contextual Content Analysis // Quality & Quantity.– 1990.– No 24.– PP. 245-265.– www.cres.com/online-papers/qq.doc

систематичного кодування чотири рівні контексту: традиційний, практичний, емоційний і аналітичний.

Важливу роль у методиці контент-аналізу відіграє структуризація контексту відповідно до деякого рівня знань. З появою комп'ютерних систем стало можливим використання для структуризації контексту семантичних мереж, гіпертексту, методу масованої онтології концептуальних значень, частотно-імовірнісних і логіко-статистичних моделей і методів, методу рубрикацій, методу автоматичного нелінгвістичного аналізу неструктурованої текстової інформації, реалізованого на основі нейромережних алгоритмів тощо.

Одним із способів автоматичної нелінгвістичної структуризації текстової інформації є нейромережні алгоритми. Як значеннєвий портрет тексту розглядаються не просто списки ключових слів, а мережі понять – сукупності ключових слів або стійких словосполучень, пов'язаних між собою. Кожне поняття має свою вагу, яка визначає його значущість у тексті. Зв'язок між поняттями теж має вагу. Використання зв'язків дозволяє більш точно зважувати поняття тексту. У кінцевому підсумку, отриманий структурний портрет тексту являє собою тематичне дерево, яке разом із мережею понять описує зміст тексту і контексту.

Ще одним способом структуризації контексту є використання семантичних мереж. Семантична мережа служить для представлення знань у вигляді спрямованого графа, у якому поняття представляються як вузли, а відношення між поняттями – зв'язками між ними. Однією з переваг семантичних мереж виступає те, що близькі поняття прагнуть зв'язатися в єдиний пучок мережі.

Концептуально близьким до семантичних мереж є семантичні моделі, але, на відміну від перших, моделі представляють собою заздалегідь розроблений обмежений набір відношень (шаблонів), наприклад, універсальні для семантичних мереж бінарні відношення типу: "частина – ціле", "загальне – часткове", "причина – наслідок", а також деякі типові небінарні відношення: "джерело – продукт – споживач", "об'єкт – процес – суб'єкт", "вхід – процес – вихід".

Той чи інший спосіб структуризації тексту і контексту на даний час став обов'язковим атрибутом більшості програм контент-аналізу. Для прикладу, програма NUDIST будує ієрархії категорій, тоді як система ATLAS/ti формує неієрархічні зв'язки категорій – мережні структури, а система TextAnalyst перетворює початковий текст у гіпертекст, побудований на основі категорій, а також будує деревоподібну структуру і семантичну мережу категорій і контексту.

Наступною складовою методики контент-аналізу, яка інтенсивно розвивається, є намагання формувати висновки. На початкових стадіях розвитку для формування висновків у рамках кількісного контент-аналізу обмежувалися залученням методів статистичного аналізу. Однак поступово, особливо з розвитком якісного контент-аналізу, стали використовуватися додаткові засоби резюмування істотних відношень концептів досліджуваної проблеми.

Серед засобів вирізняються різноманітні техніки візуалізації та методи багатовимірної картографії для представлення асоційованої інформації великих масивів даних. Особливої уваги заслуговує візуальне представлення контекстуальної інформації у вигляді семантичних мереж, неймереж тощо. Приклади використання візуалізації наводяться в роботі ¹⁷³.

Характеризуючи застосування картографії в якісному контент-аналізі, автори роботи ¹⁷⁴ відзначають такі додаткові переваги, які отримують аналітик:

- картографічне представлення результатів аналізу забезпечує швидкі і зручні засоби знаходження значущих тем серед великих масивів інформації;
- значущі теми виявляються з даних швидше, ніж це зробив би дослідник без картографії;

¹⁷³ Wong P.C., Whitney P., Pacific J.T. Visualizing Association Rules for Text Mining / Northwest National Laboratory.– 1999.–
<http://www.pnl.gov/infoviz/InfoVis1999Association.pdf>

¹⁷⁴ Miller M.M., Riechert B.P. Identifying Themes via Concept Mapping: A New Method of Content Analysis / University of Tennessee, Knoxville.– 1994.–
<http://excellent.com.utk.edu/%7Emmmiller/pestmaps.txt>

– методи показують не лише відносну важливість тем, а й зв'язок між ними.

Крім цього, варто додати таке: картографічне представлення визначає не лише найвагоміші, а й всеможливі теми, тобто забезпечує виявлення та ранжування всіх, навіть незначних (одноразових) або ж несподіваних тем. Як бачимо, при тому, що сама процедура аналізу фактично залишилася незмінною, можливості аналітика щодо вироблення нових висновків істотно розширюються.

Зауважимо, що все це стало можливим лише завдяки появі комп'ютерних програм контент-аналізу.

Останнє, на чому слід зупинитися, це питання надійності і валідності результатів. Як ми бачили, проблеми надійності належать до найкритичніших. Хоча усвідомлення необхідності забезпечення надійності зростає, увага до неї є недостатньою. Так, у роботі ¹⁷⁵ у 200 публікаціях за 1994-1998 рр., присвячених контент-аналізу, вивчалася присутність оцінок надійності інтеркодування. Автори дійшли висновку, що лише 69% статей містили інформацію про перевірку надійності. Та й тут цьому важливому аспекту методики було присвячено в середньому на статтю лише 4,5 речення.

Проблеми надійності, як правило, виростають із неоднозначності слів, недостатньо чітко визначених категорій, недоліків інструментарію контент-аналізу, низької кваліфікації кодувальників тощо. К.Нойєндорф ¹⁷⁶ наводить також ще одну практичну причину: під час розподілу робіт між кількома кодувальниками забезпечення високого рівня надійності може служити критерієм організованості їхньої роботи.

У контент-аналізі показником надійності виступає надійність інтеркодування. За Дж.Мангеймом ¹⁷⁷, надійність інтеркодування (intercoder reliability – у перекладі з англійської часто використовують терміни "надійність", "стійкість",

¹⁷⁵ Lombard M., Snyder-Duch J., Bracken C. C. Content analysis in mass communication: Assessment and reporting of intercoder reliability // Human Communication Research.– 2002.– No 28.– PP.587-604.

¹⁷⁶ Neuendorf K.A. The Content Analysis Guidebook...

¹⁷⁷ Мангейм Дж.Б., Рич Р.К. Политология. Методы исследования. – М.: Весь Мир, 1997. – С.532.

"відтворюваність") – це узгодженість, досягнута стосовно значень, які приписують одній і тій же ознаці або схожим ознакам різні незалежні спостерігачі.

Перевірка надійності має гарантувати від помилок. Без забезпечення надійності результати аналізу та інтерпретація даних реально не мають сили. Якщо аналітики не можуть демонструвати високу надійність, то при використанні результатів потрібно проявляти особливу обережність. У¹⁷⁸ наводяться такі превентивні заходи для досягнення прийнятних рівнів надійності:

- максимально детально визначити зміст категорій;
- провести ґрунтовне навчання кодувальників;
- перед початком роботи переконатися, що кодувальники належно знають кодувальні інструкції;
- відвести час для обговорень і вирішення проблемних питань;
- обов'язково провести експериментальне/попереднє кодування на основі деякої підмножини реальних даних.

Безпосередньо в процесі роботи надійність перевіряється кодуванням одного і того ж тексту з використанням однієї і тієї ж кодувальної інструкції кількома людьми або повторним кодуванням його тим самим кодувальником через деякий час. Ознакою надійності є близькість результатів. Для оцінки близькості, а відтак і надійності, використовуються індикатори або індекси надійності.

Намагання віднайти єдиний універсальний індекс надійності інтеркодування поки що не увінчалися успіхом. Запропоновано чимало різноманітних мірил узгодженості результатів кодувань. У¹⁷⁹ наводиться їхня кількість – 39 індексів. Однак при цьому, зазначають автори, лише небагато з них використовуються на практиці, а саме:

- процент погодженості між кодувальниками;
- індекс Голсті (Holsti's index);
- індекс Скотта пі π (Scott's index π);

¹⁷⁸ Department of Journalism Studies...

¹⁷⁹ Lombard M., Snyder-Duch J., Bracken C.C. Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects...



Рис. 3.6. Напрями вдосконалення кількісного контент-аналізу

- індекс Коена каппа (Cohen's kappa κ);
- індекс Кріппендорфа α (Krippendorff's alpha α).

Для остаточної переконаності в досягненні цілей дослідження перевіряють валідність або обґрунтованість методики. У широкому розумінні обґрунтованість означає відповідність отриманих за допомогою методики емпіричних даних основним цілям дослідження. За Дж.Мангеймом¹⁸⁰, валідність (validity) – виражає ступінь відповідності міри тому поняттю, який вона повинна відобразити. Розрізняють зовнішню валідність (виражається в кореляції деякого показника вихідного поняття з показниками інших понять, які, згідно з теорією, пов'язані з вихідним поняттям) та внутрішню валідність (виражається в кореляції кількох показників одного й того ж поняття). На практиці валідність перевіряється використанням експертних оцінок або шляхом розв'язання досліджуваної проблеми іншими методами. Одна з перевірок обґрунтованості результатів полягає в тому, що тенденції змін показників в одному періоді можуть служити передбаченням змін тих же показників або подібних до них в іншому періоді.

Підсумки всього сказаного щодо напрямів удосконалення кількісного контент-аналізу представлено на рис. 3.6.

3.5. "Якісний" розвиток контент-аналізу

Як уже зазначалося, ХХ ст. вирізняється неперервною "боротьбою" між якісними і кількісними дослідженнями. Основні відмінності між відповідними методами досліджень у соціології представлені в табл. 3.3.

Зі змісту викладеного в попередніх параграфах, а також у результаті порівняння його з характеристиками, поданими в табл. 3.3, можна побачити, що контент-аналіз розвивався в напрямі значного посилення його інтерпретаційних

¹⁸⁰ Мангейм Дж.Б., Рич Р.К. Политология. Методы исследования. – М.: Весь Мир, 1997. – С.526.

Таблиця 3.3

Порівняльна таблиця якісних і кількісних методів
у соціологічних дослідженнях

Підстави порівняння	Кількісні методи	Якісні методи
Ціль застосування	Макродослідження	Мікродослідження
Ціль застосування	Дати пояснення причин досліджуваного явища	Зрозуміти досліджуване явище
Дослідницькі завдання	– виміряти параметри явища; – встановити взаємозв'язок між параметрами	– виявити загальну картину явища; – концептуалізувати явище, інтерпретувати його
Позиція дослідника	"сторонній" спостерігач	"співчуваючий" учасник
У центрі уваги дослідників	– соціальні структури та інститути; – об'єктивні фактори; – загальні соціальні процеси	– людина; – суб'єктивні фактори; – особливі, окремі процеси
Дослідницькі гіпотези формулюються	до початку збору даних	у міру оволодіння даними
Дослідницькі інструменти	– розробляються до початку досліджень; – формалізовані, багато в чому однакові для всіх дослідників	– визначаються як до початку, так і в ході досліджень; – неформалізовані, відбивають індивідуальний дослідницький досвід
Дослідницькі процедури	стандартизовані; передбачається їхне дублювання	менш стандартизовані; дублюються рідко

Підстави порівняння	Кількісні методи	Якісні методи
Одиниці аналізу	факти, події, висловлювання, акти поведінки	суб'єктивні значення фактів для індивіда
Логіка аналізу	Дедуктивна (від абстракції до фактів шляхом) операціоналізації понять	Індуктивна (від фактів до концепцій)
Основні способи аналізу	– класифікація шляхом ототожнення випадків; – статистичними способами; – систематизація	– опис випадків без їхнього ототожнення; – шляхом узагальнення виявлених оцінок; – уява
Дані дослідження представляються у вигляді	статистичних розподілів, показників за деякою шкалою, індексів тощо	висловлень, документів, інтелектуальних продуктів групових дискусій тощо
Валідність досягається достовірним	повторенням установлених зв'язків	описом випадків і виявлених думок
Стиль дослідження	твердий, холодний	м'який, теплий

Джерело: Таблиця підготовлена С.Григор'євим і Ю.Растовим¹⁸¹ на основі робіт К.Пламмера¹⁸² і В.Семенової¹⁸³.

¹⁸¹ Григорьев С.И., Растов Ю.Е. Основы современной социологии: Ученое пособие // Алтайский государственный университет. Центр мультимедиа технологий "Ирбис". – 2001. – <http://irbis.asu.ru/mmc/grig/4.1.ru.shtml>

¹⁸² Plummer K. Documents of Life. An Introduction to the Problems and Literature of a Humanistic Method. – London: Unwin Hyman Ltd., 1990. – P.6.

¹⁸³ Семенова В.В. Качественные методы: введение в гуманистическую социологию / Ин-т социологии РАН. – М.: Добросвет, 1998. – 289 с.

можливостей, тобто баланс між "кількісним" і "якісним" аналізом схилився в бік якісного. Однак навіть цей процес відбувався через протистояння думок.

Прихильники кількісного контент-аналізу, подібно до ¹⁸⁴, уважали, що кількісний аналіз потрібен на початковій стадії, тоді як час для якісного аналізу настає опісля. Іноді вони включають якісний контент-аналіз у свій методологічний арсенал з метою посилення надійності кількісних досліджень ¹⁸⁵. У запасі "якісників" свої аргументи: контент-аналіз проводиться лише для того, щоб дати відповідь на задані питання, а не формувати їх; відповіді ж істотно залежать від характеру питань ¹⁸⁶.

Однак більшість дослідників схилиються до думки, що обидва підходи є взаємодоповнювальними. Як доказ, наводяться аргументи на кшталт ¹⁸⁷: формування категорій аналізу більшою мірою пов'язане із "якісним" дослідженням, тоді як аналіз сукупності ознак категорій у тексті потребує найчастіше "кількісного" аналізу, тобто "якісний" аналіз присутній у "кількісному" і навпаки.

Насправді важко провести чітку межу між кількісним і якісним аналізом, оскільки науковці й до цих пір не виробили єдиного розуміння їхньої сутності. Початково покладалося в основу відмінностей те, що "кількісний контент-аналіз вивчав присутність/відсутність специфічних слів, повідомлень, образів або тем, які явно проглядаються в досліджуваних текстах, тоді як якісний аналіз цікавився прихованими та неявними повідомленнями або значеннями в межах текстів. Останнє потребувало відповідної системи інтерпретації, яка б базувалася не на простому підрахунку частот, а на вивченні більш тонких аспектів побудови тексту, розміщення його частин і змісту" ¹⁸⁸.

¹⁸⁴ Qualitative social science research methodology.–

<http://faculty.ncwc.edu/toconnor/308/308lect09.htm>

¹⁸⁵ Ньюман Л. Неопросные методы исследования...– С.120.

¹⁸⁶ Yahya H. Content analysis in human communication...

¹⁸⁷ Gallagher C, Marken M., Kim M., Phillipson T., Dodge T. The Use of Content Analysis in Researching Adolescent Issues...

¹⁸⁸ Content Analysis of Introductory Geology Textbooks.–

<http://www.scsv.nevada.edu/~promise/content.html>

Навіть донині окремі науковці дотримуються подібної думки, наприклад: "Відмінність між ними в тому, що кількісні методи зосереджуються на частоті появи характеристик або ознак, тоді як якісні методи концентруються на тому, присутня чи ні певна ознака. Якісні методи також називають нечастотними, не кількісними або нестатистичними. Відсутність/присутність характеристики використовується як певний культурний індикатор для вироблення висновків у якісній техніці. У кількісній техніці висновки базуються на розподілі частот"¹⁸⁹.

Однак з часом прийшло ширше розуміння якісного контент-аналізу, у якому більша роль відводиться технології здійснення і використання аналізу. Кількісний контент-аналіз практично не дає відповіді на питання: звідки з'являються категорії чи як розвивається система категорій. "Визначення категорій – це мистецтво", – заявляє К.Кріппендорф¹⁹⁰. "Однак у межах структури якісних підходів вони знаходяться в центрі інтересів, розвиваючи аспекти інтерпретації категорій якомога ближче до первинного текстового матеріалу і формулюючи їх у термінах матеріалу, – продовжує Ф.Мейрінг. – Для цих можливостей якісний контент-аналіз виробив процедури індуктивного розвитку категорій"¹⁹¹.

Якісний контент-аналіз використовує свою неформальну логіку, щоб дістатися до того, що насправді є реальним: якість, значення, контекст, образ дійсності або те, що люди фактично роблять, а не те, про що вони лише говорять. В основу якісного контент-аналізу найчастіше покладають "обґрунтовану теорію" Б.Глезера і А.Страуса, загальними принципами якої є:

- ітеративний характер процедури, коли аналітичний процес чергується зі збором інформації;

- предмет дослідження (сукупність його характеристик або властивостей) формується поступово, у міру розуміння і появи теорії, у процесі ітерацій, коли накопичується теоретична вибірка, яка репрезентує не об'єкт, а предмет дослідження;

¹⁸⁹ Gallagher C, Marken M., Kim M., Phillipson T., Dodge T. The Use of Content Analysis in Researching Adolescent Issues...

¹⁹⁰ Krippendorf K. Content Analysis: An Introduction to Its Methodology...– P.76.

¹⁹¹ Mayring P. Qualitative Content Analysis...

– на всіх стадіях аналітичного процесу здійснюється постійний порівняльний аналіз.

Метою такого контент-аналізу стає не перевірка наперед сформульованих теоретичних гіпотез, а власне систематичне і поступове вироблення певної теорії (гіпотез), у міру заглиблення в матеріал. Створення обґрунтованої теорії проходить три етапи: індукцію, дедукцію та верифікацію. Причому важливо, що вони задіяні в дослідженні не послідовно, а присутні на всіх стадіях дослідження.

Більш конкретно якісне дослідження можна описати так. На відміну від кількісних досліджень, в основу яких покладені ідеї ущільнення інформації про конкретні об'єкти, факти, процеси з метою отримання узагальнених характеристик досліджуваного явища для перевірки наперед висунутих гіпотез, якісні дослідження навпаки – пов'язані з відкриттям нових властивостей або особливостей саме конкретних фактів і процесів, а відтак, на основі цих емпіричних даних, з генеруванням гіпотез або ж "теорій".

Логіка якісного дослідника текстів не починає одразу з кодування на основі наперед розробленої схеми категорій аналізу, тут початком служить систематична оцінка всіх даних, необхідна для того, щоб уявити собі перспективи. Послідовний перегляд тексту супроводжується відзначенням характерних рис кожного окремого фрагмента і виробленням відміток щодо того, де саме (у якому місці тексту) ці особливості з'являються. У певному розумінні це є знахідка-кодування, яка виступає, як підготовка для інтерпретуючого аналізу тексту. Інтерпретуючий аналіз починається з обережного перегляду та аналізу відмічених особливостей, порівняння необроблених первинних даних (текстів) з метою ідентифікації якихось припущень щодо них.

Головний момент у якісному аналізі – порівняти різні текстові сегменти і мемо (примітки до сегментів тексту або ідеї, зроблені дослідником у процесі читання) для того, щоб виділити спільність, відмінності або зв'язки між ними. Мета цього – ідентифікувати структури і сконструювати "значущі взірці

фактів"¹⁹². Саме ці взірці (шаблони) стають основою формування гіпотез або "теорій". У процесі подальшого вивчення первинних даних (необроблених текстів, а не кодів), їхнього повторного перегляду відбувається роз'яснення, очищення і модифікація гіпотез. Тобто текст є безпосереднім джерелом інформації, яке не лише формує, але й змінює початкову гіпотезу. Як влучно стверджує У.Келле, "у багатьох якісних підходах, особливо тих, які мають сильні інтерпретаційні та феноменологічні коріння, перевірка інтерпретуючої гіпотези вимагає повноцінного дрібнозернистого аналізу текстових даних"¹⁹³.

Якісні дослідження виробили ряд власних прийомів і методів. Серед них:

- побудова індексів (на окремому папері дослідник відзначає місце, де обговорюється певна тема);

- техніка "виріж і встав" (cut-and-paste – матеріал вирізується або робляться виписки фрагментів тексту для полегшення їхнього наступного порівняння; в українському науковому фольклорі ця техніка отримала назву "метод рі-кле" – "ріж і клей");

- формування конкордансів¹⁹⁴;

- складання тематичних картотек;

- складання в процесі читання мемо для окремих фрагментів тексту.

Ф.Мейрінг¹⁹⁵ визначає якісний контент-аналіз у межах зазначеної структури, як методично керований емпіричний аналіз текстів із врахуванням їхнього контексту, що впливає з правил контент-аналізу і етапів моделі здійснення, без зайвого

¹⁹² Kelle U. Computer-Aided Qualitative Data Analysis: An Overview // Text Analysis and Computers / Cornelia Zuell, Janet Harkness, Juergen H.P. Hoffmeyer-Zlotnik (Eds.). Zentrum für Umfragen Methoden und Analysen (ZUMA).– Mannheim (Germany): ZUMA, 1996.– P.38.

¹⁹³ Kelle U. Computer-Aided Qualitative Data Analysis...– P.56.

¹⁹⁴ Конкорданс (concordance – відповідність) – алфавітний список важливих слів, який уключає безпосередні контексти їх використання. Часто конкорданси також містять додатковий матеріал, зокрема коментарі, означення, різноманітні індекси. Окрім зазначеного традиційного конкордансу, розрізняють його інші види: білінгвістичний, тематичний тощо.

¹⁹⁵ Mayring P. Qualitative Content Analysis...

визначення кількісних характеристик. При цьому якісний контент-аналіз зберігає найкращі риси кількісного контент-аналізу, що стає його перевагою порівняно з іншими методиками:

- пристосування матеріалу до моделі комунікації (ураховуються особливості комунікацій, досвіду, почуттів і думок комунікатора, ситуація створення тексту, соціокультурний фон, безпосередньо текст і ефект повідомлення);

- чітке дотримання правил аналізу (матеріал аналізується крок за кроком, чітко слідуючи правилам процедури, виділяючи одиниці аналізу);

- категорії постійно тримаються в центрі уваги (особливості інтерпретації тексту, прямуючи за питаннями дослідження, підганяються під категорії аналізу, які спочатку ретельно підбираються, а потім виправляються завдяки зворотному зв'язку);

- перевірка отриманих результатів згідно з критерієм надійності і валідності (процедура претендує на те, щоб бути зрозумілою для різних суб'єктів, а також намагаються отримані висновки підтвердити за допомогою інших методів).

Перелічені риси контент-аналізу стали основою для якісно орієнтованої процедури інтерпретації тексту. На думку Ф.Мейрінга¹⁹⁶, незважаючи на різноманіття процедур якісного контент-аналізу, проглядаються два принципово різних підходи до його здійснення: індуктивне визначення категорій (рис. 3.7) і дедуктивне застосування категорій (рис. 3.8). Основна ідея першого полягає в тому, щоб на основі теоретичних міркувань і змісту проблеми сформулювати критерій, який дозволить визначати істотні аспекти досліджуваного тексту. Відповідно до критерію опрацьовується весь матеріал, і крок за кроком виводяться категорії. Завдяки зворотному зв'язку категорії переглядаються і остаточно узагальнюються до рівня основних категорій з перевіркою їхньої надійності. У ході дослідження, якщо потрібно, визначаються кількісні характеристики категорій.

¹⁹⁶ Mayring P. Qualitative Content Analysis...

Другий підхід передбачає, що дедуктивне застосування категорій здійснюється на основі попередньо визначених особливостей аналізу, пов'язуючи їх з текстом. Якісний крок аналізу полягає в методично керованому призначенні категорій до проходження тексту. Тут головна ідея – дати явні означення, приклади і правила кодування для кожної дедуктивної категорії, точно визначаючи, які обставини походження тексту можуть кодуватися категорією.



Рис. 3.7. Алгоритм індуктивного визначення категорій у якісному контент-аналізі



Рис. 3.8. Алгоритм дедуктивного визначення категорій у якісному контент-аналізі

Як зазначалося, якісний контент-аналіз фактично є новим різновидом контент-аналізу, який органічно поєднує "обґрунтовану теорію" з кількісним контент-аналізом. Однак є чимало публікацій про спільне використання контент-аналізу з іншими методиками, зокрема якісними. Так, у роботі¹⁹⁷ описується спільне використання тієї ж "обґрунтованої теорії" та кількісного контент-аналізу для вивчення особливостей

¹⁹⁷ Bernard H.R., Ryan G. Text Analysis: Qualitative and Quantitative Methods...– P.595-646.

