
**Танатар Н. В.
Федорчук А. Г.**

Интеллектуальные поисково-аналитические системы мониторинга СМИ

В публикации проанализированы проблемы современных поисковых систем, сделан вывод о необходимости перехода от поиска по ключевым словам к контент-мониторингу текстов публикаций. Приведен краткий обзор систем мониторинга и анализа содержания средств массовой информации (СМИ), представленных в русскоязычном секторе Интернета (Россия и Украина). Сделаны выводы об основных принципах и тенденциях мониторинга содержания информационных потоков.

Сегодня государственным и крупным коммерческим структурам становится все труднее следить за динамично меняющимся информационным полем, которое содержит массу оперативных и обзорных материалов. Кроме того, регулярное ознакомление с публикациями СМИ для любой серьезной деятельности необходимо, но не всегда достаточно. Большие массивы информации должны подвергаться качественному анализу, поскольку на основе информации из открытых источников можно не только анализировать состояние дел, но и строить прогнозы развития ситуации, что жизненно важно для принятия верных решений.

По утверждению специалистов, около 80% информации, требуемой для поддержки процесса стратегического управления, может добываться из открытых источников, таких, как Интернет, традиционные СМИ, информационно-рекламные материалы компаний и т. д. По оценкам зарубежной прессы, 9 из 10 крупных американских компаний тратят ежегодно в среднем 1 млн долл. на мониторинг. Вложенные средства окупаются за счет правильных решений, принятых на основе отобранной и проанализированной

информации. Сегодня контроль внешней информационной среды становится жизненно необходимым. Поэтому высшие менеджеры крупных и быстро развивающихся компаний создают специальные аналитические службы, помогающие собирать и фильтровать информацию из десятков и сотен источников [1, 2]. Можно сказать, что постановка анализа информации на регулярную, промышленную основу – это признак определенного уровня зрелости фирмы и ее руководства. Для того, чтобы обеспечить регулярный и полный анализ информации, необходимы, как минимум, три компонента:

- создание аналитического подразделения и разработка методики его работы;
- оснащение этого подразделения специальными поисковыми и аналитическими инструментами;
- подключение к существующим информационным источникам, сбор и анализ всей поступающей от них информации.

В условиях небольшого количества информации аналитики могли бы работать вручную или с помощью типовых офисных программ. Но практика показывает, что уже через несколько дней работы такая служба начинает понимать, что существующий в сегодняшнем мире объем доступной информации невозможно обрабатывать по старинке. Очевидно, что такие службы необходимо вооружать современными поисковыми и аналитическими инструментами.

Проблемы поиска

Современные поисковые системы в Интернете сегодня столкнулись с целым рядом проблем [3]. Среди наиболее часто упоминаемых можно назвать разрастание общего объема материалов, расположенных в Интернете, переход к новым форматам, увеличение количества нетекстовых материалов и т. д. Конечно, каждая из этих проблем очень серьезна, и в этих направлениях ведутся достаточно продуктивные работы с привлечением лучших мировых специалистов. Однако основной проблемой, с которой встретились современные поисковые системы (ПС), присутствующие на современном рынке средств текстового поиска, и с которой

сталкивается практически каждый пользователь сети – это проблема избытка информации.

Проявилась эта проблема при наращивании объема индексируемых текстовых документов. Рост объема текстов, проиндексированных поисковой системой, привел к тому, что практически любой запрос выдает выборку из многих тысяч, а то и десятков тысяч подходящих документов. Естественно, это превышает максимум, который способен обработать один человек за разумное время. Такой максимум для профессионального аналитика находится в пределах нескольких сотен документов, для непрофессионала, естественно, границы существенно ниже, как минимум, на порядок. Проблема эта не является новой, для ее решения и создавалась технология ПС. Но на сегодняшнем этапе эта технология уже не может адекватно справиться с соответствующими задачами.

