
Федорчук А. Г.
Танатар Н. В.

**Автоматизированные
технологии формирования
информационно-аналитических
систем на основе исследования
информационных потоков СМИ**

Предлагаются качественно новые подходы к обобщению больших информационных массивов, генерируемых СМИ. На основе использования классификаторов фасетно-иерархического типа осуществляется отбор материалов с их индексацией и занесением в БД для дальнейшей многоаспектной аналитико-синтетической обработки и изучения. Процедура формализации и квантификации текста гарантирует объективность полученных аналитических материалов.

События последних лет наглядно продемонстрировали, что украинское общество находится сегодня на непростом пути поиска своего будущего. Продолжаются поиски своего места в мире, в истории, поиски национальной идеи и путей объединения востока и запада Украины, политической, экономической и социальной модели общественной жизни. Выборы в органы власти, как правило, обостряют существующие противоречия и побуждают общество к более четкому и прагматичному определению своих приоритетов, жизненных позиций и политических пристрастий. Последние выборы в Верховный Совет Украины не стали исключением. К тому же, в условиях слабого политического структурирования и изменений политической модели власти в Украине, общество продемонстрировало несвойственную ему активность в определении характера и направлений развития своей страны.

Существенную роль в процессах, происходящих в Украине в последнее время, играют средства массовой информации (СМИ). Медиа-пространство отображает все разнообразие мыслей и взглядов, существующих в обществе, и одновременно принимает участие в их формировании и кристаллизации, влияя на политическую и социальную жизнь общества. Таким образом, исследование информационного пространства и, в частности, информационных потоков, которые генерируют СМИ, сегодня приобретает особенное значение. Именно СМИ становятся источником и объектом анализа в сфере изучения общественно-политических процессов, происходящих в Украине.

Каким образом СМИ влияют на политические результаты? Какие политические «функции» они выполняют? Представители разных научных школ дают разные, иногда противоречивые ответы на эти вопросы. Например, приводят очень длинный, но неполный перечень потенциальных функций СМИ. Среди них: обеспечение информационной базы для реализации гражданами политического выбора; обеспечение канала коммуникации между политиками, политическими движениями, органами государственной власти и общественностью; помощь другим институциям в осуществлении таких функций, как отбор политических лидеров и репрезентация их обществу; влияние на процесс принятия политиками решений; помощь в организации дискуссий между различными политическими силами; определение приоритетов общественных дискуссий путем освещения отдельных политических вопросов; сохранение status quo; поддержка интеллектуального влияния правящих элит; создание рынка политических идей и т. д.

Демократизация общественно-политической жизни в Украине актуализирует проблему открытости коммуникативного пространства государства, которая является предметом научных исследований государственных и общественных институций, объектом повышенного внимания широкой общественности как в Украине, так и за ее пределами. Такая тенденция обуславливает необходимость усовершенствования методик обработки политической ин-

формации, которая отображает межинституциональные связи составляющих политической системы, с одной стороны, и основных институтов государственной власти и гражданского общества, – с другой.

Особенности изучения документных потоков социально-политической тематики

Решение проблемы исследования влияния СМИ на социально-политическую жизнь страны требует знаний особенностей процессов создания, движения и куммулирования документов в обществе, которые происходят посредством функционирования документных потоков (ДП) и массивов (ДМ).

Безусловно, одной из главных особенностей нашего времени есть постоянный рост темпов производства информации. Это объективный процесс и, в общем, позитивный. Но, следует отметить, что прогресс в отрасли производства информации ведет, по мнению большинства ученых, к снижению общего уровня информированности общества. Резкий рост темпов ее производства привел к целому ряду проблем: непропорциональный рост информационного «шума» из-за слабой структурированности информации; появление избыточной информации; несоответствие формально релевантной информации (тематически соответствующей) реальным нуждам потребителей информации; многоразовое дублирование информации в разных источниках [1].