конфліктів з діловими партнерами в спільних китайсько-американських підприємствах. Ще одним прикладом може служити робота А.Уілсона¹⁹⁸, у якій контент-аналіз разом з дискурс-аналізом використовується для соціального осмислення текстових даних. Тут дискурс-аналіз забезпечує формування категорій і визначення взаємозв'язку між ними, а контент-аналіз аналізує на цій основі великі масиви текстових даних.

Узагальнюючи, В.Рао і М.Вулкок¹⁹⁹ стверджують, що спільне використання контент-аналізу з іншими методами здійснюється в трьох формах: паралельна (parallel), класична або послідовна (sequential) та повторювана (iterative). У паралельних підходах кількісні і якісні дослідження працюють незалежно, але отримані результати порівнюються та об'єднуються на стадії аналізу. На думку авторів, такий підхід дає найкращі результати для масштабних проєктів.

Для решти підходів характерний більш тісний діалог між методами на всіх стадіях дослідницького циклу, що є цілком можливим у випадку більш скромних за масштабами проєктів. Від такого поєднання забезпечується найбільша економія ресурсів і часу під час проведення дослідження. Послідовний підхід реалізується такими кроками:

1) формується розуміння проблеми у результаті якісного обговорення (наприклад, під час роботи фокус-групи, групи експертів, інтерв'ювання);

2) створюється інструментарій для аналізу;

3) формулюється гіпотеза та проводиться кількісне дослідження з використанням виробленого інструментарію.

Повторювальний підхід схожий на послідовний, однак він передбачає постійне повернення на початок для з'ясування і вирішення непорозумінь, виявлення аномалій тощо. Тут якісні

¹⁹⁸ Wilson A. Towards an Integration of Content Analysis and Discourse Analysis: The Automatic Linkage of Key Relations in Text / April 1993.–

<http://www.comp.lancs.ac.uk/computing/research/ucrel/papers/techpaper/vol3.pdf>

¹⁹⁹ Rao V., Woolcock M. Integrating Qualitative and Quantitative Approaches in Program Evaluation.– June 14, 2002.–

[http://lnweb18.worldbank.org/ESSD/essdext.nsf/61DocByUnid/8CDB76BF252A827E85256C0D0059AC4B/\\$FILE/Tool%20Kit%20Mixed%20Methods%20Chapter.doc](http://lnweb18.worldbank.org/ESSD/essdext.nsf/61DocByUnid/8CDB76BF252A827E85256C0D0059AC4B/$FILE/Tool%20Kit%20Mixed%20Methods%20Chapter.doc)

дані оцінюються як апіорні, що поновлюються кількісними дослідженнями.

* * *

Як бачимо, зміни, що мали і мають місце в період 1990-2000 рр., якнайтіснішим чином пов'язані з удосконаленням комп'ютерної техніки і програмного забезпечення, розвитком мережних технологій, використанням контент-аналізу до нових проблем і носіїв інформації. Серед тенденцій розвитку контент-аналізу можна виділити такі:

– удосконалення методики реалізації кількісного контент-аналізу під впливом потреб досліджень, накопиченого досвіду використання контент-аналізу, а також обґрунтованої критики та закидів у недосконалості з боку прихильників якісних методик досліджень;

– розвиток якісного контент-аналізу на вимогу тих же прихильників якісних досліджень; цьому сприяла також поява нових можливостей, пов'язаних з комп'ютеризацією та розвитком інтелектуальних розрахунків;

– комп'ютеризація та поява нових можливостей, пов'язаних з мережними технологіями і розвитком штучного інтелекту, перетворює контент-аналіз у технологію;

– контент-аналіз отримує всезростаюче масове застосування, активно інтегрується з іншими методиками соціальних та інших наукових досліджень;

– перетворення контент-аналізу в технологію масового застосування та інтеграція з іншими методиками наукових досліджень.

Узагальнено 1990-2000 рр. можна назвати періодом перетворення контент-аналізу в технологію масового використання.

Розділ 4. КОМП'ЮТЕРИЗОВАНИЙ КОНТЕНТ-АНАЛІЗ

Маючи відмінну від свого попередника – ручного контент-аналізу – технічну основу, комп'ютерний аналіз не лише виробив власну методику і технологію застосування, а й далі розвивається за своєю логікою. Це і є предметом нашого розгляду.

4.1. Перші спроби механізувати і автоматизувати контент-аналіз

Перші контент-дослідження з використанням технічних засобів у гуманітарній сфері пов'язані з роботами італійського священика Роберто Бузи, інформацію про діяльність якого можна віднайти в роботі Дж.Бредлі і Г.Рокуелла²⁰⁰, а також на сайті Григоріанського університету в Римі²⁰¹.

Отець Р.Буза ставив перед собою завдання з допомогою технічних засобів створити конкорданси повного зібрання листів італійського теолога XIII ст. Фоми Аквінського. Докторська дисертація Р.Бузи, захищена в 1946 р., присвячена вивченню поняття "присутність" у розумінні домініканського мислителя. Створивши і проаналізувавши друковані індекси латинських слів *praesens* і *praesentia*, Р.Буза дійшов висновку, що їхнє використання у Ф.Аквінського було тісно пов'язане применником *in*. Більше того, Р.Буза готовий був повірити, що подібні функціональні слова дають чимало інформації про зв'язок між концептуальним світом автора і словами, які він використовує для його опису. Однак достатніх дослідницьких ресурсів для створення ручним способом конкордансу таких загальних латинських слів, як *in*, *sum* або *et* (у перекладі відповідно "в", дієслово-зв'язка "бути", "і"), Р.Буза не мав.

²⁰⁰ Bradley J., Rockwell G. TACTweb: workbook...

²⁰¹ Roberto Busa: biography / Humanities and information science. Roma Univ. La Sapienza (24/11/95).–

<http://www.mediamente.rai.it/mmold/english/bibliote/biografi/b/busa.htm>

Попри труднощі, у кінці 1940 рр. він поставив перед собою набагато грандіозніше завдання – створити "Index Thomisticus", який би містив повний конкорданс 10,6 млн. слів Фоми Аквінського. Зрозуміло, що вирішення цього завдання без використання певного виду технічних пристроїв було неможливим.

Робота, розпочата з використанням перфораторів і сортувальних машин, була практично завершена через 33 роки на базі великих універсальних ЕОМ серії ІВМ. Разом з іншою інформацією, Індекс складався з майже 70000 сторінок. Було створено два конкорданси. Один, який формувався безпосередньо машиною, містив повний список відповідностей для всіх словоформ і отримав назву "нелематизованого". Другий конкорданс утворювали так звані "лематизовані" відповідності, у списку яких кожне слово зустрічалося лише один раз у стандартній формі, наприклад, іменники – лише в однині, дієслово – у невизначеній формі і т.д. Створити "лематизовану" відповідність без допомоги людини ЕОМ була не в змозі. За оцінками Бузи, на всю роботу використано понад 1 млн. людино/годин, головним чином для введення і перевірки даних, а також здійснення лематизації.

Зазначимо, що в 1992 р. отець Р.Буза заснував школу лінгвістики і герменевтики на філософському факультеті Грегоріанського папського університету в Римі, і продовжував працювати в рамках проекту "Totius Latinitatis Lemmata"²⁰².

Останнє публічне згадування про його знамениту роботу маємо за 2002 р., коли невтомному Р.Бузі виповнилося майже 90 років. У повідомленні Католицької інформаційної служби AGNUZ від 1 лютого 2002 р.²⁰³ говорилося про зустріч Святішого отця Іоанна Павла II з групою науковців, які "комп'ютеризували" твори св. Фоми Аквінського, і членів Асоціації з комп'ютеризації лексикологічних аналізів (CAEL),

²⁰² Roberto Busa: biography...

²⁰³ Папа за использование компьютера на службе у христианской доктрины / Католическая Информационная Служба "AGNUZ". – Ватикан. – 1 февраля 2002 г. –

<http://agnuz.info/print.php?year=2002&month1=February&day=1&files=101.txt&print=news>

які спонсорувала видання "Index Tomisticus", з нагоди завершення роботи. Окрім друкованого варіанту з 56 енциклопедичних томів, було створено електронну версію видання на CD-диску. Святіший отець підтримав ініціативу проф. Р.Бузи та міжнародної команди молодих учених щодо нового проекту "Лексикон св. Фоми Аквінського", який повинен дати сучасне тлумачення всіх термінів, використаних у роботах і листах середньовічного теолога.

Спроби механізувати окремі операції в процесі здійснення контент-аналізу не призвели і не могли призвести до істотних змін у його використанні і методиці. Водночас, слідом за збільшенням кількості самих текстів зростала потреба в контент-аналізі, який би швидко й ефективно аналізував тексти різної природи і великих обсягів. Поява комп'ютерної техніки створила реальні, хоча спочатку неясні і неусвідомлені за своїми майбутніми наслідками, можливості побороти "прокляту проблему розмірності".

Перші нерішучі кроки з автоматизації контент-аналізу швидко змінилися невиправданим оптимізмом. Однак сьогодні, через понад сорок років від перших спроб, незважаючи на реальні досягнення, стало очевидним, що, як висловилося К.Нойєндорф²⁰⁴, автоматичний контент-аналіз є "химерою". "Метод контент-аналізу, у якому використовується комп'ютер, найчастіше бачиться як механізований або автоматизований", – стверджує німецький дослідник Е.Мергенталер²⁰⁵. Тобто мова йде лише про той чи інший рівень використання комп'ютера людиною в процесі аналізу.

Під комп'ютеризованим контент-аналізом далі будемо розуміти аналіз текстових матеріалів, у якому всі кроки виявлення змістовних властивостей (характеристик) і визначення сукупностей концептуальних показників тексту здійснюються за допомогою алгоритмів, тобто явних логічних або статистичних операцій.

Дж.Макнамара розрізняє два рівні використання комп'ютерних програм:

²⁰⁴ Neuendorf K.A. The Content Analysis Guidebook... – P.40.

²⁰⁵ Mergenthaler E. Computer-Assisted Content Analysis... – P.3-32.

– запам'ятовування і зберігання даних, наприклад, отриманих дослідником ручним способом результатів кодування або нотаток, даних аналізу та звітів, зокрема у формі таблиць, діаграм і графіків;

– автоматичний перегляд текстів, розпізнавання та кодування слів і фраз; цей рівень може приводити до автоматичного кодування й аналізу або поєднання автоматичного і ручного перегляду та кодування²⁰⁶.

У.Келле поділяє програми аналізу текстів на три покоління:

1) текстові процесори і системи управління базами текстових даних, які забезпечують загальні менеджерські функції щодо текстів;

2) програми, пов'язані з технікою кодування і знаходження в текстах інформації, яка відповідає певним критеріям, зокрема побудова різного роду індексів і конкордансів, які є традиційними для ручної методики;

3) програми, які хоча й базуються на тих самих принципах, що і програми другого рівня, однак функції яких є значно ширшими і дозволяють істотно змінювати або посилювати можливості ручних методів контент-аналізу²⁰⁷.

З позицій сьогодення, на нашу думку, варто виділяти чотири покоління програмного забезпечення контент-аналізу (рис. 4.1). Критерієм такого поділу мають бути не стільки часові рамки, скільки та роль, яку відіграють програмні засоби в процесі аналізу. Розглянемо далі покоління програм контент-аналізу більш детально.

Перше покоління важко назвати програмами контент-аналізу, оскільки це лише використання різноманітних програм для цілей контент-аналізу. За влучним висловом Г.Бернарда і Г.Раяна, "подібно до ранніх текстових редакторів і систем управління базами даних, перше покоління текстових процесорів проектувалося як допомога нам робити те, що ми і

²⁰⁶ Macnamara J.R. Media Content Analysis: Its Uses, Benefits & Best Practice Methodology / CARMA International (Asia Pacific).– Chippendale (Australia): CARMA, 2003.– P.8.– www.masscom.com.au/book/papers/media_content.html

²⁰⁷ Kelle U. Computer-Aided Qualitative Data Analysis...– P.34.

Програми I покоління (1950 рр. – 1960 рр.)

- вузькоспеціалізовані програми для проведення окремих розрахунків або програми загального користування (текстові і табличні процесори, системи керування базами даних)
- призначені для проведення числових розрахунків, статистичного аналізу, побудови примітивних графіків
- можливість виконання окремих менеджерських функцій щодо тексту (зберігання, копіювання, формування індексів)

Програми II покоління (1966 р. – середина 1980 рр.)

- спеціалізовані програми кількісного аналізу, обмежені обробкою текстів і які не виходять за межі ручної технології
- основна увага на кодування, пошуку ключових слів і фраз, виведення інформації у формі різноманітних індексів, конкордансів
- дозволяють здійснювати ручне, автоматизоване і автоматичне кодування з використанням словників
- реалізуються різноманітні стратегії пошуку, формування таблиць конкордансу, здійснення аналізу колокацій
- можливість роботи з електронними текстами
- використовують винятково кількісні аналітики

Програми III покоління (середина 1980 рр. – 1990 рр.)

- спеціалізовані програми, обмежені обробкою текстів у рамках якісної та кількісної технології аналізу
- характеризується надзвичайним розмаїттям програм
- виходить за межі ручної технології і значно посилює аналітичні можливості дослідника, програма виступає своєрідним досвідченим експертом
- реалізуються функції всіх етапів досліджень, збільшилося число виконуваних функцій (структуризація даних, візуалізація результатів, формування та перевірка гіпотез, формування висновків і звітів), принципові зміни в реалізації тих функцій, які були реалізовані раніше (кодування, конкорданс)
- активно застосовуються кількісними дослідниками, привернули увагу "якісників"

Програми IV покоління (від середини 1990 рр.)

- вийшли за межі винятково обробки текстів
- технологічні програми, які в основі мають контент-аналіз і реалізуються для роботи в режимі реального часу
- реалізуються у вигляді масштабних систем зі складними математичними і лінгвістичними алгоритмами аналізу, для яких характерний розвинений графічний інтерфейс, доступ до різних джерел даних, функціонування в архітектурі клієнт-сервер

Рис. 4.1. Покоління програм контент-аналізу

так уже робили"²⁰⁸. Виходячи з можливостей перших універсальних ЕОМ, дослідники "доручали" їм виконувати числові розрахунки, здійснювати статистичний аналіз, будувати примітивні графіки вже після того, як прочитані і закодовані людьми тексти були введені з кодувальних карток на машинні носії. Як правило, спочатку це були спеціально написані під конкретні завдання програми, а дещо пізніше з'явилися більш універсальні програми: спеціалізовані пакети прикладних програм, системи управління базами даних, табличні процесори тощо. Що стосується безпосередньої обробки текстів, програми першого покоління дозволяли зберігати і архівувати самі тексти, будувати і видруковувати прості індекси і конкорданси, підшукувати цитати і виводити їх на друк у форматі на зразок KWIC-таблиць.

Тут варто процитувати Л.Федотову: "...приклад із машинною обробкою текстів ілюструє дуже важливу для розуміння сутності контент-аналізу думку... Аналіз змісту як метод не володіє магічними якостями – ви не отримаєте з нього більше, ніж уклали в нього. Якщо щось значуще, важливе, незвичайне не передбачалося процедурою, то воно не з'явиться в результаті аналізу, яким би складним і копітким він не був"²⁰⁹.

4.2. The General Inquirer і особливості другого покоління програм контент-аналізу

Появу другого покоління програм комп'ютеризованого контент-аналізу пов'язують з Гарвардським університетом (США), група науковців якого під керівництвом Ф.Стоуна в 1961 р. почала розробку принципів систематичного комп'ютеризованого контент-аналізу. Уже в 1966 р. була завершена перша версія спеціалізованої програми кількісного контент-аналізу текстів the General Inquirer, яка розв'язала міф про

²⁰⁸ Bernard H.R., Ryan G. Text Analysis: Qualitative and Quantitative Methods...– P.625.

²⁰⁹ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации...– С.111.

те, що універсальні ЕОМ можуть використовуватися лише для статистичного аналізу.

Перевіряючи можливості програми, автори проекту повторили чимало своїх і чужих досліджень, використовуючи категорійний апарат попередніх ручних методик. Була проведена серія аналізів текстів газет, наукових робіт і публіцистичних творів, промов кандидатів на президентську посаду від демократичної і республіканської партій США, особистих документів (листів, щоденників, автобіографій)²¹⁰. Одне з оригінальних завдань, яке не було дублюванням ручного дослідження і на якому тестувалася система, є дослідження 66 посмертних записок самогубців, з яких 33 – написані тими, хто справді здійснив самогубство, а решта – особами, що лише його симулювали. The General Inquirer проаналізувала тексти і в 91% випадків виявила записки справжніх самогубців²¹¹.

Процес аналізу в the General Inquirer відбувався таким чином. На початковому етапі компілюється словник категорій аналізу. Другим етапом і новим завданням комп'ютера виступає кодування тексту, яке полягає в тому, що система переглядає кожне слово і порівнює його з словоформами (категоріями) словника. Якщо словоформа віднайдена, то лічильник для відповідної словоформи збільшується на одиницю. У кінцевому підсумку отримують частотний розподіл категорій. Залежно від системи, у даний базовий алгоритм можуть вводитися нові правила, наприклад, для врахування контексту використання слів, усунення їхнього двозначного розуміння, виявлення в тексті не лише слів, а й цілих фраз. На третьому етапі програма виводить результати аналізу. Оскільки the General Inquirer є інструментарієм недіалоговим, то результатом стає, окрім звичайних таблиць з даними розрахунків, видрукування індексів і конкордансів. Як правило, інформація виводиться у форматі KWIC (ключові слова в контексті). Крім того, the General Inquirer дає можливість проводити нескладний

²¹⁰ Stone Ph., Dunphy D., Smith M., Ogilvie D. The General Inquirer...

²¹¹ Ogilvie D.M., Stone P.J., Schneidman E.S. Some Characteristics of Genuine Versus Simulated Suicide Notes // The General Inquirer: A Computer Approach to Content Analysis / P.J. Stone, D.C.Dunphy, M.S.Smith, D.M.Ogilvie, eds.– Cambridge: M.I.T. Press, 1966.– PP.527–535.

статистичний аналіз. Якщо виникає потреба, то він дозволяє експортувати дані у форматі інших програм (статистичних пакетів, електронних таблиць або програм ділової графіки).

Як зауважують науковці, наприклад, у²¹², підхід Ф.Стоуна добре спрацьовує для задач на зразок аналізу тем, пов'язаних із визначенням категорій аналізу, але він є недостатнім при вирішенні проблем, у яких потрібно відшукувати співвідношення між поняттями. Тим не менше, система стала прототипом програм комп'ютеризованого контент-аналізу і була величезним досягненням соціальних наук. Вона та її наступники продемонстрували широкі можливості маніпулювання текстами, їхнього кодування, виділення категорій аналізу, пошуку відповідностей тощо.

Узагальнюючи, можна стверджувати, що програми другого покоління за своїми функціями так і не вийшли за межі логіки ручної технології контент-аналізу і практично не змінили її. Вони були допоміжним інструментарієм, який просто полегшував виконувати рутинну роботу, яку аналітик робив і до їхньої появи. Як правило, це були програми кількісного контент-аналізу, однак деякі елементи якісного аналізу вже можна було виконати. Свою увагу програми насамперед зосереджували на кодуванні, а також пошуку ключових слів або фраз і виведенні результатів пошуку на друк. Досвід реалізації цих елементів методики комп'ютеризованого контент-аналізу достав теоретичне продовження і призвів до появи нових різновидів контент-аналізу.

Виходячи з правил і процедур кодування, розрізняють:

- ручне кодування;
- автоматизоване кодування, у якому комп'ютер допомагає людині знаходити відповідні категорії;
- автоматичне кодування без будь-якого втручання людини.

²¹² Carley K. Formalizing the social expert's knowledge // *Sociological Methods and Research*. – Vol. 17. – 1988. – PP. 165-232; Alexa M. Computer-assisted text analysis methodology in the social sciences / *Zentrum für Umfragen Methoden und Analysen (ZUMA)*. ZUMA-Arbeitsbericht 97/07. – Mannheim (Germany): ZUMA, 1997. – 40 p. – http://www.gesis.org/Publikationen/Berichte/ZUMA_Arbeitsberichte/97/97_07.pdf

Для комп'ютерних систем, які працювали в пакетному режимі, що виключав можливість організації діалогу людини і програми в процесі роботи, ручний режим роботи полягав у введенні кодів, підготовлених людиною самостійно. Такий варіант роботи характерний для програм першого покоління. Програми контент-аналізу другого покоління, які також не могли вести діалог з людиною, окрім ручного режиму роботи, реалізували автоматичний режим кодування, в основному для кодування на рівні слів. Автоматизоване кодування базується на діалозі людини з комп'ютером, і тому з чисто технічних причин реалізувати його в програмах другого покоління не було можливості.

Комп'ютеризований контент-аналіз запропонував два принципово відмінних підходи до автоматичного кодування, які умовно називають "a-priori" (або дедуктивний) та "a posteriori" (або індуктивний)²¹³. Підхід "a-priori", запропонований Ф.Стоуном при розробці the General Inquirer, більш відомий і поширений. Модель контент-аналізу, яка реалізується подібного роду системами, належить до категорії інструментального контент-аналізу. У чистому вигляді первинним тут виступає теорія, покладена в основу дослідження. Саме теорія визначає всі структурні компоненти дослідження: схему класифікації категорій аналізу, послідовність правил наступного кодування текстів, а також висновки, що будуть отримані в результаті дослідження. Фактично релевантність категорій базується на розумінні контексту аналітиком, його інтересів, інтуїції, досвіду і вмінь, цілей дослідження. Зауважимо, що аналітик у процесі дослідження може вносити зміни в класифікаційну схему, залежно від нового, глибшого розуміння тексту після отримання перших результатів, знаходження і виправлення недоречностей, помилок тощо.

Формалізоване представлення правил і умов кодування значною мірою реалізувалося через словники. "Уже з перших спроб використання електронно-обчислювальних пристроїв для роботи з текстом ставали очевидними переваги і недоліки

²¹³ Alexa M. Computer-assisted text analysis methodology in the social sciences...– 40 p.

обробки текстового матеріалу за допомогою машин – вони забезпечували адекватність аналізу величезних текстових матеріалів, однак вимагали значних зусиль для підготовки програм до роботи – складання "словника" з урахуванням усіх синонімічних варіантів понять, які треба буде відшукувати в тому морі слів, які пропускаються через машину"²¹⁴.

На практиці словник являє собою сукупність кількох тисяч словоформ, що належать до різних категорій. Категорії утворюють систему, яка виражає сутність певної проблеми, описує деяку тему або комплекс тем. Зазвичай, у словнику задаються слова для 60-150 категорій. Категорії підбираються або індуктивним способом, на основі тексту, або дедуктивно, на основі більш загальних теоретичних міркувань, які диктують вибір категорій. Кожній категорії в якості розшифрування задаються своєрідні "носії" змісту в реальній мові – слова в усіх своїх формах вираження або словоформи. "Побудова такого словника аналогічна побудові "тезауруса" – мови певної галузі людського знання, коли ключовим словам цієї сфери знань відповідає синонімічний ряд загальноживаних слів"²¹⁵. Комп'ютерний словник, по суті, являє собою різновид комп'ютеризованої кодувальної книги (codebook).

Прикладом словника може служити так званий Гарвардський психосоціологічний словник (the Harvard Psychosocial Dictionary), створений для роботи з the General Inquirer. Він розроблявся для використання в соціальних дослідженнях. Категорії словника формувалися дедуктивним способом, зокрема, виходячи із відомих теорій Г.Лассуелла, Т.Парсонса, Ч.Осгуда та інших. У процесі роботи категорії переглядалися й уточнювалися.

Однією з найбільш відомих є версія словника від 1975 р., яку розробники назвали Harvard IV. Вона включала 182 категорії, що спочатку об'єднували 4000 словоформ (входів у словник), а станом на 1990 р. – 8500. Найбільшою за числом

²¹⁴ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации...– С.108.

²¹⁵ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации...– С.107.

входів є категорія "negative" (негатив), на яку посилається 2291 словоформа. На відміну від своїх попередників, Harvard IV враховує багатозначність слів. Крім того, користувачі можуть самостійно вводити нові категорії з будь-якою кількістю відповідних словоформ.