Конечно, эта проблема коснулась не только ПС, занимающихся индексированием материалов в Интернете, но и ПС, работающих с любыми другими текстовыми базами, например, электронными библиотеками. К таким системам относятся известные библиотеки НЭБ-НСН, Интегрум-Техно в России, Лексис-Нексис, Рейтер на Западе. Пользователей у таких ПС, естественно, существенно меньше, чем у ПС в Интернете, хотя объем документов, доступных в ПС Лексис-Нексис, не только существенно выше объема Интернет, но даже скорость поступления новых документов в эту библиотеку до сих пор превосходит скорость поступления в Интернет. Однако, даже гораздо более строгая организация данных в таких библиотеках, почти полное единство форматов внутри одной библиотеки, не являются панацеей от общей проблемы современных ПС.

Сегодня рынок предлагает довольно большое количество поисковых систем. Но работа многих из них построена достаточно примитивно. Система ищет слово или словосочетание. В результате человек получает сотни, а чаще тысячи ненужных документов, содержащих это слово. При этом документ, относящийся к теме, но не содержащий в явном виде этого термина, будет пропу-

шен. Зачастую существует и обратная задача – узнать, о чем пишется в наборе документов или статей, выявить важные темы. С этими задачами или проблемами аналитические центры сталкиваются каждый день. И центров, и информации становится все больше. Недавно возникнув, спрос на новые интеллектуальные поисково-аналитические системы растет довольно быстро.

Инструменты мониторинга СМИ

Под «удобными инструментами для работы с информацией» понимаются компьютерные программы, помогающие собирать и сортировать материалы СМИ. Таких программ на рынке около сотни, но все они занимают, как правило, организацией хранения средних или больших объемов информации, обладают простыми поисковыми возможностями и/или тематическим рубрикатором, не предлагая какого-либо механизма качественного анализа. И это не случайно, ибо качественный анализ текста предполагает оценку таких нюансов, как эмоции, угрозы, характер отношений между объектами. В этих сферах человеческий мозг, способный накапливать опыт и обладающий интуицией, предпочтительнее. Хотя скорость оценки при этом и невелика, мы выигрываем в качестве.

Таким образом, при решении интеллектуальных задач, компьютерная программа призвана не заменить человека, а максимально облегчить его труд: во-первых – обеспечить его ограниченной выборкой документов, отсеяв лишнее, во-вторых – предоставить условия для проведения оценок.

В русскоязычном секторе Интернета можно обнаружить не меньше десятка баз данных, ориентированных на сбор и хранение материалов СМИ. Наиболее известные из них: «Артефакт» компании «Интегрум-Техно»; система «Парк», созданная «Парк.Ру»; электронный киоск Russian Story; Национальная электронная библиотека от НСН; «Медialogия»; УИС РОССИЯ; каталог СМИ от «СМИ.Ру». Работа со всеми перечисленными источниками, кроме «СМИ.РУ», платная, и все их владельцы предлагают не только возможность ознакомиться с материалами СМИ, но и некоторые мониторинговые услуги.

Аналитические программы для работы с текстами

Аналитические системы различаются по виду обрабатываемых данных – полнотекстовых или фактографических. Методы обработки фактографических данных известны достаточно давно. Среди них в последнее время особой популярностью пользуются OLAP-анализ и Data Mining (выявление последовательностей, ассоциаций, деревья решений и т. д.). Эти методы в той или иной мере сейчас поддерживаются всеми современными системами. Частично они реализованы в MS OLAP Services и в продуктах компании Business Objects. Наиболее полно – в системе PolyAnalyst компании Megaruter.

Методы анализа текстов распространены гораздо меньше. Это в основном тематическое рубрицирование входящего потока документов и подсчет статистики встречаемых слов и словосочетаний. Наиболее известными производителями этих систем и отдельных компонентов являются канадская фирма Hummingbird (продукт Hummingbird Knowledge Management), а также российские компании Media Lingva («Классификатор»), Megaruter (TextAnalyst) и «Гарант-Парк-Интернет» (ее продукт реализован на основе технологий американской фирмы InterMedia).

Для успешного применения такого рода методов входящий поток всегда подвергается предварительной обработке, включающей просмотр оператором, контроль орфографии, нормализацию регистра и т.д. Для последующего контекстного поиска или сложных процедур анализа, как правило, проводится полнотекстовое индексирование содержимого документов.

