

УДК 004.774.6CMS:004.912

КЛАСИЧНИЙ КОНТЕНТ-АНАЛІЗ ТА АНАЛІЗ ТЕКСТУ: ТЕРМІНОЛОГІЧНІ ТА МЕТОДОЛОГІЧНІ ВІДМІННОСТІ

Іванов Олег Валерійович – кандидат соціологічних наук, доктор філософії (PhD) в галузі масових комунікацій, начальник науково-дослідної лабораторії, заступник начальника науково-дослідного відділу прикладних соціологічних досліджень Науково-дослідного центру гуманітарних проблем Збройних Сил України

У статті розглядаються концептуально-методологічні засади методу контент-аналізу у співвіднесенні з методиками і техніками кількісної обробки текстових даних. Автор пропонує визначення понять "методологія", "метод", "методики", "техніки" в контексті дослідження текстів у соціальних науках. Досліджується базове термінологічне поле співвіднесення "класичного" контент-аналізу, "текст-майнінгу", "комп'ютерного контент-аналізу" в межах якісної та кількісної парадигми, "аналізу текстових даних", "аналізу природної мови", "кількісного контент-аналізу", "якісного контент-аналізу", теоретико-методологічні передумови їхнього застосування. Представлений світовий досвід класифікації методів кількісного аналізу текстів у соціальних науках, наявні суперечності між різними школами контент-аналізу. Обґрунтовується необхідність оновлення вимог до контент-аналітичного дослідження в умовах переходу до інформаційного суспільства та застосування комп'ютерних технологій в дослідженнях комунікації.

Ключові слова: контент-аналіз, текст-майнінг, комп'ютерний контент-аналіз, аналіз текстових даних, кількісний контент-аналіз, якісний контент-аналіз.

В статье рассматриваются концептуально-методологические основания метода контент-анализа в соотнесении с методиками и техниками количественной обработки текстовых данных. Автор предлагает определение понятий "методология", "метод", "методики", "техники" в контексте исследования текстов в социальных науках. Исследуется базовое терминологическое поле соотнесения "классического" контент-анализа, "текст-майнинга", "компьютерного контент-анализа" в рамках качественной и количественной парадигм, "анализа текстовых данных", "анализа естественного языка", "количественного контент-анализа", "качественного контент-анализа", теоретико-методологические предпосылки их применения. Представлен мировой опыт классификации методов количественного анализа текстов в социальных науках, имеющиеся противоречия между различными школами контент-анализа. Обсуждается необходимость обновления требований к контент-аналитическому исследованию в условиях перехода к информационному обществу и применения компьютерных технологий в исследованиях коммуникации.

Ключевые слова: контент-анализ, текст-майнинг, компьютерный контент-анализ, анализ текстовых данных, количественный контент-анализ, качественный контент-анализ.

The paper is devoted to conceptual and methodological review of the method of content analysis in relation to the procedures and techniques of quantitative analysis of text data. The author provides definitions of the terms "methodology", "method", "procedure", "technique" in the context of study of texts in social sciences. The basic terminological field of correlation of "classic" content analysis, "text-mining", "computer content analysis" within the framework of qualitative and quantitative paradigms, "analysis of textual data", "analysis of natural language", "quantitative content analysis", "qualitative content analysis" and the theoretical and methodological background of their application are considered. The world experience of classification of methods of quantitative analysis of texts in social sciences and the existing contradictions between the various schools of content analysis are briefly presented. The necessity to update the requirements to the content analysis research in conditions of transition to the information society and use of computer technology in the study of communication is proved.

Keywords: content analysis, text mining, computer content analysis, text data analysis, quantitative content analysis, qualitative content analysis.

Розвиток комп'ютерних технологій та, зокрема, мережі Інтернет як практично невичерпного і постійно оновлюваного джерела текстів для аналізу не лише відкриває нові можливості для дослідників комунікації, а й

спричиняє появу незнаних до того викликів для методологів соціальних наук. Йдеться не лише про питання технології проведення контент-аналітичних досліджень в епоху інформаційного буму, а й про термінологічні суперечності, що виникають під час обговорення методів, методик і технік дослідження змісту.