Добре відомими є такі словники загального призначення, як Словник цінностей Лассуелла (the Lasswell Value Dictionary), Словник регресивних образів (the Regressive Imagery Dictionary), Стенфордський політичний словник. Так, Стенфордський словник включає 6584 слова, для яких визначено їхнє емоційне забарвлення. У ньому виділені слова, що характеризують позитивне сприйняття (977 слів), негативне сприйняття (1513 слів), силу (1391 слово), слабкість (579 слів), активність (1218 слів), пасивність (722 слова). Групи експертів, оцінюючи кожне слово за емоційною інтенсивністю, проводили "зважування" за трибальною шкалою. У такий спосіб сконструйовано "семантичне поле", що має три виміри:

- позитивне-негативне сприйняття;
- сила-слабкість;
- активність-пасивність²¹⁶.

Окрім загальних, часто використовуються спеціалізовані словники, які будують під конкретну, чітко визначену проблему. Наприклад, в Ульмському університеті²¹⁷ розроблено тематичний Словник стану неспокою (the Anxiety Theme Dictionary). Ще одним прикладом спеціалізованого словника є так званий Словник Ікара, побудований Д.Огілві для перевірки гіпотези про існування так званого "комплексу Ікара", психологічного явища, яке виявили психологи під час спостереження за своїми пацієнтами. Д.Огілві перевіряв гіпотезу, аналізуючи текстові матеріали, створені в рамках різних культур. Як вихідний матеріал було відібрано 626 казок з 44 примітивних суспільств. Попередньо для контент-аналізу було створено словник, який охоплював 74 категорії і 2500 слів,

²¹⁶ Иголкин А.А. Масс-медиа в условиях глобализации...– С.65.

²¹⁷ Grünzig H.J. Themes of anxiety as psychotherapeutic process variables // Methodology in psychotherapeutic research / Minsel W.R., W.Herff (eds.). Proceedings of the 1st European Conference on Psychotherapy Research, Trier, 1981. – Lang, Frankfurt, 1983.– PP.135-142.

які їх наповнюють. При розробці системи категорій використовувалися різні джерела, роботи дослідників, які описували комплекс Ікара, життєві історії людей, у яких явно проглядався цей комплекс, тексти спостережень за ними психотерапевтів тощо. Оскільки початково комплекс Ікара визначався як спрямованість на "політ, падіння, вогонь, воду, безсмертя і нарцистизм", то гіпотезу про існування комплексу на рівні аналізованих казок можна було, на думку автора, вважати доведеною, якщо ці теми "пов'язані" з конкретними фольклорними одиницями або на певному етапі з'являється можливість на основі одних тем передбачити появу в тексті суміжних тем. Підсумки аналізу отримувалися після розрахунку кореляцій між 74 категоріями і в результаті факторного аналізу кореляційної матриці²¹⁸.

Автоматичний варіант кодування, який використовує елементи штучного інтелекту, відрізняється від попереднього варіанта, зокрема способами, якими вирішуються проблеми неоднозначності слів. Комп'ютерні системи зі штучним інтелектом намагаються враховувати як синтаксис, так і лексичні особливості слів. Програми цієї категорії, наприклад, пізніші версії the General Inquirer зі словником Harvard IV включають правила, які розрізняють серед різних варіантів те значення слова, яке відповідає контексту.

Ще один момент був усвідомлений з появою програм контент-аналізу: що саме кодувати в тексті? кодувати всі слова чи ні? З цієї позиції виділилося два напрями контент-аналізу: вибірковий і "тотальний". Перший, реалізований Ф.Стоуном, передбачав кодування в тексті не всіх слів, а лише незначної частини, які виражають сутність досліджуваної проблеми. До свого словника Ф.Стоун включив 5 тис. слів, якими найчастіше щоденно послуговуються²¹⁹.

²¹⁸ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации...– С.108.

²¹⁹ Mergenthaler E. Computer-Assisted Content Analysis...– P.8.

Різновид контент-аналізу, який отримав назву "тотальний", був запропонований Ю.Лаффалем. Ю.Лаффаль²²⁰ намагався кодувати майже кожне слово тексту, за винятком функціональних слів, які мають найбільшу частоту в мові, тобто аналіз охоплював широкий діапазон змістовно пов'язаних іменників, прикметників, дієслів тощо. Його словник спочатку включав 114 категорій. За інформацією Г.Бернард і Г.Раян, станом на початку 1990 рр. словник Лаффала включав 43 тис. слів, кожне з яких асоціювалося з 1-5 категоріями із 168 можливих²²¹. Для порівняння: аналіз, запропонований Ф.Стоуном, використовував для кодування близько 10% тексту, тоді як за Ю.Лаффалем покриття тексту категоріями складало майже 90%. Зауважимо, що "тотальний" контент-аналіз знайшов найбільше застосування в психотерапії, коли акцент робиться на вивченні мови пацієнтів.

Альтернативним до систем, побудованих за зразком Ф.Стоуна, проте більш "просунутим", виявився різновид систем автоматичного контент-аналізу "a posteriori", який не потребував попередньої побудови словника, тим самим виключаючи присутність людини навіть на рівні формування категорій. Цей різновид комп'ютеризованого аналізу Г.П.Айкер і Н.І.Гарвей ще в кінці 1960 рр. назвали аналізом, якого "не торкається рука людини"²²². На відміну від підходу "a priori", він насамперед "керується" даними, а не якоюсь теорією. Тут схема категорій аналізу формується в результаті перегляду досліджуваного тексту. При цьому від аналітика на вході не вимагається жодної додаткової інформації, окрім досліджуваного тексту. Фактично такі системи реалізують репрезентативну модель контент-аналізу.

²²⁰ Laffal J. Total or Selected Content Analysis // International Conference on Computational Linguistics, Sönga-Såby-Kuragård, Sweden, 1-4 September, 1969.– acl.ldc.upenn.edu/C/C69/C69-2401.pdf

²²¹ Bernard H.R., Ryan G. Text Analysis: Qualitative and Quantitative Methods...– P.629.

²²² Iker H.P., Harway N.I. A Computer Systems Approach Toward the Recognition and Analysis of Content // The Analysis of Communication Content / Gerbner G.A. et al. (eds.).– Wiley & Sons, 1969.

Сьогодні, за твердженням П.Мохлера і Ц.Зуеля, автоматичні системи цього різновиду контент-аналізу переживають період ренесансу²²³. Прикладом їх можуть бути програми the Words, TextSmart, DICTION. Логіка the Words, розроблена Г.П.Айкером і Н.І.Гарвеєм в кінці 1960 рр., полягає в такому. Спочатку текст поділяється на окремі сегменти, для яких формується таблиця частот усіх слів, за винятком функціональних і з урахуванням синонімів. По кожному сегменту відбирається *n* слів з найбільшою частотою, які утворюють *n* міні-категорій. Далі обчислюється на основі всіх сегментів матриця взаємних кореляцій між цими категоріями, яка піддається факторному аналізу. У результаті визначаються актуальні або неактуальні міні-категорії (або теми) тексту.

Зазначимо також, що автоматичне кодування забезпечує більшу точність, надійність і економію ресурсів порівняно з ручним. Як зауважує М.Алекса, "при ручному кодуванні, коли чітко не визначені правила кодування, неминучим є розходження між думками різних кодувальників. Вимога щодо явного, одноваріантного і точного вибору категорій є критичною в процесі кодування. У цьому відношенні автоматичне кодування має очевидні переваги порівняно з ручним. Автоматичне кодування виконується на основі явно сформульованих правил і однозначних умов"²²⁴. Основним джерелом труднощів автоматичного кодування є двозначність слів, використання метафор, займенників, синонімів тощо.

Іншою визначальною компонентою програм другого покоління є реалізація стратегій пошуку даних. Серед них виділяється побудова таблиць конкордансу. Важливість їх яскраво свідчить хоча б те, що в назвах багатьох програм присутнє слово конкорданс. Так, програма СОСОА (Count and Concordance generation for the Atlas) є складовою програми the Atlas. У 1978 р. комп'ютерний центр Оксфордського

²²³ Mohler P., Zuell C. A Popperian Critique of Automatic Content Analysis // Journées Internationales d'Analyse Statistique des Données Textuelles / ZUMA.– 2000.– No 5.– <http://www.cavi.univ-paris3.fr/lexicometrika/jadt/jadt2000/pdf/10/10.pdf>

²²⁴ Alexa M. Computer-assisted text analysis methodology in the social sciences...– 40 p.

університету на зміну COCOA випустив OCP (the Oxford Concordance Program), а пізніше Micro-OCP для мікрокомп'ютерів. Добре відома система TACT (Text-Analysis and Concordance Tools) у своїй назві також має слово конкорданс.

Основна ціль конкордансів – спрямувати увагу на безпосереднє лінгвістичне середовище вибраного слова. Логіка пошуку полягає в тому, що спочатку дослідник виявляє потенційно цікаве слово, далі знаходить відповідний йому конкорданс, що дає змогу визначити шаблони (патерни), характерні для даного слова і в яких даному слову відводиться цілком визначена роль.

Існує декілька форматів конкордансу. Один із способів демонстрації контексту зустрічання слів є формат KWOC (keyword-out-of-context – ключове слово поза контекстом) – перелік слів із вказанням місцезнаходження. У ньому ключове слово показується справа або зліва від контексту, а контекст подається у вигляді цілого речення, яке може займати декілька рядків. Більш поширеним є альтернативний формат KWIC (keyword-in-context – ключове слово в контексті), який займає лише один рядок із ключовим словом у центрі цього рядка (однакова кількість слів справа і зліва від ключового слова).

Разом з конкордансом здійснюють колокацію (collocate analysis) – статичний аналіз появи комбінацій слів. Як ми вже говорили, завданням колокації є визначення слів, які зустрічаються в близькому околі заданого вузлового слова. Після вибору вузлового слова або групи вузлових слів, об'єднаних спільною ідеєю або спільним об'єктом, визначаються всі слова, які знаходяться зліва і справа від вузлових слів у межах заданої відстані. Ця віддаль істотно залежить від мови, наприклад, емпірично доведено, що для англійської мови значущими є зв'язки в межах п'яти слів зліва і права від вузлового слова. Зібраний у такий спосіб контекст утворює мінітекст. Далі проводиться аналіз частоти зустрічання слів у мінітекстах. Різниця між очікуваною частотою появи тих чи інших слів і реальною частотою зустрічання служить мірою "притягання" слів вузлом.

Потреби комп'ютерної обробки породили необхідність створення електронних текстів, які можуть без зусиль автоматично опрацьовуватися комп'ютером. У 1970 рр. не лише було розроблені методи і теоретичні пропозиції щодо формування текстових архівів (наприклад, ²²⁵), а й розпочато практичну реалізацію цих ідей. До числа найстаріших і найвідоміших відноситься Оксфордський текстовий архів, який нині включає великі високоякісні, добре задокументовані колекції грецької, латинської та англійської літератури. Оксфордський архів із 1976 р. займається розповсюдженням електронних текстів серед зацікавлених осіб (ознайомитися з можливостями архівом можна на відповідному сайті ²²⁶).

Електронні архіви почали створюватися в багатьох університетах, установах, організаціях. Як наслідок, сьогодні, на відміну від 1960-1970 рр., коли доступних електронних текстів було мінімум і їхнє створення було процесом складним і довготривалим, доступ до машинних текстів став набагато простішим. Зараз існує надзвичайно багато електронних текстових архівів з різноманітних джерел, різного призначення і на різних мовах. Нинішні повнотекстові бази даних включають не лише електронні копії оригіналів першоджерел (скановані тексти, рукописи, малюнки, аудіо-, відео-, теле-, фотозображення), а й незалежні від них засоби інтерпретації. Швидкий доступ у режимі on-line дає можливість аналітикам без особливих зусиль використовувати тексти, у тому числі й для здійснення контент-аналізу з використанням комп'ютерів.

Доступність електронних текстів вплинула на розробки програм контент-аналізу. "Відтепер проблеми, пов'язані з набором і редагуванням текстів, не є домінуючими, замість цього дослідники фокусуються на нових шляхах використання комп'ютера для кращого розуміння текстів; у доповнення до звичайного пошуку слів і фраз у тексті стало можливим більше використання багатомірних статистичних методів, що

²²⁵ DeWeese L.C. Computer content analysis of day-old newspapers: a feasibility study // Public Opinion Quarterly.– Vol. 41.– 1977.– PP.91-94; DeWeese L.C. Computer content analysis of printed media: a limited feasibility study // Public Opinion Quarterly.– Vol. 40.– 1976.– PP.92-100.

²²⁶ The Oxford Text Archive.– <http://ota.ahds.ac.uk/ota/index.html>

дозволяє зробити більше, ніж просто знайти що-небудь у тексті"²²⁷.

На завершення відзначимо, що можливості програм контент-аналізу другого покоління насамперед привернули увагу дослідників, що працювали у сфері формалізованих методів досліджень текстів. Якісні ж дослідники навіть не робили спроб застосовувати їх у своїй аналітичній роботі. Тим не менше, численні контент-дослідження, як правило, мас-медійних джерел і в основному англомовних текстів, дозволили виробити методологічні засади нового підходу до емпіричних досліджень у рамках суспільних наук. Л.Н.Федотова зауважує, що в 1974 р. на робочій нараді з проблем контент-аналізу в Італії було представлено навіть декілька міжнародних проєктів, виконаних з використанням ЕОМ, зокрема проєкт міжнародного дослідження газетних заголовків з метою визначення уваги друкованих ЗМІ до місцевих, національних і світових подій, порівняння уваги американських і європейських газет до проблем "Спільного ринку", порівняння висвітлення ходу ведення війни в Нігерії газетами різних країн тощо²²⁸. Активність дослідників була настільки високою, що чимало науковців період 1960–1970 рр. називають "золотим віком" комп'ютеризованого контент-аналізу.

Однак зусилля й оптимізм, інвестовані в 1960-х рр. у комп'ютеризований контент-аналіз, на 1970-х рр. не поширилися. Головним чином це пов'язують, на думку М.Алекси, із повільним розвитком обчислювальної техніки, обмеженістю доступу до ЕОМ (доступ в межах обчислювальних центрів), а також відсутністю достатньої бази електронних текстів, не кажучи вже про труднощі з переведення в машинний формат розмовних текстів²²⁹. Відтак зменшилася кількість наукових публікацій, теоретичні дослідження поступилися маломасштабним прикладним, поступово наростав методологічний застій. Щоправда, саме в це десятиріччя

²²⁷ Bradley J., Rockwell G. TACTweb: workbook...

²²⁸ Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации...– С.110.

²²⁹ Alexa M. Computer-assisted text analysis methodology in the social sciences...– P.8.

комп'ютеризований контент-аналіз почав застосовуватися в психології та психотерапії, а також отримав ширше розповсюдження в Європі. Така ситуація зберігалася аж до середини 1980 рр.

4.3. Третє покоління програм контент-аналізу

Від середини 1980 рр. намітився значний прогрес у розвитку комп'ютеризованого контент-аналізу. Його стимулювали декілька переходів: у 1980 рр. – від великих ЕОМ до персональних комп'ютерів, а потім – від операційної системи MS DOS до MS Windows з його графічним інтерфейсом і дружнім ставленням до користувача. Ці технічні і технологічні зміни створили можливості для забезпечення вільного діалогу людини з комп'ютером, який став стрижнем усієї дослідницької роботи, а не лише адміністрування текстових даних. Органічним доповненням був бурхливий розвиток Інтернет, широке розповсюдження електронних архівів-бібліотек, загальна доступність електронних текстів і можливість доступу до текстових архівів через Інтернет у режимі on-line.

Одразу зауважимо, що саме третє покоління програм контент-аналізу привернуло увагу якісних дослідників, які побачили в них величезні, нездійсненні до цих пір можливості неформалізованої обробки текстів. "Стало зрозуміло, що, хоча комп'ютери не можна використовувати безпосередньо для інтерпретаційного аналізу текстів, проте вони, без сумніву, можуть надавати значну допомогу в процесі інтерпретації"²³⁰. Це особливо виявилось очевидним, коли виникла потреба обробки значної кількості неструктурованих текстових даних. Навіть незважаючи на те, що використання комп'ютерних програм у якісних дослідженнях дуже тісно пов'язане зі специфікою цього роду досліджень.

Програмне забезпечення третього покоління характеризується надзвичайним розмаїттям. Серед найвідоміших тогочасних систем – Atlas/ti, HyperResearch,

²³⁰ Kelle U. Computer-Aided Qualitative Data Analysis...– P.36.

Aquad, NUD'IST. Незважаючи на те, що нове покоління програм будувалося на тих же принципах, що і їхні попередники, однак воно дозволяло здійснювати контент-аналіз, який виходив за межі ручної технології і значно посилював аналітичні можливості дослідника. Орієнтовані на якісні та кількісні дослідження, програми третього покоління мають засоби для:

- формування текстів і створення на їхній основі цілих проектів;

- вивчення частоти і контексту використання слів, зокрема, даючи відповіді на запитання: як часто категорії присвоюються словам або текстовим сегментам? які категорії і як часто вони з'являються разом? які зв'язки існують між категоріями або текстовими сегментами;

- створення і підтримка категорій і схем класифікації;

- присвоєння однієї або більше категорій рядкам символів, словам, фразам, реченням, параграфам або цілим текстам;

- зберігання приміток ("memo") до текстів, кодування текстових сегментів;

- отримання різних форматів перегляду текстів, частин текстів або груп текстів;

- експортування кодів для подальшої обробки їх іншими програмами, а також формування звітів з проведеного аналізу;

- підтримка командної або спільної роботи в рамках проекту і злиття в один кількох проектів²³¹.

Стосовно якісного контент-аналізу, то в програмах було реалізовано ряд уже згадуваних нами ручних методів і прийомів якісних досліджень, а також нових можливостей. Функції програм якісного контент-аналізу можна згрупувати таким чином:

- управління текстовими даними та супроводжуючою їх інформацією;

²³¹ Alexa M., Zuell C. Commonalities, differences and limitations of text analysis Software: The results of a review / Zentrum für Umfragen Methoden und Analysen (ZUMA). ZUMA-Arbeitsbericht 99/06.– Mannheim (Germany): ZUMA, 1999.– P.2.–

http://www.gesis.org/Publikationen/Berichte/ZUMA_Arbeitsberichte/99/99_06.pdf

– систематична оцінка фрагментів тексту, побудова категорій аналізу на її основі, а також конструювання взаємозв'язків між категоріями аналізу;

– пошук фрагментів тексту на основі сукупності критеріїв;

– перевірка інтерпретуючих гіпотез.

Найпершим завданням, яке постало перед програми якісного контент-аналізу, стало управління неструктурованими текстовими базами даних. Використання традиційних СУБД не забезпечувало потрібних операцій. Ідея програм якісного контент-аналізу полягала в тому, що окрім бази даних з основним текстом, створювалися спеціальні файли або бази даних, які містили адреси сегментів тексту (наприклад, номер запису початку і номер кінця) та імена кодів, які асоціювалися з даним сегментом. У табл. 4.1 наводиться приклад такого файлу. Якщо кодом CZR позначено категорію "Чеська Республіка", а MGR – "міжнародні міграції", то, як видно з таблиці, в обох документах ці теми присутні. Зокрема, категорія "Чеська республіка" у першому документі зустрічається у фрагменті, який починається з 205 рядка, а завершується 219 рядком. Причому, у другому документі мова йде міжнародні міграції саме в Чеській республіці, про це свідчить той факт, що обидва коди присутні у фрагментах, які мають спільні рядки. За допомогою даного файлу можна виводити окремо ті сегменти тексту, в яких присутні потрібні коди.

Таблиця 4.1

Використання кодів як вказівників сегментів текстів

Номер документа	Назва коду сегмента тексту	№ першого рядка сегмента в документі	№ останнього рядка сегмента в документі
1	CZR	205	219
1	MGR	45	87
2	CZR	50	54
2	MGR	43	67
2	CZR	156	190

Такий допоміжний файл використовувався для пошуку і виділення потрібних фрагментів тексту. Доповнюючи ці файли новими записами, можна постійно розширювати пошукову базу, не зачіпаючи самого тексту. Такий принцип був закладений у перших пакетах програм Qualpro, the Ethnograph, Textbase Alpha.

Вирішення першого завдання дозволило без особливих проблем виконувати й інші функції аналізу: пошук фрагментів тексту; побудова конкордансів; здійснення колокацій; перевірка інтерпретуючих гіпотез шляхом пошуку сегментів з однаковими кодами; введення, редагування та зберігання теоретичних коментарів до фрагментів текстів і т.д. Істотним доповненням до них стали різноманітні представлення взаємопов'язаних категорій шляхом різних способів візуалізації. Представлення теорій у вигляді мережі взаємозв'язаних категорій є чудовим інструментом демонстрації структури новоствореної теорії. З допомогою допоміжних файлів показників можна легко встановлювати зв'язки між сегментами текстів, мемо, кодами. Так, програма NUD'IST дозволяє будувати ієрархічні та мережні структури категорій, програма Atlas/ti формує різноманітні неієрархічні мережі. Детальне знайомство з реалізацією зазначених алгоритмів можна знайти в²³².

Зміни відбулися на всіх етапах технології досліджень. Насамперед, вони торкалися кодування. Інтеграція ручного і автоматичного кодування стало новою технологією багатьох програм, наприклад, PLCA (Program for Linguistic Content Analysis), MECA (Map Extraction, Comparison and Analysis). Відтак стало очевидним, що традиційна відмінність між автоматичним і ручним контент-аналізом стирається. "Підтримка або не підтримка дружної і доступної технології більше не є обмеженням для дослідника, не змушує вдаватися до єдиного режиму аналізу"²³³. Від поєднання обох підходів особливо вирає аналітик.

Змінилася така функція, як використання конкорденсу. Вона набула "діалогових" властивостей. Хоча частота появи слів

²³² Kelle U. Computer-Aided Qualitative Data Analysis...— P.33-63.

²³³ Alexa M. Computer-assisted text analysis methodology in the social sciences...— P.9.

і далі залишалася домінуючим джерелом інформації і служила одним з індикаторів потенційного інтересу до категорії аналізу або теми, однак відтепер, маючи на екрані слова (категорії) і їхню частоту, дослідник оперативно з допомогою KWIC-таблиці переглядав, у якому контексті з'являлося відібране ним слово. Ця процедура змінила перспективи використання конкорданс, що підсилило переконливість висновків, заготовки яких виробляла система. Такий режим реалізується через систему взаємозв'язаних вікон у багатьох програмах, зокрема навіть під MS DOS у програмі TACT.

Збільшилося число нових функцій, які реалізують програми. З'явилася можливість перевіряти гіпотези, розширилися інтерпретаційні можливості за рахунок методів пошуку спільного входження кодів (слів), побудови концептуальних моделей, які зв'язують поняття в семантичні мережі, матричного, логічного та картографічного аналізу. Чимало програм приділяють увагу підрахункам показників надійності, наприклад, програми AGREE, Krippendorf's alpha 3.12a, PRAM (Program for Reliability Assessment of Multiple Coders), або окремі модулі статистичних пакетів програм, зокрема SPSS і Simstat (див. ²³⁴).

Такі можливості послужили основою створення цілого ряду програмних систем, які дозволяли вирішувати найрізноманітніші специфічні проблеми в процесі аналізу. Так, побудована на концепції "concept mapping", програма the VBPro дозволяє шляхом картографічного представлення ідентифікувати домінуючі теми і взаємозв'язки між темами для великих масивів даних. Цей різновид текстового аналізу використовується для аналізу медійних повідомлень.

Інший різновид аналізу представляє програма Minnesota Contextual Content analysis (МССА), яка дозволяє вимірювати соціальну різницю (відстань) між статусом людей в організації, наприклад, лікарями і пацієнтами в лікарні, менеджерами та іншими працівниками фірми, ураховуючи стилістичні

²³⁴ Lombard M., Snyder-Duch J., Bracken C.C. Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects. – 2004. – <http://www.temple.edu/mmc/reliability/>

особливості мови в процесі бесіди, а також контекстуальну інформацію. На відміну від стандартного контент-аналізу, дана методологія враховує чотири контекстуальних виміри: традиційний, практичний, емоційний і аналітичний. Здійснюючи кластерний аналіз, система дозволяє кількісно оцінити ступінь близькості між представниками різних соціальних груп.