Краткий обзор систем мониторинга и анализа содержания СМИ

«Астарта» (www.cognitive.ru/products/astarta.htm)

Компания Cognitive Technologies предлагает информационно-аналитическую систему (ИАС) «Астарта» [4]. Это экспертный рубрикатор, предназначенный для сбора неструктурированной информации, хранения и семантического анализа текстовых материалов. Под анализом в данном случае понимается автоматическое рубрицирование и группировка, а также интеллектуальная выборка

информации по заданной теме. Главной отличительной особенностью данной ИАС является автоматическое обучение рубрикатора, т. е. автоматическое построение списка терминов и понятий, определяющих принадлежность документа к рубрике. Технологической основой для «Астарты» является комплекс средств для создания электронных архивов «Евфрат». В «Астарте» есть подсистема, умеющая обрабатывать материалы из СМИ в потоковом режиме.

«Галактика-Zoom» (zoom.galaktika.ru)

Программный комплекс «Галактика-Zoom», предназначенный для аналитической обработки текстовых неструктурированных документов, предлагает корпорация «Галактика». Программа может собирать информацию на Интернет-сайтах или брать документы из подключаемых баз данных. Авторы обещают пользователю информационную поддержку принимаемых управленческих решений благодаря быстрому поиску и контент-анализу (КА) отобранной информации. Список тем формируется не случайным образом, а с учетом частоты употребления в текстах с более общими терминами. Рубрикатор, таким образом, создает структуру, включающую рубрики и подрубрики. Такой подход позволяет составлять информационный портрет представляющего интерес объекта [5–6].

«Медialogия» (www.medialogia.ru)

Информационно-аналитическая система «Медialogия» одноименной компании разработана группой российских ученых, аналитиков, а также специалистов в области информационных технологий [7]. Прототипом «Медialogии» является система IBS-Media, более известная как модуль ситуационных центров, создаваемых и продвигаемых отделением систем имитационного моделирования компании IBS. Разработчики исповедуют смешанный подход к оценке статей и других объектов. Это значит, что часть наименее интеллектуальной черновой работы выполняет программа. Более тонкие оценки, отражающие характер упоминания объектов в статье и то, в каких отношениях они между собой находятся, берет на себя человек.

На текущий момент система осуществляет мониторинг более чем 24 000 объектов, фиксируя статистическую и аналитическую информацию из тысячи источников (центральная и региональная бумажная пресса, сообщения информационных агентств, материалы теле-, радиопередач, Интернет-источники). На обработке сообщений задействовано несколько сотен квалифицированных операторов, безостановочно просматривающих до десяти тысяч сообщений в сутки. Система позволяет классифицировать публикации по значимости, определять отношение СМИ к объектам, анализировать характеристики PR-кампаний, устанавливать отраженные в СМИ связи между объектами и т. д.

Технологии Hummingbird (www.hbsltd.biz/products_km.asp)

В системе, которая создавалась по заказу российской компании HBS для анализа региональных СМИ, была реализована технология обработки электронных текстов с использованием поискового сервера Hummingbird SearchServer (ранее Fulcrum SearchServer) компании Hummingbird. Поступающая информация автоматически рубрицируется, а затем подвергается OLAP-анализу. В пакетах Hummingbird SearchServer и Hummingbird KnowledgeServer есть средства, с помощью которых пользователи могут сами создавать новые или настраивать имеющиеся деревья рубрик.

Для повышения качества авторубрицирования в системе реализованы развитые возможности предобработки поступающих документов. В частности, проводится контроль орфографии, используется фильтр стоп-слов, осуществляется нормализация регистра и др. При выполнении поисковых запросов можно подключать тезаурус (словарь синонимов). Выборка документов производится с помощью контекстного поиска или OLAP-анализа.

TextAnalyst (www.analyst.ru)

Программа TextAnalyst от НПИЦ «МикроСистемы» является инструментом для анализа содержания текстов, смыслового поиска информации и формирования электронных архивов [8]. Она также способна строить семантические деревья, но не по объектам, а по отдельным статьям, в результате чего создается смысло-

вой портрет каждого текста в терминах основных понятий и их смысловых связей в виде иерархии тем и подтем конкретной статьи. В TextAnalyst есть также модуль, автоматически генерирующий реферат текстового документа в терминах наиболее информативных фраз. Программа не предназначена для потоковой обработки материалов СМИ, но может брать с диска файлы в формате txt и rtf и после анализа текста сохранять результаты в отдельном файле.