В силу изложенных факторов традиционные информационно-поисковые системы постепенно теряют свою актуальность. Причина этого не столько в физических объемах информационных потоков, сколько в их динамике, то есть в постоянном систематическом обновлении информации, которое далеко не всегда имеет очевидную регулярность. Обобщение больших информационных массивов, постоянно генерируемых в СМИ, требует качественно новых подходов.

Задание систематического отслеживания тенденций и процессов в динамической информационной среде может быть решено с использованием мониторинга содержания ДП. Это перспектив-

ное направление в специальной литературе определяется как контент-мониторинг (КМ) и используется для непрерывного получения качественных и количественных срезов ДП на протяжении неопределенного заранее промежутка времени. Самой главной методологической составляющей КМ является контент-анализ (КА) [2].

**Использование метода контент-анализа
для формирования документных массивов
политической направленности на материалах прессы**

Пресса – настоящий барометр политической погоды. По оперативности подачи материалов и масштабности влияния на население газетные издания конкурируют разве что с радиовещанием и телевидением, но им нет равных среди документных средств коммуникации. Однако, специализированным аналитическим службам, занимающимся анализом и прогнозированием социально-политической ситуации в Украине, довольно сложно исследовать весь разнообразный и изменяющийся поток газетной периодики. Вместе с тем, при составлении полноценного аналитического обзора или прогноза развития общественно-политической ситуации лучше всего иметь дело с текстами первоисточников, а не с их рефератами или аннотациями. При анализе содержания коммуникаций важно иметь количественные показатели упоминания на страницах газет того или другого события, персоны, знать в каком контексте о них говорится, каким образом информация интерпретируется прессой. В связи с этим возник вопрос усовершенствования способов обработки и отображения такой информации, повышения ее наглядности и доступности. В Отделе организации и использования документального фонда Национальной библиотеки Украины им. В. И. Вернадского (далее ООИДФ НБУВ), учитывая перечисленные требования, для удовлетворения потребностей заказчиков разработаны информационно-аналитические системы по ключевым фрагментам публикаций (ИАС). Технология формирования этих систем предусматривает отбор материалов с их дальнейшей индексацией и занесением в

базы данных (БД) вместе с выхдными данными. Таким образом, формируется документный массив ИАС для дальнейшего его изучения.

Контент-анализ оказался наиболее подходящим для проведения данной процедуры, так как объектом исследования данного метода выступает именно текст [3]. За время своего почти 100-летнего существования КА занял достойное место среди других современных методов исследования СМИ. Он закрепился как научный метод изучения документов, характеризующийся принципами объективности, детального и полного рассмотрения объекта исследования, строгого соблюдения процедуры. Большие возможности КА объясняются органическим соединением качественной и количественной составляющих исследования. Высокие требования к этому методу делают выводы таких исследований более надежными и достоверными, чем в случае использования качественных методов.

Сущность КА состоит в приведении всех текстов исследуемой проблемы к набору определенных элементов, которые в дальнейшем подлежат подсчету и анализу. В основе количественной составляющей КА – частота появления в документах определенных характеристик содержания, тогда как его качественная составляющая основывается на самом факте присутствия или отсутствия в тексте одной или нескольких характеристик содержания. На основе изменений, которые отображены в текстах, можно делать выводы о тенденции развития общества или функционирования соответствующих социальных институтов. Именно поэтому объектом анализа в таких исследованиях является текст публикации.

КА, который в рамках исследования ДП предусматривает анализ массива текстовых документов – результатов мониторинга информационного пространства, используется в ООИДФ НБУВ, начиная с 1993 года. На основе данного метода формируются БД ИАС разной социально-политической направленности [4–7].

При создании и функционировании системы контент-мониторинга по ключевым фрагментам публикаций используются следующие принципы КА:

- обработка больших потоков газетной периодики;
- выделение из публикаций компактных фрагментов, которые по содержанию отвечают теме исследования;
- сортировка этих фрагментов и объединение их в однотипные по содержанию подгруппы.