У вітчизняних публікаціях дедалі частіше зустрічаються такі назви методів і методик, як "текст-майнінг" [1, с. 80-88], "комп'ютерний контент-аналіз" [2, с. 397-401], "аналіз текстових даних" [3, с. 39-42], "кількісний контент-аналіз" (на противагу якісному) [4, с. 17-26]. Вказані словосполучення в своїй більшості є запозиченнями з англійської літератури, де дана проблематика обговорюється значно ширше і протягом тривалішого часу, ніж у нас. Водночас, бракує чіткого термінологічного розмежування та співвіднесення даних методів з класичним контент-аналізом, основи якого були закладені ще в середині минулого століття. У вітчизняних дослідженнях текстів¹, судячи з практики вживання наведених назв, контент-аналіз вважається своєрідною "методологічною парасолею", яка охоплює всі засоби кількісно-якісного аналізу текстів в соціальних науках: зустрічаємо такі словосполучення, як "статистичні методи контент-аналізу", лінгвістичний аналіз як вид комп'ютерного контент-аналізу та ін. Відповідно, до всіх даних засобів застосовуються процедурні вимоги, розроблені для класичного контент-аналізу.

Мета статті – представити доцільне в сучасних умовах концептуальне визначення контент-аналізу як методу у співвіднесенні з методиками і техніками кількісного аналізу (а точніше — кількісної обробки) текстів. Така спроба класифікації є актуальною не лише з точки зору потреби у термінологічній уніфікації, але і як початок обговорення оновлених вимог до контент-аналітичного дослідження з врахуванням методологічних напрацювань останніх років.

Перш за все, варто з'ясувати, що ми розуміємо під методологією, методом, методикою та технікою. Методологія — базове поняття, яке можна розглядати як систему найзагальніших принципів, підходів, методів, що складають основу даної галузі науки. Методологія нерозривно пов'язана з теоретичними засадами даної науки (наукової галузі). Якщо ми розглядаємо соціологію масових комунікацій як галузь соціологічної науки, а дослідження текстів масової комунікації як один з напрямків даної науки, то контент-аналіз є базовою методологією даного напрямку. Класичний контент-аналіз², власне, і розроблявся для дослідження текстів мас-медіа [5, с. 175]. Таке визначення досить поширене в англійській літературі [6, р. 9], однак варто зазначити, що протягом останніх 20 років термін "методологія" все частіше вживається як синонім "методу", а розрізнення робиться лише у випадках дослідження методів (як, наприклад, теоретичний аналіз в даній статті), а не методології (методів) дослідження³.

Для вітчизняної наукової традиції (і на пострадянському просторі загалом теж [8, с. 19]) більш прийнятно визначати контент-аналіз як метод [5, с. 175]. Під методом ми розуміємо спосіб досягнення цілі, сукупність прийомів і операцій теоретичного чи практичного освоєння дійсності, а також людської діяльності, організованої певним чином. Метод у науці — це також і заданий відповідною гіпотезою шлях осягнення предмета вивчення [9, с. 628]. Контент-аналіз як метод є не лише сукупністю загальних принципів осягнення текстів з точки зору соціальних теорій, а й певною сукупністю процедур опрацювання цих текстів. Наразі існує велика кількість підходів до проведення контент-аналітичного дослідження, тому тут ми зупинимось лише на тих процедурних етапах, які характеризують класичний контент-аналіз та задають його специфіку відносно інших методів дослідження текстів.

Перший етап — теоретико-методологічний, на якому визначаються предмет та об'єкт дослідження, формулюються гіпотези. Специфіка контент-аналізу в тому, що головним предметом аналізу є зміст, заданий соціальним контекстом, в якому існує текст, а метою аналізу — висновки щодо соціальної реальності [10, р. 33]. Гіпотези дослідження іманентно включають контекст джерел появи тексту, його значення, інтенцій комунікатора або потенційних ефектів. Текст як об'єкт аналізу розглядається в найширшому розумінні — писемний текст, аудіо-, відеозаписи, зображення і т.д., тобто як певна смислова єдність [11, с. 45, 12, S. 1-15].