Semantic Explorer (www.neurok.ru/products)

Semantic Explorer – это клиент-серверный программный комплекс компании «НейрОК». Клиентский интерфейс Semantic Explorer ориентирован на работу с семантикой документов и поиск по смысловым и тематическим ассоциациям. В отличие от TextAnalyst семантическая карта строится не по отдельным документам, а по их базе. На такой карте (карте Кохонена) каждый документ имеет свое уникальное положение. Причем близкие по смыслу документы располагаются рядом.

Большое внимание компания уделяет технологиям Интернет-агентов, которые призваны производить в Интернете «осмысленный» поиск. Если связать таких агентов в единую сеть, то можно создать распределенное индексное поле обработанной ими совместно информации, существенно облегчающее поиск.

TopSOM (research.metric.ru)

Компания «Гарант-Парк-Интернет» уже не первый год ведет исследования в области интеллектуального поиска и тематического анализа текстовых документов. Исследователи предлагают линейку продуктов, одним из которых является TopSOM, базирующийся на технологии нейронной сети Кохонена.

Все множество документов разбивается на небольшое количество классов схожих по содержанию документов. Эти классы отображаются на плоскость таким образом, что близкие классы соответствуют близким областям плоскости. Задача нелинейного отображения многомерного семантического пространства в пространство малой размерности решается нейросетевым алгоритмом. Такое отображение позволяет наглядно представить тематический

состав большой (десятки тысяч текстов) коллекции документов в целом и помочь пользователю сориентироваться в океане информации.

Convera RetrievalWare (www.convera.com/press/webinar/comm.html)

Информационно-поисковая система Convera RetrievalWare – продукт американской компании Convera Technologies. Однако российская компания «Весть-МетаТехнология» не только локализовала интерфейс и документацию, но и адаптировала поисковый механизм системы к работе с русскоязычными документами. Для этого была составлена семантическая сеть словаря русского языка, которая содержит около 100 тыс. слов и идиоматических выражений и более 350 тыс. связей между ними, подключена библиотека морфологического анализа, внесены изменения в ядро оригинальной системы. А компания «Одеон», авторизованная для работы на рынках СНГ, существенно переработала механизм поиска и семантико-морфологического разбора. В частности, новый механизм позволяет при поиске и анализе текста создавать семантическую сеть не только из синонимов, но и антонимов, морфем, вариантов произношения, сленга, а также выделять сущности и передавать результаты в СУБД. Новый словарь содержит более 1,2 млн взаимосвязанных словоформ.

IntellSoft Vision (www.intellsoft.ru/vision)

Компания IntellSoft предлагает программу IntellSoft Vision, помогающую в выборе объектов, представляющих интерес, и в реализации стратегий. Программа обеспечивает:

- ситуационно-аналитический мониторинг;
- интеллектуальный анализ данных;
- оценку тенденций материалов СМИ;
- интеграцию с внешними информационными ресурсами.

Проект ВААЛ (www.vaal.ru)

В рамках проекта ВААЛ созданы две системы контент-анализа: ВААЛ-2000 для психолингвистического исследования русскоязычных текстов и Vaal Toolbox для аналогичных исследований англоязычных источников информации. ВААЛ-2000 позволяет восполь-

зоваться установленными аналитическими моделями или создать собственные модели. В его распоряжении, например, методика психоанализа, критерии которого оценивают наличие в тексте слов, которые относятся к сексуальной символике (по З. Фрейду), архетипам (по К. Юнгу) и выражению агрессивности. А эмоционально-лексический анализ позволяет выявить эмоциональную насыщенность прямой речи по 15 наиболее значимым для российской культуры критериям.