Использование методологии КА в нашем случае обусловлено, прежде всего, необходимостью получения результатов отбора материалов, независимых от взглядов или впечатлений информационного аналитика. Объективность при этом достигается тем, что личные убеждения исследователя нивелируются процедурой формализации и квантификации текста. В рамках публикации выделяются отдельные фрагменты, отображающие разные аспекты исследуемой проблемы, которые потом в системе сортируются и при необходимости подсчитываются. Реализация этой процедуры требует специально разработанного детального классификатора, по которому аналитик отбирает и индексирует материал в БД [8].

Для формального описания документов при формировании БД системы по ключевым фрагментам публикаций были выбраны классификаторы фасетно-иерархического типа, где каждому элементу исследуемой проблемы отвечает свой фасет с фиксированным месторасположением в фасетной формуле, а совокупности значений в рамках каждого элемента – соответствующий фасетный класс. Кроме того, с помощью отдельного фасета (фасет модальности) передается отношение автора публикации (позитивное, нейтральное, негативное) к описанному им событию или персоне.

Предусматривается дальнейшее развитие и усовершенствование классификатора в зависимости от развития проблемы – на уровне структуры добавляются новые элементы с совокупностью соответствующих им значений. Параллельно фасетная формула пополняется аналогичными этим элементам фасетами и фасетными классами. Что касается значений элементов, выделенных ранее, то некоторые из них со временем теряют свою актуальность и изымаются, или модифицируются, вместе с тем появля-

ются и новые значения. То есть развитие проблемы проявляется в наращивании индексов в фасетной формуле и их конкретного содержательного наполнения в классификаторе.

Отбор ключевых цитат происходит следующим образом. Из текста выделяются фрагменты публикаций, которые в рамках сформированного классификатора несут черты целостного наименьшего модуля исследуемой проблемы. После этого в рамках выделенного модуля устанавливаются отдельные элементы проблемы (соотносятся с конкретными значениями классификатора) и связи между ними. При этом семантические элементы текста и классификатора или совпадают, или имеют одинаковое содержательное наполнение. В результате такого анализа отобранной цитате приписывается набор конкретных индексов (кодов), каждый из которых отвечает своему значению в классификаторе и занимает строго фиксированное место в совокупности выбранных фасетов формулы. С помощью этой процедуры совершается формальное описание конкретного фрагмента текста, если же их несколько в публикации, то совокупность фасетных формул всех выделенных фрагментов обеспечивает формальное описание всего документа в контексте исследуемой проблемы. Следовательно, информация, которая не касается исследования, в БД не попадает. Таким образом, предложенный способ обработки первоисточников выполняет функцию информационного «сита», которое пропускает только релевантную исследованию информацию в виде «выжимок» оригинального текста. Окончательный формат каждой записи БД включает все выделенные в тексте и заиндексированные по содержанию цитаты с полным библиографическим описанием публикации.

Среди достоинств данной технологии при формировании БД прежде всего следует выделить сведение информационного «шума» к минимуму. Кроме того, при сохранении оригинального текста в виде его «выжимок», сформированные БД компактны и удобны для использования.

На этапе обработки БД программные средства дают возможность разбивать входной документ (библиографически описан-

ные заиндексированные цитаты) на три независимые составляющие. Каждая из них функционирует в системе самостоятельно. Первая составляющая – содержательная информация (цитата публикации), вторая – ее полное библиографическое описание и третья – структурная характеристика выделенного фрагмента текста, которая формируется с помощью фасетной формулы. Такая процедура позволяет подавать весь документный массив в виде публикационной матрицы, которая может анализироваться как отдельно по каждой из выделенных составляющих, так и во всей совокупности. В автоматизированной процедуре анализа входной информации заложены большие возможности для аналитико-синтетической обработки БД с получением вариантов структурирования текстовой составляющей и разных количественных характеристик ДП (графики, диаграммы, статистические таблицы).

Особенность данной технологии КМ газетной периодики состоит еще и в том, что анализ и синтез ключевых цитат публикаций из широкого круга источников осуществляется по продуманным и заданным схемам на основании разработанного классификатора. Иначе говоря, цитаты, лаконично передающие заложенную в первоисточнике идею, могут объединяться разными способами в рамках параметров, представленных в фасетной формуле.