Контент-аналіз виник у рамках позитивної методології вивчення актів комунікації, тому другий процедурний етап задається логікою кількісного дослідження: це розробка категорій аналізу, обґрунтування вибірки, одиниць аналізу і рахунку. Методологічними максимами тут є валідність, надійність та об'єктивність. Для контент-аналізу ці вимоги набувають певної специфіки через якісну сутність тексту [10, р. 33]. Об'єктивність є виключно інтерсуб'єктивною і досягається через формалізацію

¹ Варто зазначити, що значна частина зарубіжних дослідників, які працюють з текстами, також вживає "контент-аналіз" у різних, часто несумісних контекстах.

² Під класичним контент-аналізом ми маємо на увазі методологічні засади формалізованого аналізу текстів закладені в середині ХХ ст. Г. Ласуеллом та Б. Берельсоном, оскільки їх праці є відправною точкою для всіх сучасних контент-аналітиків, а термін "контент-аналіз" концептуалізується в контексті загальної методології соціальних наук.

³ Див. зауваження щодо проблем використання слова "methodology" у [6, р. 1027].

тексту, а "валідність" має сприйматись у епістемологічному розумінні як "валідність, доведена для нас" [13, р. 95-99]. Досягнення інтерсуб'єктивної узгодженості відбувається на наступному етапі претесту.

Далі проводиться збір і кодування даних згідно розроблених дослідником правил опрацювання інформації [14, S. 212]. На наступному етапі, в рамках класичної методології контент-аналізу, проводиться статистичний аналіз зібраних даних [15, р. 49-69] та їх інтерпретація згідно теоретико-методологічних рамок, що задані на першому етапі.

Виходячи з викладеного, ми можемо визначити контент-аналіз як метод (або методологію) отримання надійних та валідних висновків з текстів у соціальному контексті їх використання з допомогою інтерсуб'єктивно узгоджених правил систематизації інформації.

Деякі фрагменти описаної процедури викликають суперечності: якісний характер об'єкту аналізу, використання статистики та потреба в теоретично навантаженій інтерпретації. Останнім часом існує тенденція, у тому числі у вітчизняній соціології [4, с. 17-26], до визначення контент-аналізу як кількісно-якісного методу. Ми вважаємо таке визначення, до певної міри, тавтологічним і таким, що не вирізняє контент-аналіз серед інших методів соціальних наук. По-перше, будь-яке дослідження соціальної реальності передбачає якісну інтерпретацію, якщо не "сирих" даних, то результатів їх статистичної обробки. Під час здійснення дослідником-"кількісником" експлораторного аналізу даних чи використання візуалізації даних для визначення тенденції, взагалі не існує числового стандарту, який би визначав відсутність чи наявність паттерну. По-друге, якісний характер даних не виключає їх аналізу без, власне, інтерпретативного "прочитання", що ми побачимо далі, розглядаючи технологію опрацювання текстових масивів. По-третє, в умовах сучасних міждисциплінарних досліджень та поєднання методів ("mixed-methods"), "методологічна" чистота кількісних та якісних методів є, радше, архаїзмом. Тому метод наразі варто визначати радше за теоретико-методологічними та епістемологічними засадами, на яких він базується (для контент-аналізу характерний радше позитивний підхід до тексту), ніж за набором технік та методик. Крім того, класичний контент-аналіз є таки кількісним методом, навіть з цієї точки зору [16, р. 40-52].