Украинские системы мониторинга СМИ

InfoStream

В Информационном центре «ЭЛВИСТИ» создана самая известная в Украине система контент-мониторинга InfoStream, с помощью которой охватываются новости из более 3 тыс. отечественных и зарубежных Web-сайтов, осуществляется их обработка и обобщение [9]. Система InfoStream предназначена для нахождения в сети Интернет оперативной информации по интересующим пользователя тематикам, доставки результатов поиска, предоставления пользователю одновременного доступа в режиме поиска к информации с многих Web-сайтов, смысловой обработки и, вследствие этого, минимизации усилий пользователя на отсеивание дублирующейся информации, шума. В настоящее время система InfoStream сканирует порядка 50000 уникальных сообщений в сутки.

«Медіа-простір»

Корпорация «Медіа-простір» осуществляет анализ информации в центральной и региональной печати, интернет-изданиях, а также в передачах новостей и аналитических передачах украинского и российского радио и телевидения (около 550 источников).

Центральный офис, расположенный в Киеве, осуществляет отбор материалов, появляющихся в центральных СМИ, а 25 региональных бюро во всех областях Украины отслеживают информационные сообщения в местной прессе и ежедневно пересылают информацию в центральный офис. Основная услуга агентства – предоставление медиа-обзоров по любой теме и за любой интересующий клиента период времени.

Систематизация информационных сообщений происходит по четырем объектам привязки:

- предметные сферы;
- политические субъекты;
- личности;
- территории.

Информационно-технологическое обеспечение позволяет:

- в процессе привязки предметных материалов производить погружение до третьего уровня;
- осуществлять перекрестную привязку.

Информационно-технологическое обеспечение позволяет производить мониторинг СМИ по двум направлениям:

- количественный анализ информационного пространства в виде статистических выкладок как в абсолютном, так и сравнительном разрезе;
- вычленение и анализ тенденций социально-экономического развития на основе анализа гиперинформационных массивов.

Системы контент-мониторинга по ключевым фрагментам публикаций

Отдел организации и использования документального фонда Фонда Президентов Украины Национальной библиотеки Украины имени В. И. Вернадского, начиная с 1993 года по сегодняшний день непрерывно осуществляет контент-мониторинг газетной периодики [10–12].

У автоматизированной технологии контент-мониторинга существуют несколько важных особенностей. Прежде всего – за единицу формирования текстового информационного массива используется ключевой фрагмент публикации. Формирование банка ключевых фрагментов публикаций является объединением двух взаимосвязанных автоматизированных процессов: аналитико-синтетической переработки и многоуровневой процедуры контент-анализа (КА) текстов публикаций. Индексация ключевых фрагментов публикаций происходит при помощи многофасетной классификации.

Уникальность предложенной технологии состоит в объедине-

нии содержательных и количественных методов КА. Последовательность этапов содержательного анализа проблемы, которая исследуется конкретной информационной системой, условно можно поделить на содержательный (качественный) анализ совокупности публикаций и формализованный (количественный) анализ информационных массивов: индексного, библиографического и массива текстов ключевых фрагментов публикаций.

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 соответствуют 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.

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.

Тенденции и перспективы

Многие из перечисленных продуктов снабжены собственными или встроенными авторубриками и аннотаторами, а это означает, что указанные инструменты постепенно становятся стандартными для информационно-аналитических систем. Настоятельной необходимостью для такого класса программ является умение работать с тезаурусом (словарем синонимов) и учитывать морфологию языка: без этих функций при поиске легко пропустить нужные документы.

Желая оставаться в русле основных тенденций развития ИТ-индустрии, многие разработчики переводят свои клиент-серверные приложения на Интернет-платформы. Все перечисленные в обзоре продукты, кроме «Астарты» и TextAnalyst, имеют Web-оболочку и используют для передачи данных TCP/IP. Очень желательным для заказчика свойством сложных систем является их модульность, позволяющая легко строить необходимые (нередко более дешевые) конфигурации.