Технологически предусмотрено, что главным может быть любой фасет. В выходном материале возможно выделение любого из представленных в БД аспектов проблемы (в зависимости от потребностей заказчиков) или проблема анализируется в целом, с учетом всех ее параметров. Выходные информационные продукты имеют вид структурированных ключевых цитат публикаций с соблюдением определенной тематической иерархии, дополненных статистическими данными и другими иллюстративными материалами. Для получения таблиц чисел и графических изображений используются фасетные индексы цитат и библиографические описания публикаций. Количественные подсчеты ключевых цитат позволяют сделать объективные выводы о направленности исследуемых материалов по количеству употреблений единиц анализа. Таким образом, и в технологии обработки первоисточ-

ников, и в программных средствах конвертирования информационных объектов заложен потенциал разнообразного формирования выходной информации [9].

По данной технологии на базе ДП прессы формировались системы КМ социально-политических событий общегосударственного значения, таких как выборы Президента и народных депутатов Украины, конституционные акты и т. д. Последние наработки касались выборов в Верховный Совет Украины 2006 года.

Контент-мониторинг ДП газетной периодики по теме «Выборы-2006»

Особенностью контент-мониторинга предвыборной кампании в Верховный Совет 2006 года было отслеживание информации по двум параллельным направлениям. Исследовались публикации отечественных газет, отображавшие репутацию и характер участия в предвыборных гонках, как политических партий, так и общенациональных политических лидеров Украины. Выбор данной стратегии оказался очень удачным, поскольку парламентские выборы, которые впервые в Украине проводились на пропорциональной основе, обрели довольно персонифицированный характер. Партия Регионов Украины однозначно ассоциировалась с ее лидером Виктором Януковичем, «Народный Союз «Наша Украина» с действующим Президентом Виктором Ющенко, «Блок Юлии Тимошенко» – с Юлией Тимошенко, Социалистическая партия – с Александром Морозом и т. д. Полную картину, таким образом, можно было исследовать только соотнеся два разноплановых информационных потока, освещающих политические партии и их лидеров. Одновременно контент-мониторинг каждого из этих потоков также имел отдельную ценность, поскольку позволял аналитикам делать выводы о развитии и тенденциях двух параллельных процессов в обществе.

Для иллюстрации возможностей данной технологии детально остановимся на анализе материалов за период с января по март 2006 года, касающихся политических партий и блоков-победителей в предвыборных гонках.

Немного неожиданным оказалось то, что информация, генерируемая газетной прессой в период предвыборной борьбы, по тональности была менее поляризованная, чем в период предыдущих президентских выборов. 68% публикаций было определено аналитиками как нейтральная (график 1). То есть преобладал объективный непредвзятый характер подачи журналистами информации. Доля негативно окрашенных материалов составляла всего 15 %, позитивно – 17 %. Публикации, касающиеся лидеров гонок, были более поляризованные (график 2). Так, о Партии Регионов Украины с отрицательным оттенком писалось в 18% публикаций, о «Блоке Юлии Тимошенко» – в 21%. Положительно Партия Регионов оценивалась в 22% публикаций, «Блок Юлии Тимошенко» – в 20%.

Интересен тот факт, что из региональных изданий про обе политические силы больше всех писали газеты Львовской области. При этом во львовских изданиях преобладали нейтральные по тональности публикации, а доля отрицательных характеристик обоих участников гонок оказалась почти одинаковой. Еще более неожиданную позицию заняли газеты восточного региона: донецкие и луганские издания практически не публиковали на своих страницах негативных публикаций. О «Блоке Юлии Тимошенко» газеты восточного региона практически не писали, а часть публикаций в поддержку Партии Регионов почти равнялась части нейтральной информации.