Водночас з'явилися системи контент-аналізу для ефективної роботи в окремих сферах. До більш сучасних систем можна віднести програми для дослідження медіа, наприклад, CARMA® (Computer Aided Research & Media Analysis), Precis™, Echo®Research, ІМПАКТ™, Metrica, the Delahaye Medialink system. Окрім здійснення самого контент-аналізу, ці програми включають такі модулі, як убудовані медіабази даних, що забезпечують уведення, доступ і підрахунок відповідних статистичних даних, які стосуються різних параметрів медіа-засобів²³⁵. З характеристиками цілого ряду систем контент-аналізу можна ознайомитися з оглядових матеріалів²³⁶.

Незважаючи на явні успіхи, критика в бік комп'ютеризованого контент-аналізу практично не вщухала. Основне її спрямування пов'язують з тим, що Дж.Зейдель називає "темним боком технічного поступу"²³⁷. Більш конкретно ці "темні сторони" виражаються в такому:

- програма дистанціює людину від самих даних, людина практично не відчуває самого алгоритму;
- використання програм призведе до того, що якісні дані будуть аналізуватися кількісно;

²³⁵ Macnamara J.R. Media Content Analysis...– P.8.

²³⁶ Macnamara J.R. Media Content Analysis...– 24 p.; Alexa M. Computer-assisted text analysis methodology in the social sciences...– 40 p. ; Alexa M., Zuell C. Commonalities, differences and limitations of text analysis Software...– 29 p.

²³⁷ Seidel J. Method and Madness in the Application of Computer Technology to Qualitative Data Analysis // Using Computers in Qualitative Research / N.Fielding, R.M.Lee (editors).– London: Sage, 1991.

– використання програм призведе до зростання однорідності (одноманітності) в методах аналізу, що особливо негативно позначиться на якісних дослідженнях²³⁸.

До цього додається: комп'ютер лише ідентифікує слова, а від ідентифікації слів до ідентифікації ідей, які ці слова представляють, дуже далеко.

Особливої ваги набули проблеми забезпечення валідності і надійності результатів:

- труднощі врахування контексту;
- нездатність програми розпізнати комунікативні інтенції слів;
- нездатність дослідника забезпечити вичерпне внесення в список ключових слів для певних категорій;
- нездатність розв'язати проблеми посилань перед або після слів, які з'являються в довільному місці тексту, зокрема проблема займенників;
- нездатність програм визначати межі одиниць аналізу, насамперед при якісному аналізі;
- за чисельними характеристиками, які вираховуються програмами, може втрачатися сутність категорій²³⁹.

Привертається увага і до обмежень, одним з яких є так звана "ціна" комп'ютеризації, під якою Р.Морріс розуміє:

- вартість комп'ютера достатньої для роботи конфігурації;
- час і зусилля, витрачені на вивчення роботи програм аналізу;
- час і зусилля, витрачені на створення словників категорій аналізу;
- час і зусилля, витрачені на тестування, перегляду і підтвердження категорій.

Така "ціна" робить комп'ютеризований контент-аналіз надто дорогим для досліджень, наприклад, невеликих масивів текстів.

²³⁸ Barry C.A. Choosing Qualitative Data Analysis Software: Atlas/ti and Nudist Compared // Sociological Research Online.– 1998.– Vol. 3.– No.3.– <http://www.socresonline.org.uk/socresonline/3/3/4.html> (2004.05.14)

²³⁹ Morris R. Computerized content analysis in management research: a demonstration of advantages & limitations...

На протигагу скептичному ставленню, прихильники комп'ютеризованого контент-аналізу виставляють свої потужні аргументи. Насамперед це те, що завдяки комп'ютеру контент-аналіз став для дослідників набагато доступнішим, ніж будь-коли, а можливості аналітика значно зросли. Ось як з цього приводу міркує Р.Морріс: "хоча комп'ютери самі не можуть виробляти рішення, однак їхня здатність маніпулювати текстом шляхом використання програм контент-аналізу дає змогу досліднику поглянути на дані з іншої точки зору, приводячи до вироблення висновків щодо змісту комунікацій"²⁴⁰. Одна лише можливість у процесі діалогу вибудовувати концептуальну схему аналізу є важливою сама по собі, адже рекурсивний процес вироблення, уточнення і знищення зайвих категорій долучає дослідника до процесу відкриття, сприяє поліпшенню розуміння й усвідомлення ним сутності фрагмента дійсності, представленого текстом.

Можна стверджувати, що програми контент-аналізу виступають не лише в ролі одного з інструментів аналізу, а стають повноцінним експертом, з чією думкою слід рахуватися. Такі дії, як адміністрування текстів, швидке проведення розрахунків, швидкий і зручний пошук і класифікація даних, формування та виведення на замовлення в заданій формі результатів досліджень тощо, відносять їх до інструментарію аналізу. Однак інтерпретаційні, графічні та статичні можливості, здатність формувати гіпотези і готувати варіанти висновків, перетворюють програми контент-аналізу в серйозного експерта, який пропонує свою об'єктивну оцінку результатів і виводить на варіанти рішень. Істотна перевага такого експерта проявляється на дуже великих масивах текстів, коли результати отримуються не лише швидко і в рамках допустимого бюджету, але й робляться узагальнення і висновки, які людина при таких масштабах виробити не в змозі і надійність яких гарантується довершеними правилами кодування.

На цьому фоні символічним є створення в різних країнах цілого ряду науково-дослідних центрів, які спеціалізуються на

²⁴⁰ Morris R. Computerized content analysis in management research: a demonstration of advantages & limitations...

комп'ютерному аналізу текстів. Крім названих раніше, до таких можна віднести університетські наукові центри Centre for Computer Assisted Qualitative Data Analysis Software (м.Суррей, Великобританія), Centre for Social Anthropology and Computers (м.Кент, Великобританія), добре відомі центри ZUMA – Zentrum für Umfragen Methoden und Analysen (м.Мангейм, Німеччина), Qualitative Solutions and Research (м.Ла Троб, Австралія).

4.4. Сучасні технології контент-аналізу і особливості Text Mining

Перші програми кількісного контент-аналізу зосереджували увагу в основному на підрахунку частот певних характеристик тексту. Більшість сучасних програм контент-аналізу також обмежені обробкою тексту, проте їхні можливості набагато ширші:

- зберігання даних та управління ними (дозволяє зберігати у формі текстових першоджерел або в спеціальному форматі, а також зберігати різноманітний аудіовізуальний матеріал: фотографії, діаграми, відео- та аудіозаписи, зв'язки з Web-сторінками; додатково аналітик має змогу анотувати, редагувати тексти, автоматично індексувати й записувати власну супроводжувальну інформацію);

- пошук даних (програми дозволяють шукати текстові дані за вказаними словами чи фразами, підраховувати частоту відповідних слів, шукати інформацію за заданим контекстом, а також різноманітними додатковими даними на зразок дати, хто брав інтерв'ю, звідки з'явилися дані тощо);

- кодування (процес кодування робиться відносно простим; є можливість робити окремі дані більш істотними та відносити їх до певної категорії, називати, об'єднувати і розділяти категорії, формувати концептуальну схему для розвитку теорії);

- розвиток і перевірка теорії (дозволяють застосовувати різноманітні теоретичні моделі для побудови теорій і представлення результатів);

- написання звітів (дозволяє готувати звіти для різноманітних категорій або відтворювати відповідні фрагменти

документів у формі цитат, таблиць, графічних зображень тощо; є змога формувати в програмі "журнал", у який можна записувати пояснення, ідеї, що виникають, та виводити його на друк або у файл)²⁴¹.

З появою комп'ютерного контент-аналізу змінився спосіб його використання: завдяки можливостям швидко переробляти величезні масиви інформації, контент-аналіз органічно влітається в загальну технологію, у рамках якої він застосовується. Наведемо приклад простого, однак технологічного використання контент-аналізу в Генеральному контрольному управлінні США GAO (United States General Accounting Office)²⁴². Насамперед контент-аналіз став хорошим інструментарієм для аналізу поточної інформації. Робочі папери, документи, розшифровки стенограм зустрічей, попередніх оцінок тощо з місцевих представництв GAO містять багато корисної інформації, проте її важко об'єднати та узагальнити через різноманітність і неструктурованість. Для структуризації текстових матеріалів складають загальний список проблем, які містяться в них, і підраховують частоту їхньої появи. З цим добре справляється контент-аналіз. Спрощення і категоризація також є частиною контент-аналізу. Наприклад, оцінюючи ступінь дублювання роботи, аналітики зібрали інформацію про бюджети, інтерв'ю, звіти і, унаслідок, було виділено 31 проблему, які стосувалися розміщення та будівництва житла і розвитку міст. Аналогічно, розглянувши 38 звітів з двох офісів, була вироблена система категорій, яку почали застосовувати до кожного звіту, що дозволило виявити схожі проблеми в офісах GAO.

Однак комп'ютерний контент-аналіз дуже часто призводить до більш радикальних змін, унаслідок чого технологічний ланцюжок "джерело тексту – текст – контент-аналіз"

²⁴¹ Lacey A., Luff D. Trent Focus for Research and Development in Primary Health Care: An Introduction to Qualitative Analysis.– Trent Focus, 2001.– <http://www.trentfocus.org.uk/Resources/Qualitative%20Data%20Analysis.pdf>

²⁴² Spisak T. Content Analysis: A Methodology for Structuring and Analyzing Written Material / GAO (United States General Accounting Office).– Transfer Paper 10.1.3.– Washington, D.C. – March 1989.– P.6.– http://www.coe.uga.edu/~cwise/content_analysis/GAO_methodology.pdf

замінюється простішим "джерело тексту – контент-аналіз", тобто контент-аналіз стає складовою певної технології. Прикладом такого технологічного використання можуть бути системи автоматичного розпізнавання в реальному часі текстів радіопередач з автоматичним підбором потрібної інформації. В основу цих систем убудовано контент-аналіз. Так, система WASABI (Watson Automatic Stream Analysis for Broadcast Information) вхідний звуковий потік через підсистему розпізнавання мови перетворює в текст, використовує різноманітні аналізатори для ідентифікації інформаційних елементів, автоматично генерує запити на основі цих інформаційних елементів і підбирає з інформаційної бази дані, релевантні поточному дискурсу²⁴³.

З іншого боку, комп'ютерний контент-аналіз не лише став утіленням традиційної методики, обмеженої обробкою тексту. Він має власну логіку розвитку – технологічну. Саме ця логіка розвинулася в програмах четвертого покоління. До програм четвертого покоління будемо відносити програми, які, будучи втіленням контент-аналізу, "убудовують" його в інші технології.

Прикладом технології цього покоління, що, значною мірою, побудована на принципах контент-аналізу і дістала застосування вже з середини 1990 рр., є технологія "видобування" даних або Text Mining (більш повна назва – Text Analysis and Knowledge Mining System). Text Mining – це алгоритмічне виявлення на основі статистичного і лінгвістичного аналізу, а також штучного інтелекту раніше невідомих зв'язків і кореляцій у вже існуючих неструктурованих текстових даних для проведення значеннєвого аналізу, забезпечення навігації і пошуку в неструктурованих текстах з кінцевою метою одержання нової цінної інформації – знань. Text Mining являє собою логічне продовження і поєднання цілого ряду методик і методів, зокрема технології Data Mining, контент-аналізу, статистичного аналізу тощо.

²⁴³ Brown E.W., Srinivasan S., Coden A., Ponceleon D., Cooper J.W., Amir A. Toward speech as a knowledge resource // IBM Knowledge Management.– Vol. 40.– Number 4.– 2001. – <http://www.research.ibm.com/journal/sj/404/brown.html>

Не лише Text Mining, а й її першооснова Data Mining мають надто коротку історію. Чимало науковців (зокрема²⁴⁴) вважають, що вони беруть свій початок від програм видобування інформації і близьких до них. Прикладом ранньої програми видобування інформації, який наводить М.Діксон в²⁴⁵, є написана в 1982 р. програма FRUMP. Вона використовувала ряд сценаріїв, які могли сканувати і опрацьовувати масиви новин, намагаючись на їхній основі робити описи подій.

До важливих піонерських досліджень з Text Mining М.Діксон відносить два. Насамперед, це роботи дослідницької групи з Гельсінського університету²⁴⁶, яка намагалася використати технологію Data Mining до неструктурованих попередньо необроблених текстових масивів. По-друге, це роботи Р.Фельдмана²⁴⁷, в основу яких покладено встановлення значущих для тексту понять (концептів) і визначення взаємозв'язку між документами і цими поняттями, тобто фактично здійснення класифікації тексту²⁴⁸. Система Document Explorer, запропонована Р.Фельдманом, спочатку будує базу даних на основі сукупності досліджуваних документів різних видів, у тому числі з Інтернет, а потім аналізує їх, використовуючи техніку видобування знань і графічний підхід. Зауважимо, що Р.Фельдман для позначення технології видобування інформації з документів замість Text Mining використовує, на його думку, більш точний термін Knowledge Discovery in Textual Databases (відкриття знань на основі текстових баз даних).

²⁴⁴ Dixon M. An Overview of Document Mining Technology.– October 4, 1997.– <http://www.geocities.com/ResearchTriangle/Thinktank/1997/mark/writings/dm.html> ; Wilks Y. Information extraction as a core language technology // Information Extraction: A Multidisciplinary Approach to an Emergine Information Technology.– Vol. 1299.– June 1997.– PP.1-9.

²⁴⁵ Dixon M. An Overview of Document Mining Technology...

²⁴⁶ Ahonen H., Heinonen O., Klemettinen M., Verkamo A.I. Mining in the phrasal frontier // Proceedings of PKDD'97 / 1st European Symposium on Principles of Data Mining and Knowledge Discovery.– Norway.– Trondheim.– June 1997.

²⁴⁷ Feldman R., Klosgen W., Ben-Yehuda Y., Kedar G., Reznikov V. Pattern based browsing in document collections // Principles of data mining and knowledge discovery. – June 1997.– Vol. 1263.– PP.112-122.

²⁴⁸ Dixon M. An Overview of Document Mining Technology...

Піонерськими роботами, у яких розглядається техніка видобування інформації з HTML-ресурсів, стали дослідження О.Етціоні²⁴⁹.

Технологія Text Mining відрізняється від Data Mining тим, що вона, на відміну від останньої, яка працює зі структурованими базами даних фактів, видобуває зразки (шаблони) зі звичайних текстів, призначених для читання людьми, а не комп'ютерами. Водночас, як і більшість когнітивних технологій, Text Mining – це не просто пошук серед великих масивів готової, кимсь іншим уже створеної інформації, а насамперед виявлення раніше невідомої і ніде не записаної інформації, а точніше, алгоритмічне виявлення раніше непомічених зв'язків як в самих текстах, так і внаслідок їх спільного читання. Крім того, часто на початку досліджень аналітик сам достеменно не знає, яка конкретно інформація йому потрібна та де її шукати.

Починаючи із середини 1990 рр., як напрям аналізу неструктурованих даних технологія Text Mining взяла за основу не лише методи класичного видобування знань, але й останні досягнення контент-аналізу: класифікацію, кластеризацію, виділення понять, фактів, шаблонів тощо.

Технологічними компонентами Text Mining є:

- 1) інформаційний пошук (відбір релевантних записів або текстових баз даних для наступного опрацювання);
- 2) інформаційна переробка (виділення зразків на основі відібраних даних);
- 3) інформаційна інтеграція (поєднання комп'ютерного виведення інформації з пізнавальними можливостями людини)²⁵⁰.

Саме при реалізації другої компоненти Text Mining здійснює такі види аналізу, як:

²⁴⁹ Etzioni O. The world-wide web: Quagmire or gold mine? // Communications of the ACM.– No 39(11).– November 1996.– PP.65-68.

²⁵⁰ Kostoff R. Text Mining for Global Technology Watch / Office of Naval Research (ONR) Science & Technology.– 2001.–

http://www.onr.navy.mil/sci_tech/special/technowatch/default.htm ; Kostoff R. Information Extraction From Scientific Literature with Text Mining / Office of Naval Research (ONR) Science & Technology.– 2001.–
http://www.onr.navy.mil/sci_tech/special/technowatch/default.htm

– виявлення або видобування інформації (цей аналіз передусе всій решті видів комп'ютерного аналізу неструктурованих текстів і служить основою для їхньої реалізації; у процесі аналізу ідентифікуються ключові фрази і зв'язки всередині тексту шляхом пошуку наперед визначених послідовностей слів у тексті, які називають шаблоном, або взірцем; виявлення інформації є надзвичайно корисним при роботі з великими обсягами інформації);

– відслідковування категорій або тем (визначаючи і зберігаючи деякі параметри користувача, наприклад, якого виду документи він переглядав раніше, які теми і поняття присутні в документах, що цікавлять користувача, система знаходить схожі інформаційні матеріали і регулярно передає їх для нього; прикладом зі сфери бізнесу може бути відстежування матеріалів, які містять новини про конкурентів і їхню продукцію або поширення власної продукції фірми);

– резюмування або реферування документів (основна ідея підсумовування полягає в тому, щоб за рахунок вилучення неістотних деталей і скорочення довжини тексту виділити головні моменти і загальний зміст; програма має здатність легко ідентифікувати людей, місця, події, час, однак найскладнішим є вміння аналізувати семантику повідомлення і давати йому змістовну інтерпретацію; засоби підсумовування є корисними, насамперед у тих випадках, коли користувачеві потрібно негайно оцінити, наскільки документ відповідає потребам користувача і чи варто витратити час для подальшої роботи з документом);

– класифікація або категоризація тексту (основна ідея аналізу є віднесення об'єктів з текстових масивів до заздалегідь визначених категорій; класифікуючи об'єкти, програма навіть не намагається аналізувати їх, а розглядає як звичайний набір слів, підраховує їхню частоту і на цій основі ідентифікує основні теми документів; далі можливе групування документів у Intranet-мережах і на Web-сайтах, розміщення їх у визначених папках, сортування повідомлень електронної пошти, відслідковування і вибіркоче поширення тематичної інформації користувачам тощо);

– кластеризація або групування (виділення компактних підгруп об'єктів з близькими властивостями; даний вид аналізу дуже схожий до класифікації, однак тут немає наперед визначених категорій, вони формуються на основі самих даних; інша вигідна перевага кластеризації полягає в тому, що документ може одночасно стосуватися багатьох категорій: алгоритм кластеризації для кожного документа формує цілу ієрархію категорій і вимірює вагу, з якою документ можна співвіднести з кожною категорією; відтак на цій основі здійснюється: побудова базису посилань від документа до документа, заснованих на вагах і спільному вживанні заданих ключових слів; реферування великих документальних масивів; визначення взаємозалежних груп документів; спрощення процесу перегляду при пошуку необхідної інформації, вилучення унікальних документів з колекції; виявлення дублікатів або близьких за змістом документів тощо);

– прогнозування (передбачення за значеннями одних ознак об'єкта значень інших);

– знаходження винятків (пошук об'єктів, які своїми характеристиками сильно виділяються на тлі загальної маси);

– пошук пов'язаних ознак, полів, понять окремих документів (цей інструментарій пов'язує документи спільними для них категоріями і допомагає користувачу не просто знаходити інформацію, як це робиться в традиційних методах пошуку, а безпосередньо переходити від одних понять до інших, пов'язаних з першими і не обов'язково в межах одного документа; більше того, якщо між категоріями X і Y , а також між Y і Z взаємозв'язки добре відомі, а через великі обсяги інформації дослідник може не побачити зв'язку між X і Z , зокрема, коли він лише формується або є дуже слабким, то такого сорту асоціативний механізм є надзвичайно цінним);

– візуалізація даних (засіб представлення змісту всього масиву документів і реалізації навігаційного механізму для дослідження документів і класів документів; візуалізація дозволяє представляти великі за обсягом документи у вигляді "інтерактивних" графічних зображень або мап, коли в результаті "взаємодії" з ними користувач має змогу переглядати їх з допомогою простих засобів пошуку, масштабувати дані,

створювати різноманітні підсхеми; візуалізація особливо корисна при дослідженні великих масивів документів і дослідженні концептуально взаємозв'язаних документів);

– відповіді на запитання або Q&A (намагається віднайти найкращу відповідь на задане запитання, як правило, записане природною мовою; при цьому використовуються всі різновиди аналізу Text Mining, самі ж запитання стосуються таких категорій, як люди, місця події тощо, окремо розробляються алгоритми для FAQ's – запитань, які часто задаються; підкреслимо, що технологія підготовки відповідей на запитання використовується багатьма Web-сайтами).

Приклад моделі²⁵¹ "видобування" даних представлено на рис. 4.2. Маючи на вході масив документів, програма Text Mining відбирає потрібні екземпляри документів і попередньо опрацьовує їх, перевіряючи формати і набори символів. Далі має місце етап текстового аналізу, під час якого шляхом повторення певних дій відбувається видобування інформації. На відповідній діаграмі показано лише три методи, однак їх є значно більше. Зрозуміло, що використовуються вони не всі, а лише ті, що потрібні для досягнення цілей. Підсумкова інформація записується в базу даних, звідки користувач отримує потрібні знання.

Наведені в табл. 4.2 і 4.3 дані показують, які методи Text Mining використовуються різними комерційними програмами та в різних сферах людської діяльності.

²⁵¹ Fan W., Wallace L., Rich S., Zhang Z. Tapping into the Power of Text Mining // Communications of ACM.– February 16, 2005.–
http://filebox.vt.edu/users/wfan/paper/text_mining_final_preprint.pdf

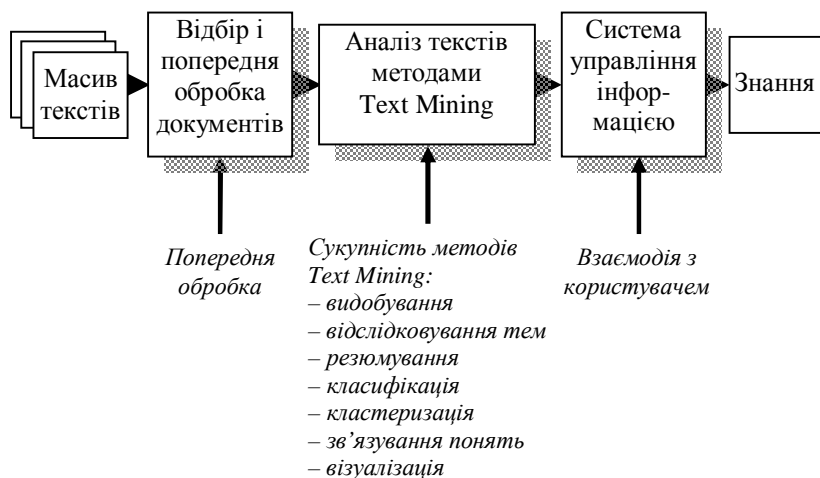


Рис. 4.2. Типова модель процесу "видобування" знань

Таблиця 4.2

Перелік методів, які пропонуються в програмах Text Mining

Методи Text Mining	Фірма-виробник програм Text Mining							
	Inxight	Autonomy	Clearforest	SAS	Convera	Megaputer	SPSS	IBM
Видобування інформації	+	+	+	+	+	+	+	+
Відслідковування тем		+						
Підсумовування (узагальнення)	+	+			+	+		+
Виділення понять	+	+	+	+	+	+	+	+
Зв'язування ознак		+	+	+				
Кластеризація		+			+	+		+
Візуалізація інформації	+						+	
Відповіді на запитання		+				+		

Сфери людської діяльності	Методи Text Mining							
	Видобування інформації	Відслідковування тем	Підсумовування (узагальнення)	Виділення понять	Зв'язування ознак	Класифікація	Візуалізація інформації	Відповіді на запитання
– Персоналізація змісту		+			+			
<i>Урядові служби</i>								
– Державна безпека: боротьба з тероризмом	+	+			+	+	+	
– Боротьба зі злочинністю: виявлення / попередження злочинів	+	+			+	+	+	
<i>Освіта</i>								
– Тематичні дослідження		+	+	+				
– Аналіз цитат	+				+		+	
– Питання, які найчастіше задаються (FAQ's)	+			+	+			+

З точки зору формату представлення даних, спочатку відбувається очищення тексту, під час якого неструктуровані документи перетворюються в стандартизовану проміжну форму, а потім здійснюється дистилляція знань, коли на основі проміжної форми виводяться зразки або знання²⁵². Проміжна

²⁵² Tan A.-H. Text Mining: The state of the art and the challenges // The Pacific Asia Conference on Knowledge Discovery and Data Mining PAKDD'99 workshop on

форма може бути документо- або концептуально-базована. У першому випадку зразки або знання виділяються, виходячи з документів, у другому – істотна інформація відповідає концептуальному представленню або образу проблемної ситуації (рис. 4.3).