Якісні методики контент-аналізу можна віднести до пост-класичного етапу розвитку цієї методології протягом останніх десятиліть [10, р. 15-16]. Якісний контент-аналіз має певні процедурні особливості [17, р. 290-296]. Дослідник має ретельно прочитати аналізовані тексти та записати висновки, які виникають у результаті його попередньої інтерпретації. Далі проводиться агрегація текстів шляхом відбору фрагментів найбільш значущих з точки зору дослідницьких питань. Агреговані дані структуруються та відображаються відповідно до теоретичної моделі. Базуючись на даному відображенні дослідник ідентифікує домінуючі теми, паттерни, зв'язки та виключення ("замовчування"). Створена в результаті аналізу схема наповнюється цитатами, що підтверджують виділені теми та зв'язки між ними. Якщо виникає необхідність, аналітичні категорії можна додавати, виключати чи модифікувати, щоб забезпечити їх максимальну вичерпність на насиченість ("data saturation").

Не дивлячись на невідповідність описаної процедури класичній, її не варто вважати окремим методом, оскільки формалізація кожного з етапів, теоретико-методологічні засади та розуміння тексту відповідають наведеному нами загальному визначенню контент-аналізу. Це, радше, *методика*, тобто набір готових алгоритмів проведення дослідження в конкретній ситуації. Так, в рамках контент-аналітичного дослідження може застосовуватись як якісна методика, так і кількісна, або їх поєднання. Тому якісний контент-аналіз — це методика аналізу текстів шляхом систематичної інтерпретації текстів через виділення в них тем, тенденцій і відповідне їх кодування. Кількісний контент-аналіз — методика аналізу текстів шляхом кількісного співставлення аналітичних категорій через підрахунок ключових слів (чи інших одиниць рахунку). Для чіткішого вирізнення можна говорити про кількісні та якісні методики контент-аналізу.

Конкретними процедурами вирізняються і *техніки* т.зв. "комп'ютерного контент-аналізу", тобто застосування під час контент-аналітичного дослідження комп'ютерних засобів. Під технікою ми розуміємо сукупність спеціальних прийомів для ефективного використання певного методу. В англійській літературі вирізняють комп'ютерний контент-аналіз, що застосовується у якісних (computer-aided) та кількісних (computer-assisted) контент-аналітичних дослідженнях [18, р. 668]. Ми пропонуємо перекладати дані терміни як "комп'ютеризований" та "автоматизований", відповідно. У німецькомовній літературі окремих термінів не вживають (раніше — "elektronische Inhaltsanalyse" [19], зараз, переважно, — "computerunterstützte Inhaltsanalyse" [20, S. 155-184]), а відмінність визначається контекстом.

Не будь-який кількісний аналіз тексту є контент-аналізом. У соціальних науках останнім часом досить популярний т. зв. "текст-майнінг"⁴. Він може розглядатися, з одного боку, як набір технік для виявлення в текстових даних нетривіальних тенденцій, які можуть зацікавити дослідника (в т.ч. контент-аналітика), а з

⁴ Усталеного перекладу, поки що, не існує. Найчастіше використовують транслітерацію англійського терміну або переклад "дубування [інформації] з тексту", "інтелектуальний аналіз текстових даних", "глибинний аналіз текстових даних".

іншого — як міждисциплінарний напрямок досліджень, що охоплює методики обробки інформації, машинного навчання, статистичної класифікації, роботи з базами даних та ін., які можуть бути застосовані до тексту.

У загальному вигляді текст-майнінг передбачає два етапи: *переробку* довільного *тексту* (text refining) у певну проміжну форму та виділення знання (knowledge distillation), тобто ідентифікацію паттернів у проміжній формі [21, р. 65-70]. Проміжні форми можуть бути напівструктурованими (наприклад, концептуальна діаграма) та структурованими (наприклад, масиви даних сполучення слів). Одиницями аналізу можуть бути документи або концепти. В першому випадку для "виділення знання" використовуються методики статистичної класифікації документів (наприклад, кластерний аналіз), у другому — статистичні моделі зв'язків між об'єктами (наприклад, асоціативний аналіз).