Главное, что объединяет все системы контент-мониторинга, это сочетание качественного и количественного анализа текстов. Доверяя в рамках проведения контент-анализа компьютеру количественные оценки, аналитические системы предоставляют человеку возможность качественной оценки исследуемых текстов, помогающей фиксировать и структурировать новый слой знаний для последующего его анализа. Таким образом, данные системы являются новым шагом в плане работы с информацией, и такой недостаток, как избыток информации, превращается в достоинство. Ведь чем большее количество разнообразных источников будет проанализировано, тем полнее и качественнее будет информаци-

онный портрет исследуемого объекта. Контент-мониторинг, таким образом, дает возможность объективации субъективных мнений, выраженных в большом количестве текстовых материалов СМИ.

Впрочем, по всей видимости, настоящий прорыв в обработке материалов СМИ будет достигнут лишь тогда, когда сами авторы станут сопровождать текст некоторой информационной структурой, описывающей смысл статьи и «знания», в ней изложенные, другими словами, когда возобладает подход, основанный на разделении данных, характеризующих содержание, представление и смысловое значение. Адептом этого подхода является один из основателей современного Интернета Тим Бернерс-Ли. Он предлагает объединить документы подобного рода в единую сеть знаний, которая будет называться Semantic Web.

Литература

1. *Опарин А.* Системы мониторинга и анализа СМИ [Текст] / А. Опарин // РСWEEK. – 2003. – № 47. – С. 19–24.

2. *Липинский Ю. В.* Средства информационного поиска и навигации в больших массивах неструктурированной информации [Электронный ресурс] / Ю. В. Липинский. – Электрон. дан. – Режим доступа: <http://www.fep.ru/text/dataarrays04.html>. – Назва з екрану.

3. *Антонов А.* Современные проблемы поисковых систем и некоторые пути их преодоления [Текст] / А. Антонов, В. Мешков // Аналитика – Капитал / ВИНТИ ; ИПКИР Минпромнауки России ; под науч. ред. А. В. Руцкого. – М. : Академия аналитики и информатики, 2000. – Т. 3 : Риски. Ситуационные центры и системы. Технологии выявления неявных проблем. – С. 234–241.

4. *Инструменты «Р»* : ИАС «Астартга» [Электронный ресурс]. – Электрон. дан. – Режим доступа: <http://it2b-pro.ru/it2b3.view4.page1.html>. – Назва з екрану.

5. *Антонов А. В.* Методы классификации и технология Галактика-Zoom [Текст] / А. В. Антонов // Междунар. форум по информ. – 2003. – Т. 28, № 4. – С.27–32.

6. *Самойлов Ю.* Система мониторинга и анализа СМИ «Медialogия» [Электронный ресурс] / Ю. Самойлов. – Электрон. дан. – Режим доступа: <http://integration.ibs.ru/content/rus/rubr65/rubr-652.asp>. – Назва з екрану.

7. *Харламов А.* TextAnalyst ver. 2.0 – Программа для персонального

анализа текстов [Електронний ресурс] / А. Харламов. – Електрон. дан. – Режим доступу: <http://www.offext.ru/Library/Data/Datakeeping/51.aspx>

8. *Моніторинг* діяльності органів виконавчої влади із застосуванням комп'ютерної системи контент-аналізу електронних ЗМІ [Текст] / Г. Леліков, В. Сороко, О. Григор'єв, Д. Ланде // Вісн. держ. служби України. – 2002. – № 2. – С.21–38.

9. *Федорчук А. Г.* Теоретико-методичні засади аналізу інформаційного потоку соціально-політичного спрямування [Текст] / А. Г. Федорчук, Н. В. Танатар // Бібліотекознавство, документознавство, інформологія. – 2004. – № 2. – С. 33–38.

10. *Танатар Н. В.* Інформаційно-аналітичні системи за ключовими фрагментами публікацій як основа для створення електронної бібліотеки соціально-політичного спрямування [Текст] / Н. В. Танатар // Наук. пр. Нац. б-ки України ім. В. І. Вернадського. – К. : НБУВ, 2001. – Вип. 6. – С. 190–200.

11. *Федорчук А. Г.* Контент-моніторинг інформаційних потоків [Текст] / А. Г. Федорчук // Библиотеки национальных академий наук: проблемы функционирования, тенденции развития : науч.-практ. и теорет. сб. – 2005. – Вып. 3. – С. 141–150.