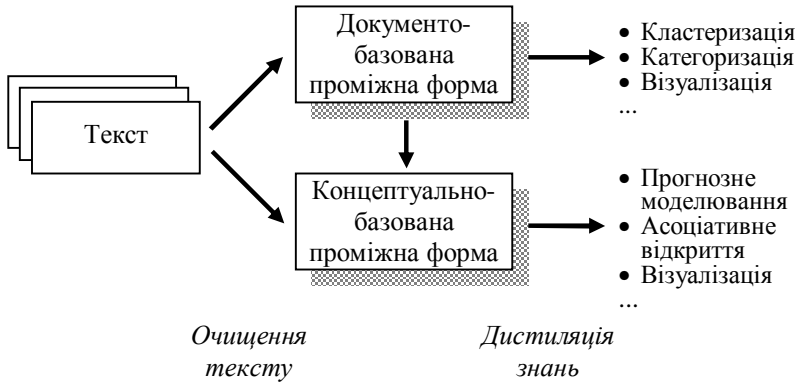


Рис. 4.3. Схема формування нових знань в Text Mining

На сьогоднішній день чимало виробників програмного забезпечення пропонують свої технології і програми Text Mining. Системи Text Mining, як правило, реалізуються у вигляді масштабних систем зі складними математичними і лінгвістичними алгоритмами аналізу, для яких характерним є розвинений графічний інтерфейс, багаті можливості візуалізації та маніпулювання даними, доступ до різних джерел даних, функціонування в архітектурі клієнт-сервер. За даними Центру технологічної політики і оцінювання (Technology Policy and Assessment Center – TPAC) Джорджіанського інституту технологій, станом на кінець 2000 р. в Інтернеті було представлено понад 70 інструментальних систем Text Mining²⁵³.

Knowledge Discovery from Advanced Databases.–1999.–PP.65-70.–
<http://citeseer.ist.psu.edu/tan99text.html>

²⁵³ Text Mining: Review of TPAC Technologies for ONR // ASDL.– Aug. 2002.–
http://www.asdl.gatech.edu/research_teams/pdf/2002/Text%20Mining%20Sum.doc

Опис функціональних можливостей окремих систем можна знайти в додатку В, а також у роботах Д.Ланде²⁵⁴.

Одним із найперспективніших напрямів узагальнення інформаційних потоків є контент-моніторинг. У найпростішому вигляді його ідею можна сформулювати як постійно здійснюваний в часі контент-аналіз неперервних інформаційних потоків. Серед принципів побудови системи моніторингу виділяють системність, адресність і предметну спрямованість. У ній текстовий потік досліджується на підставі заданих конфігураційних характеристик (наборів кількісних параметрів або слів, що супроводжують визначені теми і поняття). Він опрацьовується багаторазово, з додаванням характеристик, отриманих із самого потоку. Методологічну основу дослідження складає контент-аналіз. У результаті генерується, а потім наочно відображається узагальнена інформація. З появою систем Text Mining контент-моніторинг отримав реальну і потужну програмну основу.

До сучасних і перспективних напрямів використання Text Mining також відносяться:

- пошук всеохоплюючої і релевантної інформації на основі текстових баз даних;
- визначення інфраструктури заданих технологічних і наукових дисциплін і напрямів;
- здійснення тематичної структуризації певних сфер діяльності та взаємозв'язку між темами;
- виявлення нових напрямів досліджень, появи нових ідей у рамках певних дисциплін і на стику дисциплін;
- прогнозування технологічного розвитку²⁵⁵.

Як відзначається в²⁵⁶, особливої уваги заслуговує інноваційне прогнозування. Один із різновидів прогнозів базується на бібліометриці: підраховуючи число публікацій,

²⁵⁴ Ландэ Д. Добыча знаний // CHIP Ukraine. – 2003. – №10; Ландэ Д. Об отделении зерен от плевел // Мой Компьютер Weekly. – №33(256). – 25.08.2003.

²⁵⁵ Text Mining: Review of TPAC Technologies for ONR...

²⁵⁶ Watts R.J., Porter A.L. Innovation Forecasting // Technology Policy and Assessment Center (TPAC) at Georgia Institute of Technology. – 2002. – <http://www.tpac.gatech.edu/toa/innov.shtml>

патентів, відповідних згадувань у виступах науковців, можна виміряти та інтерпретувати напрями технологічного розвитку. Ці вимірювання узагальнюються у формі інноваційних індикаторів технологій, які можуть свідчити про стадію життєвого циклу технологій, інноваційний контекстуальний вплив даної технології на інші, розвиток ринкового потенціалу тощо. Ще один метод прогнозування базується на картографії новинних повідомлень, що дозволяє ідентифікувати споріднені групи технологій і ресурсів, взаємовпливи різних груп технологій, локалізувати дослідницькі домени та встановлювати коло їхніх інтересів.

Надзвичайно перспективним напрямом Text Mining є технологічна конкурентна розвідка (Competitive Technological Intelligence). Її значення і використання особливо зросло в 1990 рр., коли посилилася технологічна конкуренція, і компанії, університети та урядові організації відчували особливу потребу в знаннях про нові і перспективні технології. Значна частина результатів технологічною розвідкою отримується на основі пошуків з використанням технологій Text Mining.

Актуальність розробок Text Mining добре засвідчує застосування їх Федеральними службами і агентствами США. Так, дослідження GAO від травня 2004 р.²⁵⁷ засвідчило, що з 128 служб, які вивчалися, 52 використовували або планували використовувати технології Data Mining і Text Mining. Цілі їх використання надзвичайно різноманітні: починаючи від поліпшення обслуговування населення і завершуючи аналізом і виявленням терористичної і злочинної діяльності. Аналітиками GAO виявлено 199 випадків застосувань технології видобування знань, з яких 68 пов'язані з плануванням роботи і 131 – з оперативною діяльністю. Наведемо конкретні факти.

²⁵⁷ Data Mining. Federal Efforts Cover a Wide Range of Uses: Report to the Ranking Minority Member, Subcommittee on Financial Management, the Budget, and International Security, Committee on Governmental Affairs, U.S. Senate / GAO (United States General Accounting Office).– GAO-04-548.– Washington, D.C. – May 2004.– 71 p.– http://www.epic.org/privacy/profiling/gao_dm_rpt.pdf

У березні 2001 р. в ряді російських і українських інтернет-видань²⁵⁸ з'явилися повідомлення про використання Управлінням розвитку інформаційних технологій, яке є частиною директорату науки і технології ЦРУ США, Text Mining для роботи з відкритими джерелами інформації. Окремі вітчизняні публікації відсилають до першоджерела – матеріалу на сайті "Вашингтон пост". Мова йшла про застосування розвідувальним відомством трьох комп'ютерних систем – Oasis, FLUENT, Text Data Mining.

Перша система пов'язана з медіа-моніторингом як систематичних, так і випадкових джерел, які охоплюють друковані видання, цифрові матеріали, графічні зображення, аудіоінформацію 35 мовами світу. Наприклад, під час роботи з аудіоінформацією система Oasis перетворює звукові матеріали в текст, розпізнаючи мову, чоловічі та жіночі голоси, голоси різних людей, і записує їх у формі діалогів. Крім того, методика дозволяє виділяти з потоку лише ті голоси і ту інформацію, які закладені в настройках системи. На час появи зазначених інтернет-матеріалів система обмежувалася лише англомовною версією, хоча передбачалося створити версію для розпізнавання китайської, арабської й інших мов.

Комп'ютерна технологія FLUENT призначена для пошуку інформації в текстових документах. Маючи на вході ключові слова англійською мовою, система тут же перекладає їх на ряд інших мов, шукає інформацію в текстових базах даних з документами на різних мовах і повертає аналітику результати пошуку після автоматичного перекладу. FLUENT дозволяє перекладати англійською мовою з китайської, корейської, португальської, російської, сербсько-хорватської, української та інших мов.

²⁵⁸ Ландэ Д. Добыча знаний...; ЦРУ начинает просеивать информацию // Сетевой журнал. Лента новостей.– Вып. от 14.03.2001.– <http://www.setevoi.ru/cgi-bin/srch.pl?id=579> ; ЦРУ извлекает данные // Компьютер-информ.– 2001.– №6.– http://www.ci.ru/inform06_01/p245moz.htm ; Гордиенко И. Понять и заставить // Компьютерра. – 10.04.2001.– <http://www.ibusiness.ru/offline/2001/158/8585/print.html>

Ще одна програма, Text Data Mining, дозволяє автоматично створювати візуальні образи текстових документів, а також отримувати дані про частоту використання тих або інших слів.

Перелічені технології ЦРУ використовує для відслідковування незаконних фінансових операцій і наркотрафіку.

Наведемо також ще один цікавий напрям застосування Text Mining. Це так звані інтегратори новин. В умовах зростаючих інформаційних потоків звичайні пошукові системи Інтернет не дозволяють ефективно відбирати потрібну інформацію з цих потоків. На думку Д.Ланде²⁵⁹, оптимальним вирішенням проблем орієнтації в новинній інформації з Інтернет є використання нових мережних служб – інтеграторів новин, які інтегрують інформаційні потоки, здійснюючи контент-моніторинг новин у Web-просторі як базу для своєї роботи. Наприклад, Northern Light Technology є клієнтом однієї з величезних служб збору новин COMTEX, що інтегрує ресурси солідних джерел, серед яких такі світові інформаційні агентства, як Associated Press, ИТАР-ТАСС, Синьхуа. Клієнтами COMTEX, у свою чергу, є десятки новинних служб: OneSource, Screaming Media, Vertical Net, CompuServe та інші. Технологія моніторингу і подальшої синдикації Інтернет-новин передбачає етапи: "навчання" програм збору інформації структурі відібраних джерел, сканування інформації, приведення її до внутрішньосистемного формату, класифікація, кластеризація, передача користувачам через різноманітні канали, у тому числі e-mail, WWW, Wap, SMS.

На завершення відзначимо, що технологія Text Mining знаходиться лише на початку своєї "кар'єри". Однак навіть зараз вона видаються дуже перспективною. Немала роль у цьому належить контент-аналізу.

²⁵⁹ Ланде Д. Системы мониторинга Internet-контента // Бизнес-регистрация. – 2002. – №2(8). – <http://www.biz-registr.com.ua>

ВИСНОВКИ

Таким чином, проведене дослідження засвідчує, що контент-аналіз у своєму розвитку пройшов шлях від наукового методу формалізованого аналізу змісту мас-медіа до широко застосовуваної високотехнологічної методики.

Як було показано, у процесі становлення контент-аналізу можна виділити такі етапи:

– його зародження (до 1920 рр. – характеризується описовою та інтуїтивною методологією, розвиваються різноманітні підходи до аналізу та порівняння текстів у інтерпретаційних контекстах, насамперед засобів масових комунікацій, ранній газетний аналіз, графологічний аналіз, аналіз мрій);

– формування основ "класичного" контент-аналізу (1920-1940 рр. – розвиваються систематичні основи кількісного контент-аналізу, знову ж таки в рамках масових комунікацій, розвиток відбувається практично разом з теорією і практикою пропаганди);

– міждисциплінарне розширення і диференціація (1950-1960 рр. – методологія не лише розвиває свої теоретичні основи, але й знаходить шлях до різноманітних дисциплін, насамперед лінгвістики, психології, соціології, історичних наук, мистецтва тощо);

– розвиток теоретичних основ та розширення практичного застосування (1970-1980 рр. – удосконалення і застосування різноманітних моделей зв'язку, аналіз невербальних комунікацій, а також розвиток нових різновидностей, зокрема якісного контент-аналізу, удосконалення методики, насамперед за рахунок використання нових можливостей комп'ютерної техніки);

– період "глобального" розвитку (від 1990 рр. – період пов'язаний із використанням контент-аналізу в практиці діяльності найрізноманітніших суб'єктів, починаючи від науковців, які ознайомлені з його методикою та свідомо її застосовують, і завершуючи пересічними користувачами

Інтернет, практична більшість з яких, мабуть, і не підозрюють, що пошук інформації для них здійснюють програми із убудованими елементами контент-аналізу).

Характеризуючи нинішній стан контент-аналізу, варто зауважити, що тепер не наука чи засоби масової інформації є найбільшими користувачами контент-аналізу, а державні та недержавні установи, політичні партії, аналітичні центри, комерційні структури, зацікавлені в здобутті нових знань. Нинішні контент-дослідження пов'язані з переробкою величезних текстових масивів на основі Інтернет-технологій та комп'ютерних технологій видобування знань на зразок Text mining і Web-mining, побудованих значною мірою на ідеях контент-аналізу. З ними ж пов'язана перспектива розвитку контент-аналізу найближчих років.

БІБЛІОГРАФІЯ

1. A Brief Biography of Calvin S.Hall.–
<http://mind.ucsc.edu/dreams/About/calvin.html>
2. Ahonen H., Heinonen O., Klemettinen M., Verkamo A.I. Mining in the phrasal frontier // Proceedings of PKDD'97 / 1st European Symposium on Principles of Data Mining and Knowledge Discovery.– Norway.– Trondheim.– June 1997.
3. Alexa M. Computer-assisted text analysis methodology in the social sciences / Zentrum für Umfragen Methoden und Analysen (ZUMA). ZUMA-Arbeitsbericht 97/07.– Mannheim (Germany): ZUMA, 1997.– 40 p.–
http://www.gesis.org/Publikationen/Berichte/ZUMA_Arbeitsberichte/97/97_07.pdf
4. Alexa M., Zuell C. Commonalities, differences and limitations of text analysis Software: The results of a review / Zentrum für Umfragen Methoden und Analysen (ZUMA). ZUMA-Arbeitsbericht 99/06.– Mannheim (Germany): ZUMA, 1999.– 29 p.–
http://www.gesis.org/Publikationen/Berichte/ZUMA_Arbeitsberichte/99/99_06.pdf
5. Barry C.A. Choosing Qualitative Data Analysis Software: Atlas/ti and Nudist Compared // Sociological Research Online.– 1998.– Vol. 3.– No.3.–
<http://www.socresonline.org.uk/socresonline/3/3/4.html> (2004.05.14)
6. Berelson B. Content Analysis in Communication Research. – Glencoe, Illinois: Free Press, 1952.
7. Bernard H.R., Ryan G. Text Analysis: Qualitative and Quantitative Methods // Handbook of Methods in Cultural Anthropology / H. Russell Bernard (ed).– Walnut Creek: Altamira Press, 1998.– P.595-646.–
nersp.nerdc.ufl.edu/~ufruss/documents/bernard_ryan.pdf
8. Bradley J., Rockwell G. TACTweb: workbook. –
<http://tactweb.humanities.mcmaster.ca/tactweb/doc/twintro.htm>
9. Brown E.W., Srinivasan S., Coden A., Ponceleon D., Cooper J.W., Amir A. Toward speech as a knowledge resource // IBM Knowledge Management.– Volume 40.– Number 4.– 2001. –
<http://www.research.ibm.com/journal/sj/404/brown.html>

10. Carley K. An approach for relating social structure to cognitive structure // *Journal of Mathematical Sociology*.– 1986.– No 12.– PP. 137-89.
11. Carley K. Formalizing the social expert's knowledge // *Sociological Methods and Research*.–Vol.17.– 1988.– PP. 165-232.
12. Carney T.F. *Content analysis: Techniques for systematic inference from communications*.– Winnipeg: University of Manitoba Press,1972.
13. Collocations, Colligations, and Phrases.–
http://www.ling.lancs.ac.uk/staff/andrew/aw-col_files/frame.htm
14. Content Analysis in Political.–
<http://www.janda.org/workshop/content%20analysis/seminalstudies.htm>
15. Content Analysis of Introductory Geology Textbooks / Projects for Multicultural and Interdisciplinary Study and Education.– <http://www.scsv.nevada.edu/~promise/content.html>
16. Content Analysis of Introductory Geology Textbooks.–
<http://www.scsv.nevada.edu/~promise/content.html>
17. Content Analysis: An Introduction.–
http://www.coe.uga.edu/~cwise/content_analysis/index.html
18. Dahl H. A quantitative study of a psychoanalysis // *Psychoanalysis and contemporary science* / Eds. Holt R.R., Peterfreund E.– New York: Macmillan, 1972.– PP. 237-257.
19. Data Mining. Federal Efforts Cover a Wide Range of Uses: Report to the Ranking Minority Member, Subcommittee on Financial Management, the Budget, and International Security, Committee on Governmental Affairs, U.S. Senate / GAO (United States General Accounting Office).– GAO-04-548.– Washington, D.C. – May 2004.– 71 p.–
http://www.epic.org/privacy/profiling/gao_dm_rpt.pdf
20. Department of Journalism Studies, The University of Sheffield, 2002. – <http://doj.shef.ac.uk/ResearchMethods-Lecture%20Eight.doc>
21. DeWeese L.C. Computer content analysis of day-old newspapers: a feasibility study // *Public Opinion Quarterly*.– Vol. 41.– 1977.– PP. 91-94.

22. DeWeese L.C. Computer content analysis of printed media: a limited feasibility study // *Public Opinion Quarterly*.– Vol. 40.– 1976.– PP. 92-100.
23. Dixon M. An Overview of Document Mining Technology.– October 4, 1997.–
<http://www.geocities.com/ResearchTriangle/Thinktank/1997/mark/writings/dm.html>
24. Dovring K. Quantitative Semantics in 18-th Century Sweden // *Public Opinion Quarterly*. – 1954-1955. – Vol.18. No.4.
25. Eaton B.C., Dominick J.R. Product related programming and children's TV // *Journalism Quarterly*. – 1991. – No 18. – pp. 67-75.
26. Etzioni O. The world-wide web: Quagmire or gold mine? // *Communications of the ACM*.– No 39(11).– November 1996.– PP.65-68.
27. Feldman R., Klosgen W., Ben-Yehuda Y., Kedar G., Reznikov V. Pattern based browsing in document collections // *Principles of data mining and knowledge discovery*. – June 1997.– Vol. 1263.– PP.112-122.
28. Gallagher C, Marken M., Kim M., Phillipson T., Dodge T. The Use of Content Analysis in Researching Adolescent Issues. – Indiana University, 2000. – 17 p. –
www.secretstage.com/other/web/messages/school/y520/monograph.pdf
29. Gerbner G. Violence in television drama: trends and symbolic functions // *Television and Social Behavior*. Vol. 1: Media Content and Control / Comstock G.A., Rubinstein E.A. (eds.). – Washington, DC: US Government printing office, 1972. – p. 31.
30. Gerbner G., Gross, L. Living with television: the violence profile // *Journal of Communication*.. –1976. – No 26. – pp. 173-199.
31. Gottschalk L.A., Kaplan S.M. A quantitative method of estimating variations in intensity of a psychologic conflict or state // *Archives of Neurology and Psychiatry*.– 1958.– No 79.– PP.688-696.
32. Grünzig H.J. Themes of anxiety as psychotherapeutic process variables // *Methodology in psychotherapeutic research* / Minsel W.R., W.Herff (eds.). Proceedings of the 1st European Conference on Psychotherapy Research, Trier, 1981. – Lang, Frankfurt, 1983.– PP. 135-142.

33. Hall C.S. Content Analysis of Dreams: Categories, Units, and Norms // The analysis of communication content / G.Gerbner (Ed.).– New York: Wiley, 1969.
34. Hall C.S. Normative Dream Content Studies // Dream Psychology and the New Biology of Dreaming / M.Kramer (Ed.).– Springfield, IL: Charles C.Thomas, 1969.– PP. 175-184.
35. Hall C.S., Van de Castle R. The Content Analysis of Dreams.– New York: Appleton-Century-Crofts, 1966.
36. Hamesse J. Automatic Processing of Philosophical Works at the Catholic University of Louvain // A.J.Aitken et al. (eds.) The Computer and Literary Studies. Edinburgh.– 1973.– P.341-344.
37. Harway N.I., Iker H.P. Computer analysis of content in psychotherapy // Psychological Reports.– 1964.– №14.– PP.720-722.
38. Hockey S.. An Historical Perspective in Rahts. Information technology in the Humanities. P.22.
39. Holbrook M.B. More on Content Analysis in Consumer Research // Journal of Consumer Research.– 1977.– Vol. 4.– PP.176-177.
40. Holsli R. Content analysis // Handbook of Social Psychology, 2-nd ed. Vol. 2 / Ed. by Zindzey Y., Aronson E.– 1968, – P. 596-692.
41. Holsti O.R. Content Analysis for the Social Sciences and Humanities.– Reading, MA: Addison-Wesley Publishing Company, 1969. – <http://www.questia.com/PM.qst?a=o&d=54363997>
42. Iker H.P., Harway N.I. A Computer Systems Approach Toward the Recognition and Analysis of Content // The Analysis of Communication Content / Gerbner G.A. et al. (eds.).– Wiley & Sons, 1969.
43. Jackson J. Content analysis and objectivity, language and metaphor.– 1998. – <http://www.spinworks.demon.co.uk/pub/content2.htm>
44. Kassarijan H. Content Analysis in Consumer Research // Journal of Consumer Research.– 1977.– Vol. 4.– June 8-18.
45. Kelle U. Computer-Aided Qualitative Data Analysis: An Overview // Text Analysis and Computers / Cornelia Zuell, Janet Harkness, Juergen H.P. Hoffmeyer-Zlotnik (Eds.). Zentrum für Umfragen Methoden und Analysen (ZUMA).– Mannheim (Germany): ZUMA, 1996.– P. 33-63.

46. Kohlbacher F. The Use of Qualitative Content Analysis in Case Study Research // Forum: Qualitative Social Research [On-line Journal].– 2005.– December, 7(1), Art. 21.– [http:// www.qualitative-research.net/fqs-texte/1-06/06-1-21-e.htm](http://www.qualitative-research.net/fqs-texte/1-06/06-1-21-e.htm)

47. Kostenko N.V., Ivanov V.F. Experience of Content Analysis. Models and Practices: Monograph.– K.: Centre for Free Press, 2005.– 234 p.

48. Kostoff R. Information Extraction From Scientific Literature with Text Mining / Office of Naval Research (ONR) Science & Technology.– 2001.– http://www.onr.navy.mil/sci_tech/special/technowatch/default.htm

49. Kostoff R. Text Mining for Global Technology Watch / Office of Naval Research (ONR) Science & Technology.– 2001.– http://www.onr.navy.mil/sci_tech/special/technowatch/default.htm

50. Krippendorff K. Content Analysis: An introduction to its Methodology / Vol. 5. The Sage Communication Text Series. – Beverly Hills: Sage Publications, 1980.

51. Lacey A., Luff D. Trent Focus for Research and Development in Primary Health Care: An Introduction to Qualitative Analysis.– Trent Focus, 2001.– <http://www.trentfocus.org.uk/Resources/Qualitative%20Data%20Analysis.pdf>

52. Lacy S., Fico F., Simon T.F. Fairness and balance in the prestige press // Journalism Quarterly. – 1991. – No 68(3). – pp. 363-370.

53. Laffal J. An approach to the total content analysis of speech in psychotherapy // Research in psychotherapy / J.M.Shlien (Ed.).– Washington: American Psychological Association, 1968.– PP.277-294.

54. Laffal J. Total or Selected Content Analysis // International Conference on Computational Linguistics, Sångå-Såby-Kurugård, Sweden, 1-4 September, 1969.– acl.ldc.upenn.edu/C/C69/C69-2401.pdf

55. Lasswell H.D., Lerner D., Pool I.D. The Comparative Study of Symbols // Hoover Institute Studies.– Series C: Symbols.– No 1.– Stanford, CA: Stanford University Press, 1952.