Теоретико-методологічним підґрунтям текст-майнінгу є, перш за все, теорії обчислювальної лінгвістики⁵ (computational linguistics) як галузі інформатики (computer science). Об'єктом аналізу є *текстові дані* — цілісні послідовності символів (літер, цифр, знаків пунктуації і т.д.), які разом з метаданими містять змістовну інформацію. Метадані — інформація, яка не містяться безпосередньо в тексті, але яка відома про текст (наприклад, мова, час створення, авторство). Таким чином, текст-майнінг не оперує широким визначенням тексту, а стосується саме писемного мовлення.

Ще одним із напрямків кількісного аналізу тексту є аналіз природної мови (natural language processing). Це міждисциплінарна галузь науки, що охоплює методики обчислювальної лінгвістики та теорії штучного інтелекту, основним проблемним полем якої є забезпечення взаємодії людських комунікативних актів (вербальних та невербальних) та комп'ютерних систем. Отже, теоретико-методологічним підґрунтям аналізу природної мови є формальні теорії мови [22, р. 113-124, 23, р. 9977-9982] та штучного інтелекту. Розуміння тексту в даному разі є ширшим, ніж у межах текст-майнінгу, але вужчим, ніж в контент-аналізі: об'єктом аналізу є не лише писемний текст, а й звуки, відео-зображення, будь-які інші засоби передачі змістовної інформації (включаючи дотик, жести), однак повна автоматизація процесу виключає врахування теоретично навантаженого контексту.

На позначення сукупності методик і технік, що оперують числовим представленням текстів (у межах кількісного контент-аналізу, текст-майнінгу, аналізу природної мови та ін.) ми пропонуємо вживати термін *кількісний аналіз текстових даних*. Узагальнення викладених концептів та їх співвідношення представлено в таблиці 1.

Таблиця 1.

Методи, методики та техніки аналізу тексту

	Концептуально-теоретичне підґрунтя	Загальна мета
Контент-аналіз — <i>метод (методологія)</i> отримання надійних та валідних висновків з текстів у соціальному контексті їх використання з допомогою <i>інтерсуб'єктивно</i> узгоджених правил систематизації інформації.	Соціальні теорії, перш за все — теорія комунікації. Текст — з функціонально-смыслового погляду певна упорядкована група висловлювань або їх аналогів, які, завдяки семантичним і функціональним взаємовідношенням елементів являють собою завершену смыслову єдність. Текст в широкому розумінні: писемний текст, аудіо-, відеозаписи, зображення і т.д.	Висновки щодо <i>соціальної реальності</i> , з якою змістовно пов'язаний текст.
Якісний контент-аналіз — <i>методика</i> аналізу текстів шляхом систематичної <i>інтерпретації</i> текстів через виділення в них тем, тенденцій і відповідне їх кодування.		
Комп'ютеризований контент-аналіз (computer-aided content analysis) — <i>техніка</i> кодування текстів з використанням комп'ютерних засобів для систематизованого запису результатів роботи кодувальника (аналітика).		
Кількісний контент-аналіз — <i>методика</i> аналізу текстів шляхом <i>кількісного порівняння аналітичних категорій</i> через підрахунок ключових слів (чи інших одиниць рахунку).		
Автоматизований контент-аналіз (computer-assisted content analysis) — <i>техніка</i> автоматизованого кодування текстів та підрахунок в них ключових слів (чи інших одиниць рахунку) з допомогою комп'ютерних засобів.		
Текст-майнінг — (1) набір технік для виявлення в текстових даних нетривіальних тенденцій, які можуть зацікавити дослідника; (2) міждисциплінарний напрямок досліджень, що охоплює	Теорії обчислювальної лінгвістики. Текст як <i>дані</i> — цілісні послідовності символів (літер, цифр, знаків пунктуації і т.д.), які разом з метаданими містять змістовну інформацію.	Висновки щодо інформаційних паттернів в текстових даних.

⁵ Інші варіанти перекладу: "комп'ютерна", "математична" лінгвістика.