Иная картина наблюдалась в тональности освещения газетами пропрезидентской политической силы – «Народного Союза «На-

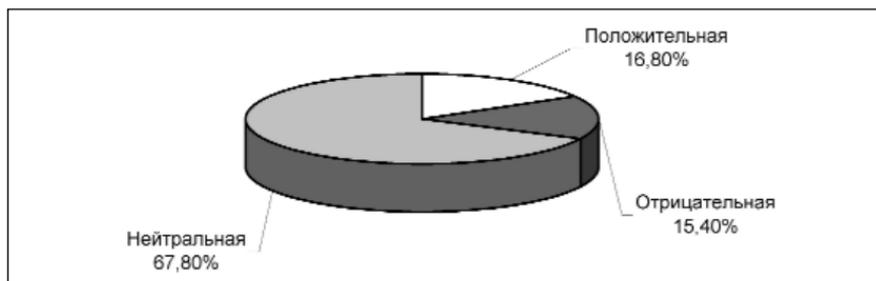
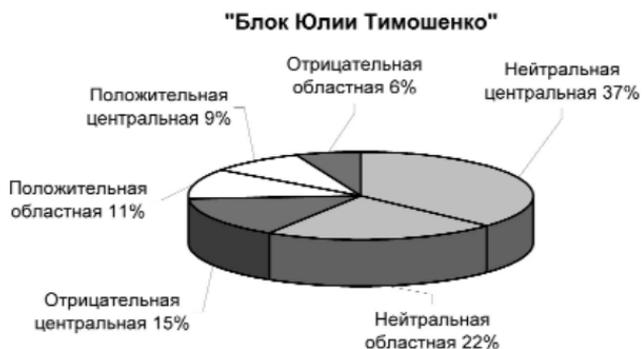


График 1. Распределение публикаций по тональности поданной информации

График 2. Тональность публикаций о лидерах парламентских предвыборных гонок



ша Украина». Значительная часть негативных оценок, особенно в общеукраинских изданиях (24%), очевидно, была вызвана позиционированием блока как провластной силы. 15% позитивно окрашенных публикаций – это фактически свидетельство слабой и невыразительной работы предвыборного штаба «Нашей Украины». Особенно заметны недостатки в работе с прессой в регионах. На страницах областных изданий было размещено только 27 % публикаций про блок (для сравнения – в «Блоке Юлии Тимошенко» доля региональных публикаций – 38%). Больше всего про «Нашу Украину» писали в газетах Львовской области. Хотя в публикациях прессы данного региона преобладала нейтральная тональность по отношению к блоку, однако именно они дали и наибольший процент негативной региональной информации.

Распределение тональности публикаций в газетах станет понятнее, если проанализировать тематическую структуру информации (график 3). Наиболее наполненной оказалась рубрика «Аналитическая и прогнозная информация» (23,6%). Журналисты активно анализировали ход предвыборной кампании, печатали статистические данные и рейтинги политических партий (блоков), прогнозы отдельных политологов, социологов или влиятельных политиков. Наполненными также оказались разделы, содержащие информацию о предвыборной и внутривнутрипартийной деятельности политических



График 3. Тематическая структура информации

сил-участников предвыборных гонок, сообщения о съездах, пленумах и т. д. Таким образом, все три наибольшие тематические блоки информации, касающиеся предвыборной борьбы, уже по определению представляли объективный характер подачи материалов.

Результативным было исследование динамики упоминания отдельных политических сил в газетах Украины. Партия Регионов Украины на протяжении почти всей кампании имела информационное преимущество над политическими соперниками, а вторым по количеству внимания газет был «Народный Союз «Наша Украина». Но в марте «Блок Юлии Тимошенко» осуществил настоящий информационный прорыв: в первую неделю месяца обойдя по количеству упоминаний «Нашу Украину», а в середине месяца – Партию Регионов.

Интересным оказалось распределение публикаций, посвященных отдельным политическим силам (график 4). Ранжирование участников гонок по количеству посвященных им публикаций почти полностью совпадает с реальными результатами выборов в парламент: Партия Регионов Украины (26,5%); «Блок Юлии Тимошенко» (25,3%); «Народный Союз «Наша Украина» (19,6%); Социалистическая партия (5,8%), Коммунистическая партия Украины (3,2%).