56. List D. Know Your Audience. Chapter 16. Content analysis. – August 2003.– <http://www.audiencedialogue.org/kya.html>

57. Lombard M., Snyder-Duch J., Bracken C. C. Content analysis in mass communication: Assessment and reporting of intercoder reliability // *Human Communication Research*.– 2002.– No 28.– PP.587-604.
58. Lombard M., Snyder-Duch J., Bracken C.C. Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects. – 2004.–
<http://www.temple.edu/mmc/reliability/>
59. Lord J.B., Eastlack J.O., Stanton J.L. Health Claims in Food Advertising: Is There a Bandwagon Effect? // *Journal of Advertising Research*.– 1987.– 27 April/May.
60. Luborsky L., Crits-Christoph P., Mintz J., Auerbach A. Who will benefit from psychotherapy? Predicting therapeutic outcomes.– New York: Basic Books, 1988.
61. Macnamara J.R. Media Content Analysis: Its Uses, Benefits & Best Practice Methodology / CARMA International (Asia Pacific).– Chippendale (Australia): CARMA, 2003.– 24 p.–
www.masscom.com.au/book/papers/media_content.html
62. Markoff J., Shapiro G., Weitman S. Toward the integration of content analysis and general methodology // *Sociological Methodology* / Heise D.R. (ed.).– San Francisco: Jossey-Bass, 1975.– PP.1-58.
63. Marshall R., Lu Z. A Content Analysis of Chinese Magazine Ads // *Journal of Advertising*.– 1988.– No 17(4).– PP. 43-48.
64. Martindale C. The therapist-as-fixed-effect fallacy in psychotherapy research // *Journal of Consulting and Clinical Psychology*.– 1978.– №6(46).– PP.1526-1530.
65. Mayring P. Qualitative Content Analysis // *Forum: Qualitative Social Research [On-line Journal]*.– 2000, June.– No 1(2).– <http://qualitative-research.net/fqs/fqs-e/2-00inhalt-e.htm>
66. Mctavish D.G., Pirro E.B. Contextual Content Analysis // *Quality & Quantity*.– 1990.– No 24.– PP. 245-265.–
www.clres.com/online-papers/qq.doc
67. Mergenthaler E. Computer-Assisted Content Analysis // *Text Analysis and Computers* / Cornelia Zuell, Janet Harkness, Juergen H.P. Hoffmeyer-Zlotnik (Eds.). Zentrum für Umfragen

Methoden und Analysen (ZUMA).– Mannheim (Germany): ZUMA, 1996.– P. 3-32.

68. Mergenthaler E. Textbank Systems. Computer science applied in the field of psychoanalysis.– Heidelberg: Springer, 1985.

69. Mergenthaler E., Kachele H. Changes of latent meaning structures in psychoanalysis // Sprache und Datenverarbeitung.– 1985.– №9.– PP.21-28.

70. Merten K. (1996) Reactivity in content analysis. Communications: The European Journal of Communication Research, 21(1), 65-76.

71. Miller M.M., Riechert B.P. Identifying Themes via Concept Mapping: A New Method of Content Analysis / University of Tennessee, Knoxville.– 1994.–

<http://excellent.com.utk.edu/%7Emmmiller/pestmaps.txt>

72. Mohler P., Zuell C. A Popperian Critique of Automatic Content Analysis // Journées Internationales d'Analyse Statistique des Données Textuelles / ZUMA.– 2000.– No 5.–

<http://www.cavi.univ-paris3.fr/lexicométrica/jadt/jadt2000/pdf/10/10.pdf>

73. Morris R. Computerized content analysis in management research: a demonstration of advantages & limitations // Journal of Management.– Winter.– 1994.–

http://www.findarticles.com/p/articles/mi_m4256/is_n4_v20/ai_16549030

74. Mustchin E. Explore the strengths and weaknesses of classical content analysis.– 1998. –

<http://www.spinworks.demon.co.uk/pub/content1.htm>

75. Mustonen A., Pulkkinen L. Television violence: A development of a coding scheme // Journal of Broadcasting and Electronic Media. – 1997. – No 41. – pp. 168-189.

76. Neuendorf K.A. Content Analysis Bibliography, 7/7/2000 // Neuendorf K.A. The Content Analysis Guidebook Online. – <http://academic.csuohio.edu/kneuendorf/content/bibs/bookbib.htm>

77. Neuendorf K.A. The Content Analysis Guidebook.– <http://academic.csuohio.edu/kneuendorf/content/resources/flowc.htm>

78. Neuendorf K.A. The Content Analysis Guidebook.– Thousand Oaks, CA: Sage Publications, 2002.

79. Neuman L.W. Social research methods: qualitative and quantitative approaches. Chapter 12. Nonreactive research and available data. – Boston: Allynand Bason, 1991.
80. Ogilvie D.M., Stone P.J., Schneidman E.S. Some Characteristics of Genuine Versus Simulated Suicide Notes // The General Inquirer: A Computer Approach to Content Analysis / P.J. Stone, D.C.Dunphy, M.S.Smith, D.M.Ogilvie, eds.– Cambridge: M.I.T. Press, 1966.– PP. 527–535.
81. Osgood Ch. The Representational Model and Relevant Research Methods // Trends in Content Analysis / Pool I. de Sola. (ed.).– Urbana, Ill.: University of Illinois Press, 1959.– PP. 33-88.
82. Osgood Ch., Saporta S., Nunnally J. Evaluative Assertion Analysis // *Litera*.– 1956.– Vol.3.– PP.47-102.
83. Overview: Content Analysis. –
<http://writing.colostate.edu/references/research/content/index.cfm>
84. Paisley W.J. Identifying the unknown communicator in painting, literature and music: the significance of minor encoding habits // *Journal of Communication*.– 1964.– No14.– PP.219-237.
85. Perez E. Managing text // *Data Based Advisor*.– 1990.– No 8(6).– PP. 83-93.
86. Plummer K. Documents of Life. An Introduction to the Problems and Literature of a Humanistic Method.– London: Unwin Hyman Ltd., 1990.– P.6.
87. Potter W.J., Linz D., Wilson B.J., Kunkel D., Donnerstein E., Smith S.L., Blumenthal E., Gray T. Content Analysis of Entertainment Television: New Methodological Developments // Duke University Conference on Media Violence and Public Policy in the Media, Durham, NC 27-29, June, 1996.
88. Qualitative social science research methodology.–
<http://faculty.ncwc.edu/toconnor/308/308lect09.htm>
89. Rao V., Woolcock M. Integrating Qualitative and Quantitative Approaches in Program Evaluation.– June 14, 2002.–
[http://Inweb18.worldbank.org/ESSD/essdext.nsf/61DocByUnid/8CD B76BF252A827E85256C0D0059AC4B/\\$FILE/Tool%20Kit%20Mixed%20Methods%20Chapter.doc](http://Inweb18.worldbank.org/ESSD/essdext.nsf/61DocByUnid/8CD B76BF252A827E85256C0D0059AC4B/$FILE/Tool%20Kit%20Mixed%20Methods%20Chapter.doc)
90. Riffe D., Lacy S., Fico F.G. Analyzing Media Messages: Using Quantitative Content Analysis in Research . – Mahwah, NJ: Lawrence Erlbaum Associates, 1998.– p.20.

91. Roberto Busa: biography / Humanities and information science. Roma Univ. La Sapienza (24/11/95).–
<http://www.mediamente.rai.it/mmold/english/bibliote/biografi/b/busa.htm>
92. Roberts C.W. Content Analysis // International Encyclopedia of the Social & Behavioral Sciences / Editors-in-Chief Neil J. Smelser, Paul B. Baltes.– Amsterdam: Elsevier Science Ltd, 2001.– P.2697-2702.
93. Roberts C.W. Other than counting words: A linguistic approach to content analysis // Social Forces.– 1989.– No 68.– PP.147-177.
94. Rosengren K.E. Advances in Content Analysis / Vol. 9, Sage annual reviews of communication research. – Beverly Hills: Sage Publications, 1981.
95. Seidel J. Method and Madness in the Application of Computer Technology to Qualitative Data Analysis // Using Computers in Qualitative Research / N.Fielding, R.M.Lee (editors).– London: Sage, 1991.
96. Simonton D.K. Thematic fame and melodic originality in classical music: a multivariate computer-content analysis // J. Pers.– 1980.– No 48. – PP.206-219.
97. Smythe D. Three Years of New York Television: 1951-1953. – Urbana, IL: National Association of Educational Broadcasters, 1954.
98. Spisak T. Content Analysis: A Methodology for Structuring and Analyzing Written Material / GAO (United States General Accounting Office).– Transfer Paper 10.1.3.– Washington, D.C. – March 1989.–
http://www.coe.uga.edu/~cwise/content_analysis/GAO_methodology.pdf
99. Stevenson R.L. Research Methodology // Foreign news and the new world information order / Robert L.Stevenson, Donald Lewis Shaw (Editors).– Ames, Iowa: The Iowa State University Press, 1986.– P.21-36.
100. Stone Ph., Dunphy D., Smith M., Ogilvie D. The General Inquirer: A Computer Approach to Content Analysis.– Cambridge, 1966.

101. Suzuki N. The Changing Pattern of Advertising Strategy by Japanese firms in the U.S. market: Content Analysis // *Journal of International Business Studies*.– 1980.– No 12.– PP.63-72.

102. Tan A.-H. Text Mining: The state of the art and the challenges // *The Pacific Asia Conference on Knowledge Discovery and Data Mining PAKDD'99 workshop on Knowledge Discovery from Advanced Databases*.–1999.–PP.65-70.–

<http://citeseer.ist.psu.edu/tan99text.html>

103. Text Mining: Review of TPAC Technologies for ONR // *ASDL*.– Aug. 2002.–

http://www.asdl.gatech.edu/research_teams/pdf/2002/Text%20Mining%20Sum.doc

104. The Analysis of communication content: Developments in scientific theories and computer techniques / Holsti R.O., Krippendorff K., Paisley W.J., Stone, P.J.(eds.) – New York: Wiley, 1969.

105. The Oxford Text Archive.–

<http://ota.ahds.ac.uk/ota/index.html>

106. Tombeur P. Research carried out at the Centre de Traitement Electronique des Documents of the Catholic University of Louvain // A.J.Aitken et al. (eds.) *The Computer and Literary Studies*. Edinburgh.– 1973.– P.335-340.

107. Watts R.J., Porter A.L. Innovation Forecasting // *Technology Policy and Assessment Center (TPAC) at Georgia Institute of Technology*.– 2002.–

<http://www.tpac.gatech.edu/toa/innov.shtml>

108. Wilks Y. Information extraction as a core language technology // *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*.– Vol. 1299.– June 1997.– PP.1-9.

109. Williams T.M., Zabrack M.L., Joy L.A. The portrayal of aggression on North American television // *Journal of Applied Social Psychology*. – 1982. – No 12 (5). – pp. 360-380.

110. Wilson A. Towards an Integration of Content Analysis and Discourse Analysis: The Automatic Linkage of Key Relations in Text / *April* 1993.–

<http://www.comp.lancs.ac.uk/computing/research/ucrel/papers/techpaper/vol3.pdf>

111. Wilson B.J., Smith S.L., Linz D., Potter J., Donnerstein E., Kunkel D., Blumenthal E., Gray T. Content Analysis of Entertainment Television: The 1994-95 Results // Duke University Conference on Media Violence and Public Policy in the Media, Durham, NC, 27-29 June, 1996.

112. Wimmer R.D., Dominick J.R. Mass Media Research: An Introduction. – Belmont, CA: Wadsworth Publishing Company, 1994.

113. Wong P.C., Whitney P., Pacific J.T. Visualizing Association Rules for Text Mining / Northwest National Laboratory.– 1999.–

<http://www.pnl.gov/infoviz/InfoVis1999Association.pdf>

114. Yahya H. Content analysis in human communication / JURNAL IAIN Sunan Ampel. Edisi XVII, Januari-Maret 2000. –

<http://www.geocities.com/HotSprings/6774/j-41.html>

115. Алексеев А.Н. Метод Жака Кайзера (Из опыта исследований французской ежедневной прессы) // Проблемы современной печати: Сб. ст. / Отв. ред. В.С.Соколов.– Л.: Изд-во Ленинградского ун-та, 1969.– С.76.

116. Арапов М.В. Пятибалльная шкала оценок: стоит ли менять? – <http://www.courier.com.ru/top/content/score.htm>

117. Воронько В., Костинский А. Компьютерный анализ текстов / Радио Свобода, 26-06-2001.–

<http://www.svoboda.org/hotnews/ticker.asp>

118. Гнатюк О.Л. Из истории американской коммуникологии и коммуникативистики: Гарольд Лассуелл (1902-1978) // Актуальные проблемы теории коммуникации: Сборник научных трудов.– СПб: Изд-во СПб ГПУ, 2004.– С. 11-20.– http://russcomm.ru/rca_biblio/g/gnatuk.shtml

119. Гордиенко И. Понять и заставить // Компьютерра. – 10.04.2001.–

<http://www.ibusiness.ru/offline/2001/158/8585/print.html>

120. Григорьев С.И., Растов Ю.Е. Основы современной социологии: Учебное пособие // Алтайский государственный университет. Центр мультимедиа технологий "Ирбис".– 2001.– <http://irbis.asu.ru/mmc/grig/4.1.ru.shtml>

121. Иголкин А.А. Масс-медиа в условиях глобализации // Вызовы глобализации: Ответы России и Кореи / Материалы

международного научного симпозиума, 5-6 июля 2001 г.– М.: РГГУ, 2001.– С. 59-72.–

<http://liber.rsu.ru/Conf/Globalization/globalization.pdf>

122. Исследование анонимных текстов на предмет выявления их авторов.– <http://bond007.h1.ru/new/glava12.htm>

123. Кармадонов О.А. "Символ" в эмпирических социальных исследованиях // Социс.– 2004.– №6.– http://socis.isras.ru/SocIsArticles/2004_06/karmadonov.doc

124. Ключкин А.И., Кончаков Р.Б., Сельцер Д.Г. Компьютерные методы в исследовании номенклатуры КПСС.– <http://elis.pstu.ac.ru/kljukin-selts.htm>

125. Ключкина О.Б. Построение теории на основе качественных данных (из истории развития методологического направления) // Социс.– 2000.– №10.– С.92-101.

126. Кононенко М.А. Контент-анализ как метод: аналитический обор. –

<http://liber.rsu.ru/Conf/Comparative/kononenko.htm>

127. Контент-анализ в информационном поле.– <http://taimer.wormhole.ru/discussion/ContentAnalysis.doc>

128. Коробейников В.С. Редакция и аудитория.– М., 1983.– С.140-141.

129. Костюк В.В. Анализ документов и текстов: Лекции по курсу / Белорусский Государственный Университет, 2003. – http://www.ffsn.bsu.by/infocom/kostuk/document_analysis/lectures.html

130. Ландэ Д. Добыча знаний // CHIP Ukraine.– 2003.– №10.

131. Ландэ Д. Об отделении зерен от плевел // Мой Компьютер Weekly. – №33(256). – 25.08.2003.

132. Ландэ Д. Системы мониторинга Internet-контента // Бизнес-регистр. – 2002. – №2(8). – <http://www.biz-registr.com.ua>

133. Мангейм Дж.Б., Рич Р.К. Политология. Методы исследования. – М.: Весь Мир, 1997. – 544 с.

134. Манекин Р.В. Контент-анализ как метод исторического исследования.–

<http://manekin.narod.ru/contentM.htm>

135. Методы сбора информации в социологических исследованиях. Кн.2. Организационно-методические проблемы

опроса. Анализ документов. Наблюдение. Эксперимент / Отв. ред. В.Г.Андреевков, О.М.Маслова. – М.: Наука, 1990. – 224 с.

136. Миронов Б. История в цифрах.– Л., 1991.– С.24-26.

137. Ньюман Л. Неопросные методы исследования // Социологические исследования. – 1998. – №6. – С.119-130.

138. Папа за использование компьютера на службе у христианской доктрины / Католическая Информационная Служба "AGNUZ".– Ватикан.– 1 февраля 2002 г. – <http://agnuz.info/print.php?year=2002&month1=February&day=1&files=101.txt&print=news>

139. Семенова В.В. Качественные методы: введение в гуманистическую социологию / Ин-т социологии РАН.– М.: Добросвет, 1998.– 289 с.

140. Страусс А., Корбин Дж. Основы качественного исследования: обоснованная теория, процедуры и техники / Пер. с англ.– М.: Эдиториал УРСС, 2001

141. Терин В.П. Основные направления исследований массовой коммуникации // Социологические исследования. – 1997. – №11. – С.25-31.

142. Федотова Л.Н. Анализ содержания – социологический метод изучения средств массовой коммуникации.– М.: Институт социологии РАН, 2001.– 202 с.

143. Федотова Л.Н. Контент-аналитические исследования средств массовой информации и пропаганды. Учебно-методическое пособие. – М.: МГУ, 1988. – 77 с.

144. ЦРУ извлекает данные // Компьютер-информ.– 2001.– №6.– http://www.ci.ru/inform06_01/p245moz.htm

145. ЦРУ начинает просеивать информацию // Сетевой журнал. Лента новостей.– Вып. от 14.03.2001.– <http://www.setevoi.ru/cgi-bin/srch.pl?id=579>

146. Юськів Б. Историчні витоки та логіка розвитку "класичного" контент-аналізу // Слов'янський вісник: Збірник наукових праць. Вип. 4. Серія "Історичні науки".– Рівне: РДГУ, 2004.– С.255-277.

ДОДАТОК А. Основні поняття і алгоритм кількісного контент-аналізу

Опис основних понять і алгоритму кількісного контент-аналізу в обробці і з доповненнями автора подається за структурою, запропонованою в роботі²⁶⁰.

1. Загальні принципи і ключові поняття

1) Систематичність:

– зміст аналізується згідно з явно описаними і послідовно застосовуваними правилами;

– вибірка формується з використанням належних процедур, наприклад, випадкова або ймовірнісна вибірка, багатоступенева вибірка, які достатньо добре представляють досліджуваний текстовий масив;

– кодування та аналіз відбуваються за єдиними правилами; кодування являє собою перетворення особливостей змісту в специфічні числові значення;

2) об'єктивність:

– особисті симпатії та ідіосинкразія дослідника не повинні впливати на результати дослідження;

– дослідження проведене одним дослідником повинно бути таким, щоб його могли відтворити інші;

– операційні означення і правила класифікації змісту повинні бути зрозумілими і достатньо добре поясненими, що дозволить різним кодувальникам досягти однакових результатів;

3) квантифікація:

– точно відтворювати числами основний зміст текстів;

– часто найважливішими в контент-аналізі є числові дані;

– числові оцінкові дані накопичуються, і це повинно робитися систематично і згідно зі строгими правилами, не допускаючи суб'єктивного сприйняття і тлумачення.

²⁶⁰ Department of Journalism Studies...

2. Кроки контент-аналізу

- 1) сформулюйте питання дослідження або гіпотезу;
- 2) визначте генеральну сукупність на основі питань;
- 3) визначте вибірку з генеральної сукупності для безпосереднього дослідження;
- 4) виберіть і дайте означення одиниць аналізу;
- 5) побудуйте систему категорії аналізу змісту;
- 6) встановіть систему квантифікації (кількісного оцінювання категорій);
- 7) проведіть навчання кодувальників і експериментальне (попереднє) дослідження;
- 8) закодуйте тексти згідно з установленими правилами кодування;
- 9) проаналізуйте отримані дані;
- 10) зробіть висновки і підготуйте пропозиції.

3. Визначення проблем, питань і гіпотез дослідження

Анкета для визначення проблем і питань дослідження:

- У чому полягає проблема або питання дослідження?
- Який феномен (об'єкт) буде вивчатися?
- Що ми сподіваємося отримати в результаті аналізу текстів? Контент-аналіз – це не теорія, а метод. І він не може вирішити за нас, що ми повинні аналізувати.
- Будь-якому контент-аналізу потрібна теоретична основа, Чи є можливість віднайти готову теорію для вашого дослідження?
- Як багато відомо про явище або тему, яка буде вивчатися? Дослідник мусить ознайомитися з існуючими фактами.
- Що повинно описувати дослідження: щось про різновиди змісту медіа; або можливі способи впливу на читачів; або способи виробництва медіапродукції?
- Може, дослідження намагається сказати щось про уявлення різних соціальних груп населення (наприклад, чоловіків і жінок, етнічних, вікових і професійних груп), або про соціальну поведінку (наприклад, насильство, здоров'я, захист

прав споживачів), або про інституції (наприклад, політичні, бізнесові)?

– Може, дослідження хоче сказати щось про редакційну політику видання чи способи подачі новин?

Рекомендації при формулюванні питань дослідження:

– підрахунок – це не самоціль, тому уникайте підрахунку заради підрахунку;

– кінцева мета аналізу повинна бути чітко сформульована;

– потрібно керуватися чітко сформульованими питаннями дослідження або гіпотезами;

– ознайомтеся з відповідною літературою;

– як правило, джерела запитань або гіпотези дослідження знаходять у літературі.

4. Визначення генеральної сукупності дослідження

Рекомендації:

– прийміть рішення щодо меж дослідження (визначитися, що знаходиться в полі зору дослідника);

– ідентифікуйте зміст аналізу;

– визначте схеми класифікації змісту;

– визначте період, який буде охоплювати дослідження.

5. Формування вибірки

Загальні рекомендації:

– після визначення генеральної сукупності (повного масиву текстів) потрібно сформувати вибірку, яку утворює певна частина генеральної сукупності (підмножина текстів), що будуть безпосередньо досліджуватися;

– якщо генеральна сукупність охоплює відносно невелику кількість текстів, то варто досліджувати всю генеральну сукупність (тобто вона виконує роль вибірки);

– якщо генеральна сукупність є дуже великою, то потрібно відібрати ту її частину, яку дослідник реально спроможний вивчити;

– найчастіше використовують багатоступенева вибірку;

– описи проблем вибираються для аналізу цілком випадково;

– у межах суцільної вибірки аналізуються всі матеріали джерела: кожна телепрограма, кожне число газети, що вийшло впродовж визначеного періоду. Іноді вона модифікується у "50-відсоткову". Вибіркова сукупність визначається як "суцільна" не лише за параметром часу. Велику роль відіграє і тематична спрямованість повідомлень предмета дослідження. Відповідно аналізується або весь зміст числа чи "телевізійного дня", або лише окремі рубрики, сюжети, матеріали, які безпосередньо стосуються предмета вивчення – приміром, міжнародних відносин, реклами, політичних цінностей, екологічних уявлень тощо. Суцільні вибірки використовують під час спостереження за перебігом передвиборчих кампаній, які мають визначені часові межі, у випадку "оперативного" аналізу – наприклад, обговорення у світовій пресі міжнародних візитів державних лідерів;

– для аналізу новин, тематичної спрямованості мас-медіа найчастіше застосовують моделі вибірки, близькі до типологічних; наприклад, з цією метою в практиці контент-аналізу прийнято спостерігати за кожним п'ятим чи шостим числом газети, зважаючи на періодичність її виходу і зміщення інформації залежно від днів тижня;

– проектуються і випадкові вибірки, у яких наперед визначається крок відбору текстових одиниць або експериментально розраховується їхня кількість, достатня для репрезентативного представлення всього масиву; таку кількість номерів газети можна потім набрати за таблицею випадкових чисел.

Побудови багатоступеневої вибірки:

– спочатку визначають основні канали або джерела інформації, наприклад, досліджувані газети;

– далі визначається сукупність повідомлень вибраного джерела інформації: вибираються часові рамки, у межах яких будуть розглядатися матеріали (в основу вибору періоду можуть бути покладені якісь події – від однієї події до іншої, наприклад,

місяць, півроку, триріччя чи п'ятиріччя, "середина 90-х рр." і, нарешті, "увесь повоєнний час");

– у межах періоду можна також зробити вибірку; досвід свідчить, що два випадково відібрані дні кожного тижня місяця забезпечують гарантію цілком збалансованої вибірки за місяць; можна "конструювати" тиждень на основі місяця: понеділок вибирається з випадково взятого тижня місяця, тоді вівторок і т.д., поки не будуть відібрані всі дні тижня.

Анкета для формування вибірки:

– Які канали телебачення, радіо, видання газет і журналів будуть розглядатися?

– Яка стратегія здійснення вибірки?

– Який період часу буде охоплено?

– Які специфічні події будуть аналізуватися?

– Які частини газети вивчатимуться?

– Які передачі на телебаченні або радіо розглядаються?