методики обробки інформації, машинного навчання, статистичної класифікації, роботи з базами даних та ін., які можуть бути застосовані до тексту.		
Аналіз природної мови (natural language processing) — міждисциплінарна галузь науки, що охоплює методики обчислювальної лінгвістики та теорії штучного інтелекту.	Формальні теорії мови та штучного інтелекту. Текст без теоретично навантаженого контексту.	Забезпечення взаємодії людських комунікативних актів (вербальних та невербальних) та комп'ютерних систем.

У сучасних умовах контент-аналіз як провідний метод вивчення змісту комунікації в соціальних науках, безумовно, передбачає використання напрацьованих з описаних вище галузей. При цьому виникають певні проблеми методологічного характеру, пов'язані із застарілістю вимог, орієнтованих на методологічну "чистоту" класичного контент-аналізу для аналізу "великих" і динамічних даних. Так, на етапі кодування великих масивів текстових даних (наприклад, висвітлення певної проблеми в десятках тисяч повідомлень соціальних мереж) доцільно застосовувати експлораторні методики текст-майнінгу, оскільки розробка вичерпної кодувальної схеми на підготовчому етапі є проблематичною. Не видається можливим в цьому випадку говорити і про інтерсуб'єктивну узгодженість категорій: перевірка узгодженості можлива лише при "ручному" кодуванні. Доцільніше говорити про валідність кодувальних моделей (а не категоріальних схем), тобто таких алгоритмів автоматизованого кодування текстів, які б максимально зберігали їх змістовну насиченість. Проблемою є також переважно дискретне розуміння категорій в контент-аналізі: кожна категорія може мати певне одичне числове вираження (наприклад, кількість слів, закодованих певним модусом). Не відповідні такому підходу є такі просторові кодувальні моделі, як шкалювання слів (word scale) [24, р. 311-331], що не базується на апріорних словникових схемах. Переглянути слід і підхід до визначення вибірки контент-аналітичного дослідження: вона не повинна бути конвенційною [25, р. 873-882], але не може бути простою випадковою для таких умовно нескінченних сукупностей, як інформація світової мережі. Доцільніше говорити про випадкову вибірку з використанням індикаторів та структурних характеристик баз даних [26].

Підсумовуючи, зазначимо, що контент-аналіз як один з найстаріших методів дослідження текстів актуальний і сьогодні. Він динамічно розвивається в середовищі міждисциплінарних досліджень та технологічно вдосконалюється за рахунок розробок в різних галузях аналізу текстових даних. При цьому не слід забувати про методологічні особливості контент-аналізу, головними з яких є зв'язок із соціальною теорією та соціальним контекстом.

Література:

1. Лисенко С. М. Розробка та дослідження алгоритмів інтелектуальної обробки текстів (text mining) / С. М. Лисенко // Наукові праці [Текст] : наук.-метод. журн. — Миколаїв : ЧДУ ім. П. Могили, 2011. — Вип. 148. — С. 80–88.
2. Семотюк О. Комп'ютерний контент-аналіз: основні завдання, сфери застосування, переваги та недоліки / О. Семотюк // Вісник Львівського університету. Серія журналістики. — 2004. — Вип. 25. — С. 397–401.
3. Павлишенко Б. М. Групування текстових даних на основі моделі семантичного контексту / Б. М. Павлишенко // Восточно-Европейский журнал передовых технологий. Информационные технологии. — 2011. — № 5/2 (53). — С. 39–42.
4. Бондар В. С. Розгляд методу контент-аналізу з погляду кількісно-якісних технік проведення / В. С. Бондар, М. А. Допіра // Наукові записки НаУКМА. — К.: Національний університет "Києво-Могилянська академія", 2007. — Т. 70, "Соціологічні науки". — С. 17–26.
5. Соціологія: Навчальний посібник / С. О. Макеєв (ред.). — К.: "Українська енциклопедія" ім. М. П. Бажана, 1999. — С. 175.
6. Neuendorf, K. The Content Analysis Guidebook / Kimberly Neuendorf. — Thousand Oaks, CA: Sage Publications, 2002. — P. 9.
7. The American Heritage Dictionary of the English Language. — Houghton Mifflin Company, 2009. — P. 1027.
8. Методологические и методические проблемы контент-анализа: тезисы докладов рабочего совещания социологов / Андрей Григорьевич Здравомыслов (Ред.). — М.: Институт социологических исследований Академии наук СССР, 1973. — Т. 1 — С. 19.
9. Новейший философский словарь / А.А. Грицанов (Ред.). — Мн.: Книжный Дом, 2003. — С. 628.
10. Krippendorff K. Content analysis : an introduction to its methodology / Klaus Krippendorff. — Thousand Oaks : Sage Publications, Inc., 2004.