Заслуживает внимания информация об активности в освещении предвыборной борьбы отдельных газетных изданий (Таблица 1).

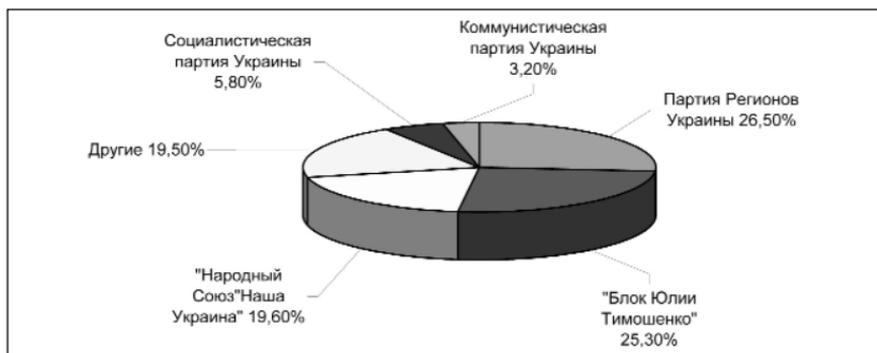


График 4. Распределение публикаций, посвященных отдельным политическим силам

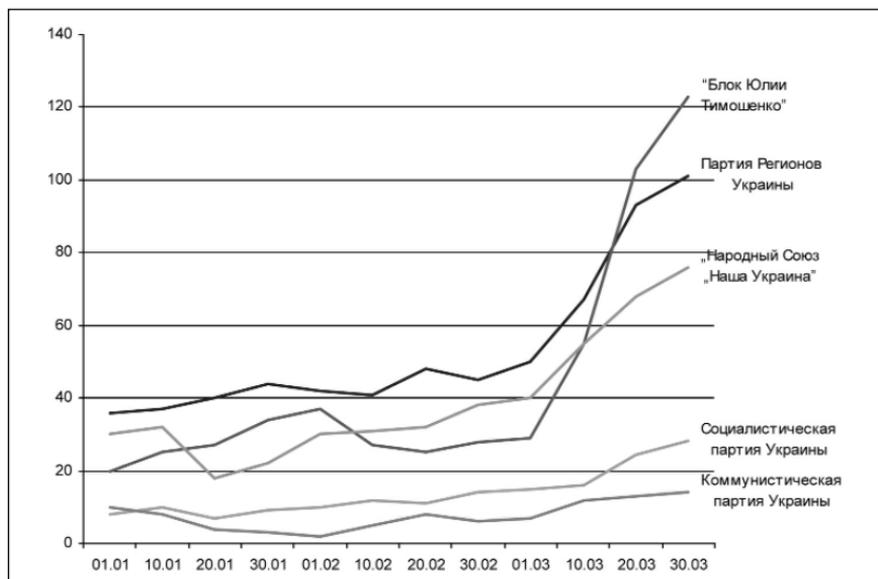


График 5. Динамика упоминания основных политических сил на страницах газет

Первые строчки в рейтинге наиболее информативных в политическом плане газет заняли ежедневные издания, причем первые три места – деловые, бизнесовые газеты. Это свидетельствует и о тесной взаимосвязи политики и экономики в наше время, и о прямой заинтересованности деловых кругов в результатах выборов. Среди еженедельников наиболее информативной была газета «2000». Большое внимание этому событию уделяли еженедельники «Українська газета», «Правда Украины», «Свобода», «Зеркало недели» и «Кієвській телеграфЪ». В первую тридчатку наиболее информативных газет из региональной прессы попали только четыре львовских и одно крымское издание. Особенно необходимо выделить традиционно активные в политической сфере «Львівську газету» и «Високий Замок».