Зауваження:

– необхідність правильного відбору пов'язана з дилемою: з одного боку, відбір конкретного джерела наперед обмежує рамки інтерпретації висновків аналітика, з іншого – аналітику потрібно коректно поширювати висновки на особливості і тенденції змісту всієї генеральної сукупності, тобто більших за обсягом і, можливо, якісно відмінних інформаційних потоків. Крім цього, вибірка повинна враховувати і кількісні критерії, зокрема, накладі газет і журналів, відвідуваність кінотеатрів тощо;

– вирішуючи питання побудови вибірки, варто пам'ятати, що немає таких процедур, які були б застосовані до будь-якої поставленої задачі. У кожному дослідженні це питання вирішується по-своєму, залежно від цілей, наявних у розпорядженні аналітика документів, необхідного рівня точності результатів;

– побудова вибірки завершується складанням списку досліджуваних документів, а також таблиці, яка включає відомості про кількість вибіркової документальної інформації і її розподіл за часовими інтервалами і видами документів.

6. Вибір одиниць аналізу

Рекомендації:

– одиниці аналізу – це малі або великі фрагменти тексту, за якими кодувальник методично спостерігає, відшукуючи в них посилання на категорії значень або їх вербально виражені ознаки;

– одиниця аналізу – це те, що підраховується;

– у записі змісту це може бути слово або символ, тема (окреме твердження щодо предмета дослідження), стаття повністю або окрема історія; для аналізу програм телебачення одиниця аналізу може охоплювати символи, події, цілі програми;

– операційні визначення одиниць аналізу повинні бути чітко і ретельно задані;

– одиниці аналізу повинні бути такими, щоб вони були очевидними, їх можна було легко ідентифікувати і виявляти; для виявлення одиниць аналізу іноді потрібна певна попередня робота.

Типологія одиниць аналізу (за К.Кріпендорфом²⁶¹):

– "фізичні" (сутності з чітко окресленими фізичними, геометричними або часовими межами, як, наприклад, екземпляри книги, номери газет, екземпляри плакатів або листівок, фотографії тощо);

– структурно-семіотичні (основні елементи семіотичних систем; наприклад, мовна лексика або граматичні показники, наприклад, заперечні частки або показники таких категорій, як віддієслівні імена);

– понятійно-тематичні (поняття, теми, проблеми, які стоять за словами, наприклад, "кримінал", "свобода слова");

– референційні і квазіреференційні (до них відносяться позначення реальних особистостей, подій, міст, країн, організацій тощо; наприклад, "наш Кобзар", "мати міст руських");

²⁶¹ Krippendorff K. Content Analysis: An introduction to its Methodology....

– пропозиційні та оцінкові (наприклад, за фразою "Україна задихається без інвестицій" стоїть конкретна констатація факту);

– макроструктурні (складні понятійні конструкції, наприклад, присутні в суспільній свідомості сучасної України: змова, оргія корупції, кримінальна революція, країна дурнів, боротьба за владу);

– результати концептуальних операцій (метафори, приклади й аналогії, наприклад, "воєнна" метафора: "війна з бідністю", "удар по губернатору", "атака з боку опозиції", "розгромна публікація");

– поетичні (кількісно вимірювані засоби художньої виразності, наприклад, каламбури, алітерації).

7. Система квантифікації (підрахунку кількості)

Види систем квантифікації (вироблені на основі практики досліджень):

– система рахунку "час – простір" (за основу беруть показники займаної площі для письмових текстів, наприклад, кількість рядків, абзаців, знаків, площа у квадратних сантиметрах, висота стандартної газетної колонки, або час аудіовізуальних повідомлень, наприклад, час висвітлення події);

– поява категорій або ознак у тексті (використовується у випадку, коли одиниці аналізу складно формалізувати, наприклад, у контент-аналізі художньо-образної інформації; при такій системі, поділивши текст на частини, визначається наявність деякої характеристики змісту в кожній частині, повторна поява даної характеристики в тій ж частині не враховується);

– частота появи одиниць аналізу (фіксується не лише присутність або відсутність одиниці аналізу, а й скільки разів вона зустрічається).

Зауваження:

- дослідник має підраховувати частоту появи одиниць аналізу в розрізі кожної категорії;
- кодувальники повинні бути проінструктовані щодо використання оцінкової шкали в процесі оцінювання певних властивостей символів або ситуацій;
- оцінювальна шкала дозволяє виявити глибину і структуру змісту, водночас вносить певний суб'єктивізм в аналіз і впливає на надійність результатів;
- частота появи теми в тому чи іншому документі може виступати показником її значущості;
- підрахунок оцінювальних характеристик тексту (всіх "за" і "проти" щодо деякого питання чи події) дозволяє досліджувати настанови автора і виявляти наміри, якими було продиктоване повідомлення.

8. Побудова категорій аналізу*Загальні рекомендації:*

- центральним моментом будь-якого контент-аналізу є система категорій;
- категорії аналізу – це найбільш загальні ключові поняття, які відповідають завданням дослідження, за допомогою яких описується предмет дослідження і згідно з якими будуть класифікуватися одиниці аналізу;
- усі категорії системи повинні бути взаємовиключними, вичерпними і надійними, правильно репрезентувати текст;
- відтак одиниця аналізу повинна асоціюватися лише з однією категорією, і в жодному разі не з кількома категоріями одночасно;
- категорії повинні покривати всі можливі одиниці аналізу в межах вибірки;
- система категорій аналізу може включати типологію передач або типи газетного матеріалу, типи осіб або поведінки осіб;
- система категорій має бути надійною в тому розумінні, що різні кодувальники повинні отримувати однакові результати;

– від вибору категорій значною мірою залежить характер отриманих результатів; тому при їхній розробці дослідник стоїть перед необхідністю по декілька разів переходити від теоретичної схеми до даних, а від них – знову до схеми для виявлення релевантної цілям дослідження системи категорій.

Анкета для визначення категорій аналізу:

– Чи аналіз буде займатися підрахунком площі на сторінках газет, яка відводиться на матеріали певного типу?

– Чи буде аналіз визначати і вимірювати специфічні слова або фрази в тексті?

– Чи буде у фокусі аналізу поява певних специфічних категорій поведінки на екрані?

– Чи буде аналіз підраховувати появу програм, газет, повідомлень, особливостей самих повідомлень? Який рівень підрахунку матиме місце?

Фактори, які враховуються під час визначення категорій:

– актори (політики, спортсмени, промисловці, науковці, злочинці тощо);

– джерела (політичні партії, експерти, певні організації тощо);

– контекст (де проводиться запис інтерв'ю – удома, на роботі, у студії тощо);

– ресурсне забезпечення знань (фактичні дані, оцінки, розрахунки, рекомендації тощо);

– присутність джерела (частково визначається контекстом і в результаті опрацювання інформації, поданої ЗМІ);

– суб'єкти, теми і проблема (якщо, наприклад, досліджується проблема насильства в ЗМІ, то ставимо запитання: які види насильства (фізичне, вербальне, шкода демонструються?);

– словниковий запас або лексичний вибір (частота, з якою певні слова використовуються; використання емоційних термінів, наприклад, можна одних і тих же людей назвати по-різному: борці за свободу, терористи, гангстери)

– спосіб вимірювання значень (наприклад, матеріал кодується як "позитивний", "негативний", "нейтральний").

Концептуальна схема (модель):

– категорії можуть поділятися на більш детальні категорії, тобто підкатегорії;

– сукупність категорій і підкатегорій утворює концептуальну схему контент-аналізу;

– концептуальна схема – це не теоретична абстракція, а змістовна конструкція, адаптована для зображення образу предмета, присутнього в текстовій реальності; важливо, щоб вона була придатною для повного та водночас економного опису предмета, зважаючи на теоретичні уявлення про нього і на досвід його текстових версій.

Способи розробки концептуальної схеми:

– категорії добираються експериментальним шляхом на підставі уявлень про предмет, наприклад, експерти називають найактуальніші соціально-економічні проблеми суспільства, і цей перелік служить номінальною шкалою категорій у контент-аналізі сучасної преси, що має на меті з'ясувати, який образ реальності преса пропонує читачам;

– категорії беруть з текстів, які вивчаються; це так званий пошуковий шлях, до якого звертаються тоді, коли є потреба висунути гіпотези про предмет дослідження;

– категорії обираються дослідником на прагматичній основі, тобто він спирається на свої теоретичні уявлення про предмет, історико-культурний і соціальний аналіз відображеної в текстах ситуації, практику контент-аналізу.

Надійність категорій:

– категорії повинні відповідати вимогам надійності (чим чіткіше визначені категорії, тим менше проблематичним стає віднесення частин змісту до визначених категорій);

– найбільш радикальним способом збільшення ступеня надійності категорії є її вичерпне означення (перерахування всіх елементів, які входять до неї, – частин змісту); однак такі випадки зустрічаються нечасто, і вичерпні означення можливі для порівняно вузького класу простих задач;

– при оцінці надійності категорій варто пам'ятати, що тонка грань їхнього означення призводить до розбіжностей між

кодувальниками при визначенні належності того чи іншого документа до певної категорії; цей недолік можна перебороти укрупненням (узагальненням) категорій, однак значне укрупнення може призвести до зменшення ступеня диференціації досліджуваного явища, і дослідник не помітить істотних для цілей дослідження розбіжностей;

– надмірна кількість використовуваних категорій також призводить до аналітичних прорахунків; оптимальну кількість визначити важко: прагнення максимально докладно описати предмет набором понять обертається надто роздрібненою категоріальною схемою, не пристосованою до класифікації змісту текстів у зручній і переконливій формі;

– висока надійність категорій звичайно властива простим формам контент-аналізу; при ускладненні з'являється більш тонка диференціація категорій зі зниженням їхньої надійності, однак це допомагає отримати більш широку і поглиблену інформацію про досліджуваний об'єкт; рішення про співвідношення потреб і значущості приймається, виходячи з поставлених перед дослідником завдань.

9. Забезпечення надійної роботи кодувальників

Рекомендації:

– дослідження повинно чітко визначати, які одиниці змісту будуть підраховуватися (слова, кількість кв.см тощо);

– кожен одиницю, яка має визначатися, потрібно означити; якщо категорії поділяються на підкатегорії, то дослідник мусить визначити кожен підкатегорію і спосіб її розпізнавання;

– потрібно також визначити числову схему, що вказує по кожній категорії, який числовий код призначається та в якому випадку;

– дослідник повинен розробити класифікатор аналізу та інструкцію для кодування, яка задає шаблони кодування для кожної категорії.

Класифікатор аналізу:

– класифікатор контент-аналізу являє собою методичний документ, призначений для пояснення процесів виділення та реєстрації ознак текстової інформації;

– класифікатор контент-аналізу – це загальна таблиця, у якій зведені всі категорії і підкатегорії аналізу й одиниці аналізу;

– основне призначення класифікатора – гранично чітко зафіксувати, у яких одиницях аналізу виражається кожна категорія, використовувана в дослідженні; класифікатор можна уподібнити соціологічній анкеті, де категорії аналізу виступають у ролі запитань, а одиниці аналізу – відповідей;

– класифікатор аналізу є основним методологічним документом, який визначає зміст усіх інших інструментів.

Етапи проектування класифікатора:

– формулювання завдань і проектування класифікатора;

– побудова категорій аналізу;

– термінологічне уточнення фіксованих ознак категорій аналізу в мові тексту документа, усунення двозначності і перетину двох або більше категорій;

– визначення випадків нетермінологічної фіксації ознак категорій аналізу;

– установлення системи квантифікації;

– конструювання інструментів для вимірювання і реєстрації одиниць аналізу (реєстраційних карток, лінійок тощо);

– побудова загальної логічної схеми категорій аналізу і системи квантифікації;

– побудова загальної схеми класифікації в послідовності, оптимальній для реєстрації ознак і їхнього наступного опрацювання;

– проектування бланків аналізу на основі класифікатора;

– апробація проектів бланків і їхнє уточнення;

– розробка програми наступної обробки бланків аналізу;

– затвердження класифікатора, його розмноження.

10. Кодування змісту

Загальні зауваження:

- процес віднесення одиниці аналізу до категорії аналізу називається кодуванням;
- особи, зайняті кодуванням, називаються кодувальниками;
- для кодування, як правило, використовується невелика кількість кодувальників;
- ретельне навчання кодувальників є істотним і приводить до надійніших результатів;
- кодувальники повинні бути добре ознайомлені з методом кодування і класифікатором змісту;
- потрібно декілька тривалих навчань, щоб гарантувати, що кодувальники зроблять роботу належним чином;
- потрібно експериментальне дослідження для перевірки надійності інтеркодування;
- кодування здійснюється на кодувальних бланках, які дозволяють кодувальникам класифікувати зміст; бланки заповнюються шляхом внесення в них певних позначок у потрібні місця навпроти наперед визначених показників (кодів категорій).

11. Аналіз даних та інтерпретація

Загальні зауваження:

- для аналізу даних може використовуватися описова статистика, наприклад, відсотки, середні арифметичні, мода, медіани;
- складніша статистика використовується для перевірки взаємозв'язків між змінними, серед них найпопулярнішою є статистика χ^2 -квадрат;
- до вивчення певних видів даних можуть застосовуватися кореляційний і дисперсійний аналіз;
- дослідник повинен бути націленим на всебічний підрахунок природи змісту в описовій формі, але також розуміти, що між різними видами змісту існують взаємозв'язки;
- результати оцінюються стосовно початкових запитань дослідження або гіпотез.

Представлення та інтерпретація результатів:

– значну роль у розумінні результатів контент-аналізу відіграє методична підготовка: зрозуміла і коректна методика дослідження, надійна процедура формалізованого читання;

– для отримання виважених і правильних висновків служить візуальне представлення результатів дослідження; на практиці найчастіше використовуються різноманітні графічні засоби: діаграми, графіки, схеми, графи;

– змістовна інтерпретація результатів залежить від цілей аналізу; вона є творчим процесом, результати якого багато в чому визначаються кваліфікацією й інтуїцією аналітиків.

12. Надійність*Загальні рекомендації:*

– поняття надійності є критичним для контент-аналізу;
– для досягнення прийняттого рівня надійності рекомендуються такі кроки:

1) визначити категорії з максимальною детальністю; невизначені або двозначно визначені категорії є найімовірнішим джерелом низької надійності;

2) ґрунтовно навчати кодувальників; перед початком збору даних переконайтеся, що кодувальники знають правила кодування і вміють їх застосовувати; відведіть час на обговорення та вирішення всеможливих проблем;

3) проводьте експериментальне дослідження; виберіть із генеральної сукупності певні тексти і дайте кодувальникам закодувати їх.

Показник надійності:

– обчислення відсотка погодженості кодувальників щодо оцінок або кодування категорій являє собою найпростіший метод перевірки надійності;

– ця міра надійності визначається як таке відношення:

$$\pi = \frac{\% \text{ фактичної згоди} - \% \text{ очікуваної згоди}}{1 - \% \text{ очікуваної згоди}}$$

– прийнятний рівень необхідної домовленості залежить від виду здійснюваного дослідження; зазвичай стандартним є мінімальний рівень 80 відсотків;

– приклад: нормативні (очікувані) результати оцінок журнальної реклами за сімома категоріями повинні бути такими:

- іміджева реклама – 30 відсотків;
- реклама продажу – 20 відсотків;
- реклама залучення до акції – 20 відсотків;
- реклама-застереження – 15 відсотків;
- утилітарне звернення – 10 відсотків;
- інше звернення – 5 відсотків.

Спочатку розраховується відсоток очікуваної згоди двох кодувальників: це сума квадратів відсотків по кожній категорії:

$$\% \text{ очікуваної згоди} = (0.3)^2 + (0.2)^2 + (0.2)^2 + (0.15)^2 + (0.1)^2 + (0.05)^2 = .20$$

Якщо кодувальники домовляються про 90 відсотків з того, що очікувалося, то π може розраховуватися таким чином:

$$\pi = \frac{0,90 - 0,20}{1 - 0,20} = 0,875 .$$

Остаточна перевірка надійності кодування:

– після задовільного завершення навчання кодувальників і проведення експериментального дослідження, наступна перевірка надійності здійснюється під час роботи з основною вибіркою;

– зазвичай рекомендується 10-25 відсотків обсягу вибірки опрацьовувати з повторним дублюючим кодуванням;

– для перевірки надійності інтеркодування використовують декілька індексів;

– індекс Олсті обчислюється так:

$$\text{Надійність} = \frac{2 * M}{N_1 + N_2} ,$$

де M є число рішень кодувальників, по яких вони дійшли згоди, а N_1 і N_2 позначають кількість кодувальних рішень (наприклад, кількість закодованих статей, якщо одиниця аналізу – стаття), які прийняв відповідно перший і другий кодувальник.

Тоді, якщо обидва кодувальники кодували по 50 одиниць аналізу і дійшли згоди щодо 35, то маємо

$$\text{Надійність} = \frac{2 * 35}{50 + 50} = 0,70.$$

13. Помилки під час проведення контент-аналізу

Під час використання контент-аналізу дослідники часто припускаються помилок. Ось перелік найпоширеніших:

- аналіз документів випереджає розробку програми досліджень;

- аналізуються документи, не пов'язані з гіпотезами дослідження, наприклад, відповідають темі дослідження хіба що за назвою;

- не перевірена справжність документів;

- не враховане їхнє авторство;

- неповно враховане їхнє призначення;

- категорії аналізу не визначені з тією точністю, яка дозволяє чітко розрізняти значенні одиниці тексту документа;

- категорії аналізу не субординовані і не приведені у відповідність з тими дефініціями й термінами, які їх операціоналізують;

- категорії аналізу не відповідають змісту і мові тексту аналізованого документа;

- одиниці аналізу характеризують категорії лише ззовні, а не за глибиною змісту;

- одиниці аналізу не дозволяють ідентифікувати зміст документа за категоріями аналізу;

- аналізуються документи при відсутності повного комплексу методичних інструментів;

- класифікатор складений з порушенням законів логіки;

- кодувальники не отримали належної методичної підготовки;

- інструкція для реєстрації та кодування недостатньо повна, складена дослідником, який не провів попередньої апробації інструментарію;

- результати контент-аналізу повторно не перевірені через інформацію, отриману іншими методами.

ДОДАТОК Б. Програми контент-аналізу

Перелік і опис програм контент-аналізу складено на основі ряду робіт К.Нойєндорф²⁶², М.Алекси²⁶³ та дослідницької групи з Великобританії ESRC²⁶⁴.

1. Програми кількісного контент-аналізу

САТРАС

<http://www.terraresearch.com/>

САТРАС читає текстові файли і формує різного роду звіти, починаючи від звичайної таблиці частот, упорядкованої за частотою або в алфавітному порядку, і завершуючи коротким підсумковим звітом "основних ідей" тексту. Програма подає зразки використання слів і здійснює кластерний або дисперсійний аналіз. Додатком до неї є програма, яка відображає дані результатів дослідження у вигляді дво- і тривимірної графіки.

Computer Programs for Text Analysis

<http://unix.dsu.edu/~johnsone/ericpgms.html>

Являє собою сукупність програм Е.Джонсона, кожна з яких призначена для реалізації однієї або кількох функцій, наприклад, аналіз зовнішніх характеристик героїв п'єси (програма ACTORS), отримання KWIC (програма CONCORD), знаходження цитат у текстах (програма DIALOG), порівняння словника двох текстів (програма IDENT). Загалом програми призначаються для літературного аналізу текстів.

²⁶² Neuendorf K.A. The Content Analysis Guidebook Online. Computer Content Analysis Programs. –

<http://academic.csuohio.edu/kneuendorf/content/cpuca/ccap.htm>

²⁶³ Alexa M. Computer-assisted text analysis methodology in the social sciences...– 40 p.; Alexa M., Zuell C. Commonalities, differences and limitations of text analysis Software...– 29 p.

²⁶⁴ Assessment and Development of New Methods for the Analysis of Media Content // An ESRC Research Methods Programme: Research Bulletin 01.– May 2003.– www.lboro.ac.uk/research/mmethods

Concordance 2.0

<http://www.rjcw.freeseerve.co.uk/>

Дозволяє формувати повні конкорданси текстів будь-якого розміру, обмежуючись лише вільним місцем на диску. Можна також оперативно отримати конкорданси для окремих слів, а також формувати web-конкорданси.

Diction 5.0

<http://www.sagepub.com/>

Diction 5.0 має серію вбудованих словників, які класифікують і знаходять документи за п'ятьма семантичними ознаками (діяльність, оптимізм, упевненість, реалізм і уніфікація) і 35 більш детальними ознаками (зокрема завзятість, відповідальність, двояке ставлення, спонування, комунікація). Після того, як текст користувача проаналізований, Diction 5.0 порівнює результати для кожної з 40 категорій словника з "нормальним оцінювальним рядом", сформованим на основі аналізу програмою понад 20 тис. текстів. Користувачі можуть порівняти свій текст або з нормативним профілем 20 тис. текстів, або з 6 окремими видами текстів (бізнес, щоденні події, розваги, журналістика, література, політика, навчання), які можуть, у свою чергу, поділятися на 36 підвидів (наприклад, фінансові звіти, комп'ютерні чати, лірична музика, газетні передовиці, новели або короткі оповідання, політичні дебати, вивчення громадської думки). Крім того, Diction 5.0 вираховує абсолютні та відносні частоти слів або стандартизовані оцінки і виводить їх за частотою або в алфавітному порядку. У результаті додаткового аналізу програма може створювати словники користувача.

DIMAP (DIctionary MAintenance Programs)

<http://www.clres.com/>

DIMAP має за ціль розробку словника. Програма включає різноманітні інструменти для комп'ютерної лінгвістики та процесорів природної мови.. З допомогою DIMAP можна будувати, керувати, редагувати, обслуговувати словники, шукати слова і порівнювати словники користувача з

традиційною лексикою. Програма також уключає модуль текстового аналізу МССА, описаний нижче.

General Inquirer (Internet version)

<http://www.wjh.harvard.edu/~inquirer/>

Ця поважна та все ще використовувана програма віднайшла нове життя у World Wide Web. Діалогова версія General Inquirer дозволяє просто і швидко зробити нескладний комп'ютерний аналіз тексту: завітавши на сайт General Inquirer і ввівши текст у відповідний блок, можна одразу отримати результати аналізу. General Inquirer кодує і класифікує текст, використовуючи словник Harvard IV-4 для оцінки такої риси, як мужність, здійснює тривимірну семантичну оцінку за методикою Ч.Осгуда, визначає рівень емоційності мови, пізнавальну орієнтацію тексту тощо. На завершення кожного аналізу програма також повертає кумулятивну статистику, зокрема прості частоти для слів.

HAMLET

<http://www.apb.cwc.net/homepage.htm>

Головна ідея HAMLET – знайти в текстовому файлі слова із заданого списку, а також підрахувати частоти в межах будь-якої вказаної одиниці контексту або поєднання слів у межах заданого діапазону слів. Індивідуальні частоти слів, частоти спільної появи слів, виражені в термінах вибраної одиниці контексту, відображаються в матриці схожості, яка може бути надалі піддана традиційному кластерному аналізу.

INTEXT/TextQuest – Text Analysis Software

<http://www.intext.de>

INTEXT являє собою програму, призначену для аналізу текстів гуманітарного і загального спрямування. Вона формує індекси, конкорданси, таблиці KWIC і KWOC, проводить аналіз читабельності, індивідуальний структурний аналіз, формує списки слів, послідовності слів, можливих перестановок слів, аналізує стилістику тощо. Програма TextQuest є Windows версією INTEXT.

LIWC (Linguistic Inquiry and Word Count software)

<https://www.erlbaum.com/shop/tek9.asp?pg=products&specific=1-56321-208-0>

LIWC має серію 68 убудованих словників, які допомагають здійснювати пошук текстових файлів і підраховувати, як часто слова відповідають кожному із 68 наперед установлених параметрів. Словники включають лінгвістичні оцінки, словесні категорії психологічних конструкцій та індивідуальні проблемні категорії. Програма також дозволяє користувачам створювати власні словники. Вона є особливо корисною для психологів, які досліджують записи розмов з пацієнтами.