11. Одинцов В.В. Стилистика текста. — М.: Наука, 1980. — С. 45.
12. Pfützte M. Grundgedanken zu einer funktionalen Textlinguistik / Max Pfützte // Textlinguistik, 1970. — № 1. — S. 1–15.
13. Іванов О.В. Кількісний контент-аналіз: проблема контексту / О.В. Іванов // Вісник Харківського національного університету імені В. Н. Каразіна. Серія: Соціологічні дослідження сучасного суспільства : методологія, теорія, методи ; вип. 30 — Харків: Харківський національний університет ім. В.Н. Каразіна, 2012. — № 999. — С. 95–99.
14. Kromrey H. Empirische Sozialforschung. Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung / Helmut Kromrey. — Stuttgart: Lucius & Lucius Verlag, 2006. — S. 212.
15. Weber R. P. Basic content analysis. / Robert Philip Weber — Newbury Park, CA: Sage, 1963. — P. 49 – 69.
16. Language of Politics: Studies in Quantitative Semantics / Harold D. Lasswell, Nathan Leites (Eds.). — Cambridge, Mass.: MIT Press, 1949. — P. 40–52.
17. Romanowski M. What you don't know can hurt you: Textbook omissions and 9/11 / Michael H. Romanowski // Clearing House, 2009. — № 82. — P. 290–296.
18. Bos W., Tarnai C. Content analysis in empirical social research / Wilfried Bos, Christian Tarnai // International Journal of Educational Research. — 1999. — № 31. — P. 668.
19. Deichsel, A. Elektronische Inhaltsanalyse. Zur quantitativen Beobachtung sprachlichen Handelns. / Alexander Deichsel. — Berlin, 1975.
20. Züll C., Mohler P. Textanalyse. Anwendungen der computerunterstützten Inhaltsanalyse / Cornelia Züll, Peter Mohler. — Opladen: Westdeutscher Verlag, 1992. — S. 155–184.
21. Tan, A. H. Text mining: the state of the art and the challenges / Ah-Hwee Tan // Proceedings of the PAKDD workshop on knowledge discovery from advanced databases. — Beijing, 1999. — P. 65–70.
22. Chomsky N. Three models for the description of language / Noam Chomsky // IEEE Transactions on Information Theory. — 1956. — Vol. 2, № 3. — P. 113–124.
23. Bates M. Models of natural language understanding / Madeleine Bates // Proceedings of the National Academy of Sciences of the United States of America. — 1995. — Vol. 92, № 22. — P. 9977–9982.
24. Laver M. Extracting policy positions from political texts using word as data / Michael Laver, Kenneth Benoit , Garry John // American Political Science Review. — 2003. — V. 97 (2). — P. 311 – 331.
25. Riffe D., Freitag A. A content analysis of content analyses: 25 years of Journalism Quarterly / Daniel Riffe, Alan Freitag // Journalism & Mass Communication Quarterly. — 1997. — V. 74. — P. 873 – 882.
26. BarYossef Z., Gurevich M. Random Sampling from a Search Engine's Index [Електронний ресурс] / Ziv BarYossef, Maxim Gurevich // Journal of the ACM. — 2008. — V. 55 (5). — Режим доступу: <http://www2006.org/programme/files/pdf/3047.pdf>.