Приведенные данные свидетельствуют о том, что предложенная технология контент-мониторинга эффективна в исследовании реакции прессы на резонансные социально-политические события. Среди других преимуществ данной системы – отслеживание эмо-

Таблица 1. Ранжирование просмотренных газет по количеству отобранных публикаций о политических партиях (первые 30 газет)

1.	Дело	169
2.	Коммерсантъ	154
3.	Экономические известия	121
4.	День	112
5.	Україна молода	109
6.	Киевские Ведомости	84
7.	Газета по-киевски	73
8.	Столичные новости	72
9.	2000	63
10.	Вечерние Вести	62
11.	Демократична Україна	62
12.	Сегодня	58
13.	Львівська газета	53
14.	Високий Замок	51
15.	Голос України	48
16.	Комсомольская правда в Украине	47
17.	Урядовий кур'єр	47
18.	Известия-Украина	44
19.	Українська газета	44
20.	Факты и комментарии	41
21.	Поступ	40
22.	Робітничка газета	39
23.	Экспрес	37
24.	Правда Украины	34
25.	Столичка	33
26.	Свобода	31
27.	Дзеркало тижня	30
28.	Кіевскій телеграфЪ	29
29.	Аргументы и факты в Украине	28
30.	Крымская правда	28

циональной окраски, тональности публикаций; отношения СМИ к тому или иному объекту исследования; взглядов объектов исследования на ту или иную общественную проблему; тенденциозности (заангажированности) или нейтральности (объективности) конкретных изданий прессы; наиболее «раскрученных» в СМИ тем и персоналий; интерпретации значимых событий и возможность сохранения контекста упоминания того или иного объекта исследования.

Литература

1. Федорчук А. Г., Танатар Н. В. Теоретико-методичні засади аналізу інформаційного потоку соціально-політичного спрямування // Бібліотекознавство. Документознавство. Інформологія. – 2004. – № 2. – С. 33-38.

2. Федорчук А. Г. Контент-мониторинг информационных потоков // Библиотеки национальных академий наук: проблемы функционирования, тенденции развития: Науч.-практ. и теорет. сб. – 2005.– Вип. 3.– С. 141–150.

3. Сорока М., Танатар Н. Використання методу контент-аналізу при створенні автоматизованих систем // Бібліотека. Наука. Культура. Інформація: Наук. праці НБУВ. – 1998.– Вип. 1.– С. 318–323.

4. Танатар Н., Федорчук А. Усе – про політичну еліту // Вісник НАН України, 1997, № 1–2 – С. 91–92.

5. Федорчук А., Танатар Н. Президентські вибори 1999 р.: контент-аналіз матеріалів преси // Українська періодика: Історія і сучасність: Доп. та повід. шостої Всеукр. наук.-теорет. конф. 11–13 травня 2000 р. – Львів, 2000. – С. 351-354.

6. Танатар Н. В. Досвід інформаційно-аналітичного забезпечення вищих органів державної влади // Документознавство. Бібліотекознавство. Інформаційна діяльність: проблеми науки, освіти, практики: Зб. матеріалів II Міжнар. наук.-практ. конф. Київ, 17–18 травня 2005 р. – К., 2005. Ч. 1. – С. 56-58.

7. Федорчук А. Г., Танатар Н. В., Третьяк Т. І., Шишкіна В. Г. Контент-моніторинг передвиборчої кампанії на базі газетної періодики // Документознавство. Бібліотекознавство. Інформаційна діяльність: проблеми науки, освіти, практики: Зб. матеріалів III Міжнар. наук.-практ. конф. Київ, 16–18 травня 2006 р. – К., 2006. – С. 98–101.

8. Танатар Н. В. Информационно-аналитическое обслуживание высших органов государственной власти с использованием автоматизированных систем ключевых фрагментов публикаций // Библиотеки национальных академий наук: проблемы функционирования, тенденции развития: Науч.-практ. и теорет. сб. – 2003. – Вип. 2. – С. 91–100.

9. Федорчук А. Г. Автоматизована технологія формування банку ключових фрагментів публікацій // Наук. праці Національної бібліотеки України імені В. І. Вернадського. – К.: НБУВ, 2001. – Вип. 6. – С. 214–220.