МССА

<http://www.clres.com/>

МССА аналізує текст, створюючи частотні й алфавітні списки, таблиці KWIC, а також здійснює кодування на основі убудованих словників. Убудовані словники спеціалізуються на відшуканні певних особливостей тексту, наприклад, вид діяльності, сленг, ознаки гумору тощо. Результати виводяться у формі вікон. Можна легко впорядковувати вікна, переходити від одного до іншого. МССА також дозволяє аналізувати записи інтерв'ю, різного роду слухань, виступів на телебаченні, роботи фокус-груп, ділових ігор, пов'язаних з участю великої кількості людей.

МЕСА (Map Extraction Comparison and Analysis)

МЕСА включає 15 шаблонів для аналізу тексту. Багато з них призначені для побудови когнітивних карт і акцентують увагу як на самих когнітивних поняттях, так і зв'язках між ними. Використовуються шаблони для здійснення класичного контент-аналізу, наприклад, програма здійснює підрахунок кількості понять по кожній карті.

MonoConc

<http://www.ruf.rice.edu/~barlow/mono.html>

MonoConc призначена для формування конкордансу. Результати можуть різними способами впорядковуватися і представлятися в конфігураціях, заданих користувачем. Програма також обчислює частоти слів для заданої сукупності текстів.

ParaConc

<http://www.ruf.rice.edu/~barlow/parac.html>

ParaConc – двомовна/багатомовна програма конкордансу, яка застосовується під час порівняльних мовних досліджень текстів.

PCAD 2000

<http://www.gb-software.com/>

PCAD 2000 реалізує контент-аналіз записів промов та інших текстів, застосовуючи підхід Готтшалка-Глесера для оцінювання розумового й емоційного стану людей. Так, до оцінюваних параметрів належить занепокоєння, ворожість, соціальне відчуження, надія, депресія тощо. Програма дозволяє порівнювати отримані оцінки з нормативними показниками для різних демографічних груп суб'єктів. Вона може робити деякі діагностичні класифікації, керуючись "Діагностичною і статистичною інструкцією психічних розладів" (DSM-IV), підготовленою Американською психіатричною асоціацією.

PROTAN (for PROTOCOL ANalyzer)

PROTAN – автоматизована система контент-аналізу. Перше завдання, на яке вона намагається відповісти: як виглядає текст. Для досягнення цього PROTAN використовує серію семантичних словників, які є складовими елементами системи. Інше завдання: з'ясувати, про що говорить текст.

SALT (Systematic Analysis of Language Transcripts)

<http://www.waisman.wisc.edu/salt/index.htm>

Програма розроблена головним чином, щоб допомогти лікарям виділити і документально підтвердити специфічні мовні проблеми пацієнтів. Вона виконує ряд видів аналізу, зокрема,

пов'язані з вимовою (наприклад, неповне висловлювання, нерозбірливість, невербальність), обчисленням довжини висловлювань, кількості і тривалості пауз, швидкості мови, частотний аналіз різних наборів слів (наприклад, негативи, сполучники, словники користувача).

SWIFT

(Structured Word Identification and Frequency Table)

Swift у процесі діалогу здійснює орієнтований на ключові слова аналіз коротких текстів. Поширюється безкоштовно, операційна система – MS DOS.

TABARI

(Text Analysis By Augmented Replacement Instructions)

TABARI є спадкоємцем програми KEDS, застосовується для аналізу коротких повідомлень, новин, різноманітних службових звітів. Вона кодує дані з міжнародними подіями (особливістю є те, що вони відображають взаємодію між учасниками), використовуючи шаблони розпізнавання і звичайний граматичний аналіз. Автори розробили цілий ряд словників для кодування подій. Наприклад, кодувальна схема KWEIS може визначити, хто діє всупереч кому, приміром, Ірак проти Кувейту. Коли така тема з'являється в газетному матеріалі, програма може автоматично кодувати агресора, жертву і дію, а також дату події.

TACT (Text Analysis Computing Tools)

<http://www.chass.utoronto.ca:8080/cch/tact.html>

TACT є системою аналіз і пошуку тексту для MS-DOS, що дозволяє робити запити європейськими мовами під час опрацювання текстових баз даних. Вона розроблена командою програмістів, проектувальників і науковців – фахівців з текстового аналізу. Крім того, система включає додаткову програму TACTweb, яка є засобом для з'єднання TACT з World Wide Web. Використовуючи формат WWW, користувачі отримують доступ до багатьох діалогових послуг TACT.

ТЕХТРАСК 7.0

<http://www.social-science-geis.de/en/software/textpack/index.htm>

Програма ТЕХТРАСК, яка спочатку проектувалася для аналізу анкет із запитаннями відкритого типу, була розширена в останні роки за рахунок уключення аналізу змісту, літературної та лінгвістичної аналітики. Зараз вона підраховує частоти слів, алфавітні списки, таблиці KWIC і KWOC, перехресні посилання, здійснює порівняння слів у двох текстах, а також кодування згідно з розробленими користувачем словниками. Результати можуть бути експортовані у формати програм статистичного аналізу. Windows-версія програми вдало використовує переваги інтерфейсу операційної системи.

TextSmart by SPSS Inc.

<http://www.spss.com/spssbi/textsmart/>

Дане програмне забезпечення, яке проектувалося перш за все для аналізу анкет із запитаннями відкритого типу, використовує кластерний аналіз і техніку багатовимірного шкалювання для автоматичного аналізу ключових слів і групує тексти в категорії. Програма може "кодувати" тексти, не використовуючи створені користувачами словники. TextSmart має приємний і простий у використанні Windows-інтерфейс, який дозволяє швидко сортувати списки слів за алфавітом і за частотою, виводити інформацію у вигляді повнокольорових діаграм, графіків, карт, схем, таблиць.

VBPro

<http://excellent.com.utk.edu/~mmmiller/vbpro.htm>

Програма формує частоти і алфавітні списки слів, KWIC-таблицю, закодовані рядки слів, використовуючи створені користувачем словники. Пакет програм уключає модуль VBMap, призначений для малювання багатовимірних карт, які вимірюють і демонструють ступінь спільної появи слів у тексті або спільної появи фрагментів тексту. Програма працює під MS DOS.

WordStat v3.01

<http://www.simstat.com/wordstat.htm>

Ця програма, будучи додатком до програми статистичного аналізу Simstat, включає декілька дослідницьких інструментів, як, наприклад, кластерний аналіз і багатовимірні статистичні моделі, для аналізу анкет із запитаннями відкритого типу й іншого виду текстів. Вона також дозволяє здійснювати кодування на основі словників користувача, підраховує частоту слів, формує алфавітні списки слів, KWIC-таблиці, проводить порівняння між підгрупами даних. Відмінності між підгрупами або числовими змінними (наприклад, вік, дата публікації) можуть бути показані візуально на діаграмах, у вигляді дво- або тривимірної графіки. Особливої уваги заслуговує спеціальний інструмент, який дозволяє користувачеві будувати загальну схему категорій, користуючись лексичною базою даних WordNet та іншими словниками (один англійський і п'ять словників інших мов).

2. Програми якісного контент-аналізу

ATLAS/ti

Це програмне забезпечення служить для інтерпретації змісту тексту, менеджменту тексту, витягування концептуальних знань з тексту, тобто побудови теорії. Сфери використання: суспільні науки, економіка, дослідження ринку, освіта, криміналістика, управління якістю, побудова баз знань, теологія. Програмне забезпечення придатне для роботи з MS DOS та MS Windows.

Code-A-Text

Code-A-Text – пакет програм, написаний для допомоги в навчанні психотерапевтів. Спочатку передбачалося, що він має полегшити аналіз терапевтичних бесід, під час яких клініцисти, викладачі і науковці намагаються зрозуміти ідеї і структури, закладені в текстах. Тепер Code-A-Text застосовується для аналізу інших видів текстів, зокрема мемо (приміток), відповідей на запитання анкет, мета-планів. Незабаром має бути

запущена версія програми, яка підтримуватиме кодування звукових файлів.

3. Програми, які допомагають проводити якісний аналіз даних

The Ethnograph v4.0

Програма забезпечує якісні дослідження і аналіз даних, полегшує менеджмент і аналіз текстових баз даних, зокрема записи інтерв'ю, роботи фокус-груп, хвилинних зустрічей, мемо, щоденники, інші документи.

Kwalitan 4.0

Kwalitan – це програма підтримки аналізу якісних даних, зокрема протоколів інтерв'ю, спостережень, існуючих письмових матеріалів, газет, річних звітів організацій, древніх рукописів тощо. Фактично, Kwalitan – це програма управління базами даних спеціального призначення. Вона була розроблена для реалізації процедур у рамках "обґрунтованої теорії".

NUD*IST

Програма допомагає дослідникам маніпулювати нечисловими неструктурованими даними шляхом індексування, пошуку і побудови "теорії". Серед інших речей, автоматизує копітку роботи з "автокодування" тексту.

winMAX

Допомагає проводити якісний аналіз даних.

ДОДАТОК В. Програми Text Mining

Перелік і опис програм Text Mining сформовано на основі робіт Д.Ланде²⁶⁵.

Intelligent Miner for Text

<http://www3.ibm.com/software/data/iminer/fortext/>

Продукт фірми IBM Intelligent Miner for Text являє собою набір окремих утиліт, які запускаються з командного рядка незалежно одна від одної. Ця система є одним з кращих інструментів "глибинного" аналізу текстів. Система складається з таких утиліт управління знаннями:

- Language Identification Tool (автоматичне розпізнавання мови, якою складено документ);

- Categorisation Tool (утиліта класифікації, тобто автоматичного віднесення тексту до деякої категорії, вхідною інформацією на фазі навчання може служити результат роботи утиліти Clusterisation Tool);

- Clusterisation Tool (утиліта кластеризації, тобто розбиття великої кількості документів на групи за близькістю стилю, форми, різноманітних частотних характеристик виявлених ключових слів);

- Feature Extraction Tool (утиліта визначення нового – виявлення в документі нових ключових слів, наприклад, власних імен, назв, скорочень, на основі аналізу наперед заданого словника);

- Annotation Tool (утиліта "виявлення змісту" текстів і складання на їхній основі рефератів або анотацій).

IBM Intelligent Miner for Text об'єднує потужний набір інструментів, які базуються в основному на механізмах пошуку інформації (information retrieval), що є особливістю всього продукту. Система включає низку базових компонент, що мають самостійне значення поза межами технології Text Mining – це

²⁶⁵ Ланде Д. Добыча знаний...; Ланде Д. Контент-мониторинг New Media.– <http://uaopt.net/files/?dwl-yal01>; Ланде Д. Об отделении зерен от плевел...; Ланде Д., Литвин А. Феномены современных информационных потоков // Сети и бизнес.– №1.– 2001.

інформаційно-пошукова система Text Search Engine, утиліта сканування Web-простору Web crawler, Net Question Solution – рішення для пошуку в локальному Web-сайті або на декількох intranet/Internet-серверах, Java Sample GUI – набір інтерфейсів Java Beans для адміністрування і організації пошуку на основі Text Search Engine.

Intelligent Miner for Text як продукт IBM включено в комплекс "Information Integrator for Content" для СУБД DB2 як засіб Information Mining ("глибинного аналізу інформації").

TextAnalyst

<http://www.megaputer.com/>

Російська компанія "Мегап'ютер Інтеллідженс", відома своєю системою PolyAnalyst класу Data Mining, розробила також систему TextAnalyst, яка вирішує такі задачі Text Mining:

- створення семантичної мережі великого тексту;
- підготовка резюме тексту;
- пошук по тексту;
- автоматична класифікація і кластеризація текстів.

Побудова семантичної мережі – це пошук ключових понять тексту і встановлення взаємозв'язків між ними. Створена мережа дозволяє не тільки зрозуміти, про що йде мова в тексті, але і здійснювати контекстну навігацію. Підготовка резюме – це виділення в тексті речень, у яких частіше за інших зустрічаються значущі для цього тексту слова. У 80% випадків цього цілком достатньо для отримання загального уявлення про текст. Для пошуку інформації в системі передбачено використання запитів природною мовою. За запитом будується унікальна семантична мережа, яка при взаємодії з мережею документа дозволяє виділити потрібні фрагменти тексту. Кластеризація і класифікація проводяться стандартними методами видобування даних.

Система Text-Analyst розглядає Text Mining як окремих математичний апарат, який розробники програмного забезпечення можуть убудовувати в свої продукти, не спираючись на платформи інформаційно-пошукових систем або СУБД. Основна платформа для застосування системи – MS Windows 9x/2000/NT.

WebAnalyst

<http://www.megaputer.com/products/wa/index.php3>

Система WebAnalyst – також продукт "Мегап'ютер Інтеллідженс", який являє собою інтелектуальне клієнт/серверне масштабоване вирішення для компаній, що бажають максимізувати ефект аналізу даних у Web-середовищі. Сервер WebAnalyst функціонує як експертна система збору інформації і управління контентом Web-сайту. Модулі WebAnalyst вирішують три завдання:

- збір максимальної кількості інформації про відвідувачів сайту і запрошені ними ресурси;
- дослідження зібраних даних;
- генерування персоналізованого контенту на основі результатів досліджень.

Вирішення цих завдань у сукупності повинно, на думку розробників системи, максимізувати кількість нових відвідувачів Web-сайту і утримати існуючі, а отже, збільшити популярність ресурсу. WebAnalyst також спроможний інтегрувати можливості Text Mining безпосередньо у web-сайт організації. Це дозволяє запропонувати індивідуалізований, автоматизований і цільовий маркетинг, автоматичний пошук і реалізацію перехресного продажу і розширити набір настроюваних користувачем даних. По суті WebAnalyst є інтелектуальним сервером додатків електронної комерції.

Технічна платформа – аналогічна TextAnalyst.

Text Miner

<http://www.sas.com/technologies/analytics/datamining/textminer/>

Американська компанія SAS Institute випустила систему SAS Text Miner для порівняння певних граматичних і словесних рядів у записаному тексті. Text Miner – винятково універсальна, оскільки може працювати з текстовими документами різних форматів – у базах даних, файлових системах і навіть у Web.

Text Miner забезпечує логічну обробку тексту в середовищі потужного пакета SAS Enterprise Miner. Це дозволяє користувачам збагачувати процес аналізу даних, інтегруючи неструктуровану текстову інформацію з існуючими

структурованими даними, такими, як вік, доход, особливості купівельного попиту тощо.

Програма Text Miner дозволяє визначати, наскільки є правдивим той або інший текстовий документ. Виявлення неправдивості документів здійснюється шляхом аналізу тексту і виявлення зміни стилю листа, що може мати місце при спробі спотворити або приховати інформацію. Для пошуку змін використовується принцип, який полягає в пошуку аномалій і трендів серед записів баз даних без з'ясування їх змісту. При цьому в Text Miner включено величезний набір документів різного ступеня правдивості, чия структура приймається як шаблони. Кожен документ, який "проганяється" на детекторі брехні, аналізується і порівнюється з цими шаблонами, після чого програма присвоює документу той або інший індекс правдивості. Особливо корисною програма може стати в організаціях, які отримують багато електронної кореспонденції, а також у правоохоронних органах для аналізу показників правдивості свідчень разом з детекторами брехні, чия дія базується, на відміну від Text Miner, на спостереженнях за емоційним станом людини.

SemioMap

<http://www.entrieva.com/entrieva/index.htm>

SemioMap – це продукт компанії Entrieva, створений у 1996 р. вченим-семіотиком Клодом Фогелем. У травні 1998 р. продукт був випущений як промисловий комплекс SemioMap 2.0 – перша система Text Mining, яка працювала в архітектурі клієнт-сервер. Система SemioMap складається з двох компонент – сервера SemioMap і клієнта SemioMap. Робота системи здійснюється в три фази:

- індексування (сервер SemioMap автоматично читає масиви неструктурованого тексту, витягує ключові фрази або поняття і створює на їхній основі індекс);

- кластеризація понять (сервер SemioMap виявляє зв'язки між витягнутими фразами і будує з них, на основі їх спільної появи, лексичну мережу – "понятійну карту");

- графічне відображення і навігація (візуалізація карт зв'язків, яка забезпечує швидку навігацію по ключових фразах і

зв'язках між ними, а також можливість швидкого доступу до конкретних документів).

SemioMap підтримує розкладання матеріалу по різних "папках" і створення окремої бази даних для кожної папки. Зв'язки між поняттями, які виявляє SemioMap, базуються на спільній появі фраз в абзацах початкового текстового масиву.

Центральним блоком SemioMap є лексичний екстрактор – програма, яка витягує фрази з текстової сукупності і виявляє спільну появу цих фраз (їх взаємозв'язки). Лексичний екстрактор базується на патентованій технології SemioLex. Вона реалізує теоретичні семантичні ідеї формування мовних комунікацій Клода Фогеля.

InterMedia Text, Oracle Text

<http://technet.oracle.com/products/text/content.html>

Засоби Text Mining, починаючи з Text Server у складі СУБД Oracle 7.3.3 і картриджа interMedia Text в Oracle8i, є невід'ємною складовою продуктів Oracle. У Oracle9i ці засоби розвинулися і отримали нову назву "Oracle Text" – програмний комплекс, інтегрований в СУБД, що дозволяє ефективно працювати з запитамі, які відносяться до неструктурованих текстів. При цьому обробка тексту поєднується з можливостями, наданих користувачу для роботи з реляційними базами даних. Зокрема, під час написання додатків для обробки тексту з'явилася можливість використовувати SQL.

Основним завданням, на яке націлені засоби Oracle Text, є пошук документів за їхнім змістом – словами або фразами, які при необхідності комбінуються з використанням бульових операцій. Результати пошуку ранжуються за релевантністю, з урахуванням частоти появи слів запиту в знайдених документах. Для підвищення повноти пошуку Oracle Text надає ряд засобів розширення пошукового запиту, серед яких можна виділити три групи: 1) розширення слів запиту всіма морфологічними формами завдяки залученню знань про морфологію мови; 2) Oracle Text допускає розширення слів запиту близькими за змістом словами за рахунок підключення тезауруса – семантичного словника; 3) розширення запиту словами, близькими за написанням і за звучанням – нечіткий пошук і

пошук співзвучних слів. Нечіткий пошук доцільно застосовувати для пошуку помилково написаних слів, а також у тих випадках, коли виникають сумніви щодо правильного написання, наприклад, прізвища, назви організації тощо.

Система Oracle Text дозволяє здійснювати тематичний аналіз текстів англійською мовою. У ході обробки текст кожного документа піддається лінгвістичному і статистичному аналізу, внаслідок чого визначаються його ключові теми, будуються тематичні резюме, а також загальне резюме – реферат.

Усі описані засоби можуть використовуватися спільно, що підтримується мовою запитів у поєднанні з традиційним синтаксисом SQL і PL/SQL для пошуку документів. Oracle Text надає можливість працювати з сучасними реляційними СУБД у контексті складного багатоцільового пошуку і аналізу текстових даних.

Можливості обробки текстової інформації російською мовою в Oracle Text обмежені. Для вирішення цієї проблеми компанією "Гарант-Парк-інтернет" розроблено модуль Russian Context Optimizer (RCO), призначений для спільного використання з interMedia Text (або Oracle Text). Крім підтримки російськомовної морфології, RCO включає засоби нечіткого пошуку, тематичного аналізу і реферування документів.

Autonomy Knowledge Server

<http://www.autonomy.com/tech/whitepaper.pdf>

Архітектура системи компанії Autonomy, відомої своїми розробками зі статистичного контент-аналізу, об'єднує інтелектуальний парсинг за шаблонами зі складними методами контекстного аналізу і для вирішення завдань автоматичної класифікації й організації перехресних посилань. Основна перевага системи Autonomy – потужні інтелектуальні алгоритми, засновані на статистичній обробці. Ці алгоритми базуються на інформаційній теорії Клода Шеннона, імовірностях Байеса та нейронних мережах.

Концепція адаптивного ймовірнісного моделювання APCM дозволяє системі Autonomy ідентифікувати шаблони в тексті

документа і автоматично визначати подібні шаблони для інших документів.

Важливий момент у системі *Autonomy Knowledge Server* – це можливість аналізу текстів та ідентифікації ключових концепцій у межах документів шляхом аналізу кореляції частот і зв'язків між термінами тексту. Компонента системи *Agentware* використовує унікальну технологію аналізу шаблонів (нелінійна адаптивна цифрова обробка сигналу) з документів і визначення характеристик, що містяться в текстах. *APCM* дозволяє ідентифікувати унікальні "сигнатури" тексту, а також створювати агентів концепцій, за допомогою яких шукаються подібні за змістом записи на *Web*-сайтах, у новинах, архівах електронної пошти і в інших документах. Оскільки система не базується на зумовлених ключових словах, вона може працювати з будь-якими мовами.

Ядро системи агентів *Autonomy* – це механізм динамічних суджень *DRE*, заснований на технології обробки шаблонів, яка використовує методи нейронних мереж. Він використовує концепцію адаптивного ймовірнісного моделювання для реалізації чотирьох головних функцій: виявлення концепції, створення агента, навчання агента і стандартний пошук тексту. *DRE* сприймає запити природною мовою або терміни, пов'язані бульовими операторами, і повертає список документів, упорядкованих за релевантністю запиту. Цей механізм є основою всіх продуктів системи агентів від *Autonomy*.

Galaktika-ZOOM

<http://zoom.galaktika.ru/content.htm>

Система *Galaktika-ZOOM* – продукт російської корпорації "Галактика". Основне призначення системи – інтелектуальний пошук за ключовими словами з урахуванням морфології російської і англійської мов, а також і формування інформаційних масивів для конкретних аспектів запиту. Обсяги знайденої інформації можуть досягати сотень гігабайт. Система орієнтована на великі інформаційні об'єкти – повідомлення і статті ЗМІ, галузеві видання, нормативну документацію, ділове листування, матеріали внутрішнього документообігу підприємства, інформацію з Інтернет. Вона надає

інструментарій для аналізу об'єктивних змістовних зв'язків відібраних даних і формування "образу" проблеми – багатовимірної моделі в інформаційному потоці у формі ранжованого списку значущих слів з даної проблематики. Значна увага приділена виявленню тенденцій динаміки розвитку досліджуваної проблеми. Система містить конвертори часто використовуваних форматів: простий текст, RTF, DOC, HTML. Galaktika-ZOOM функціонує в середовищі ОС Windows 2000.

InfoStream

<http://infostream.com.ua>

Охоплення й узагальнення великих динамічних інформаційних масивів, що безперервно генеруються в Інтернет, потребує якісно нових підходів, одним з яких виступає контент-моніторинг. Для отримання якісних і кількісних зрізів такий моніторинг повинен проводитися постійно впродовж не визначеного наперед часу. Для вирішення цього завдання в Україні в Інформаційному центрі "ЕЛВІСТІ" розроблена технологія InfoStream™. Програмно-технологічні засоби InfoStream включають три основні складові (центри):

- збору й обробки інформації;
- організації інтерактивного доступу до баз даних;
- контент-моніторингу.

Ядром механізму обробки контенту InfoStream є повнотекстова інформаційно-пошукова система InfoReS. Технологія дозволяє створювати повнотекстові бази даних і здійснювати пошук інформації, формувати тематичні інформаційні канали, автоматично проводити рубрикацію інформації, формувати дайджести, таблиці взаємозв'язків понять на основі Інтернет-інформації, гістограми розподілу зважених значень окремих понять, а також динаміки їхньої появи в мережі. Технологія InfoStream дозволяє обробляти дані у форматах MS WORD (DOC, RTF), PDF, а також усіх текстових форматах (звичайний текст, HTML, XML). Системи на основі InfoStream функціонують на платформах ОС FreeBDS, Linux, Solaris.

Наукове видання

Юськів Богдан Миколайович

**Контент-аналіз. Історія розвитку
і світовий досвід**

Комп'ютерна верстка Б.М.Юськів
Технічний редактор Б.М.Юськів
Коректор О.І.Якимчук

Підп. до друку 26.06.2006 р. Формат 60x84 Обл.-вид.арк.-7,7.
Наклад 300 прим. Зам. № 466.

Видавничі роботи: "Перспектива"^{ТМ}
Свідоцтво про внесення суб'єкта видавничої справи до
Державного реєстру видавців, виготівників і розповсюджувачів
видавничої продукції: серія РВ №26 від 6 квітня 2004 р.
Пр. Кн.Романа 9/24, м. Рівне, 33022; тел.: (0362) 24-45-09

Віддруковано засобами ризографічного друку,
ПП Самборський І.О.
Вул. Толстого, 3, м. Рівне, 33028.

