

АНАЛИЗ ТРЕНДОВ ИНФОРМАЦИОННЫХ ПОТОКОВ ПРИ ВЫЯВЛЕНИИ ИНФОРМАЦИОННЫХ ОПЕРАЦИЙ

Веб-пространство представляет собой динамическую систему из связанных по смыслу элементов (документов), образующих в динамике своей эволюции информационные потоки [1, 2]. Динамика подобных систем описывается временными рядами. Как известно «тренд» (от англ. Trend) – это общая тенденция изменений исследуемого временного ряда.

Целью настоящей работы является определение типичных трендов динамики публикаций в информационном пространстве, соответствующих информационным операциям, определяемым как «акции, направленные на воздействие на информацию и информационные системы противника и защиту собственной информации и информационных систем» [3, 4].

Основным объектом моделирования информационных потоков [5] являются тематические потоки новостей, последовательности документов, соответствующих определенной тематике. Этим потокам ставятся в соответствие временные ряды, для решения задач анализа которых все чаще применяются дисперсионный, фрактальный и вейвлет-анализ [6, 7].

Получить данные динамики тематических информационных потоков можно, например, ежедневно посещая сайты интеграторов новостей (news.yandex.ru, webground.su, uaport.net).

В настоящее время существует несколько открытых информационных сервисов, в рамках которых можно наблюдать временную динамику объемов публикаций по тематикам, определяемым запросами. Так Google Books Ngram Viewer (<http://ngrams.googlelabs.com/>), предоставляет визуализацию динамики количества книг, в которых упоминаются слова из запроса.

Сервис «Яндекс пульс блогосферы» (<http://blogs.yandex.ru/pulse/>) также позволяет отображать динамику публикаций в блогах, содержащих заданные пользователем ключевые слова.

На сайте Национального корпуса русского языка (НКРЯ) в бета-режиме запущен сервис *N*-грамм (<http://www.ruscorpora.ru/ngram.html>), близкий по функциональности сервису Google Books Ngram Viewer.

Многие современные информационно-аналитические системы содержат в своем составе средства отображения статистики вхождения в базы данных понятий, соответствующих пользовательским запросам. В частности, авторами использовалась подсистема статистики в рамках системы контент-мониторинга веб-пространства InfoStream [8], реализующая данную функциональность.

Приведенные в [4] тренды сообщений, соответствующих этапам информационной операции приведены на рис. 1. При этом аналитикам можно следует ориентироваться на такие модели, например, если мониторинг позволяет

определить фазы: «фон» – «затишье» – «артподготовка» – «затишье» – «атака», то уже по первым трем компонентам можно с большой вероятностью предсказать будущие события.

Следует отметить, что подобная динамика количества тематических сообщений при проведении информационных операций хорошо описывается известным уравнением:

$$y = A + Bx \sin(x),$$

где x – время, A и B – константы, определяемые эмпирически.

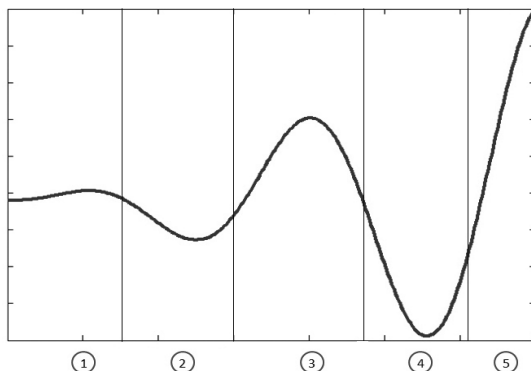


Рис. 1 – Динамика количества тематических сообщений во время проведения информационной операции: 1 – фон; 2 – затишье; 3 – «артподготовка»; 4 – затишье; 5 – атака/триггер роста

Как известно, в настоящее время инновационная деятельность также косвенно измеряется количеством публикаций, относящимся к инновациям, существует несколько моделей инновационных процессов, среди которых можно выделить модель диффузии инноваций. Вместе с тем, внедрение инноваций также можно считать информационными операциями. Поэтому обратимся к результатам соответствующих исследований. На рис. 2 приведена обоснованная в [9] диаграмма количества публикаций, соответствующая тренду инновационной деятельности.

Объединяя графики, соответствующие началу информационной операции и тренду инновационной деятельности, можно получить полный график, соответствующий отображению информационных операций в информационном пространстве (рис. 3).

Предложенные модели полностью соответствуют реальным данным, которые экстрагируются системами контент-мониторинга [10]. Поэтому приведенные зависимости могут быть использованы как шаблоны для выявления информационных операций, как путем анализа ретроспективного фонда сетевых публикаций, так и для оперативного мониторинга появления некоторых их признаков в реальном времени.

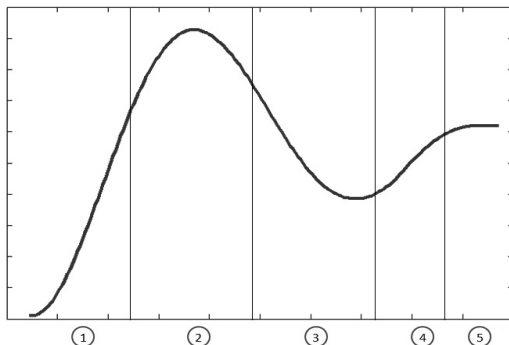


Рис. 2 – Диаграмма количества публикаций, соответствующих тренду инновационной деятельности: 1 – атака/триггер роста; 2 – пик завышенных ожиданий; 3 – утрата иллюзий; 4 – общественное осознание; 5 – продуктивность/фон

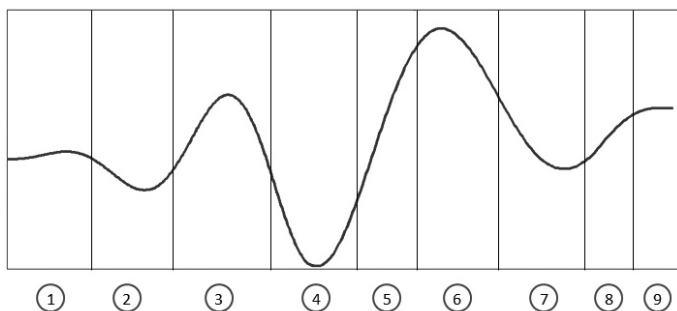


Рис. 3 – Обобщенная диаграмма количества публикаций, соответствующая всем этапам жизненного цикла информационных операций: 1 – фон; 2 – затишье; 3 – «артподготовка»; 4 – затишье; 5 – атака/триггер роста; 6 – пик завышенных ожиданий; 7 – утрата иллюзий; 8 – общественное осознание; 9 – продуктивность/фон

Как известно, для выявления информационных операций следует внимательно следить за динамикой публикаций по целевой теме и, если есть возможность, пользоваться доступными аналитическими средствами, средствами цифровой обработки данных и распознавания образов, например, вейвлет-анализом или полиномами Кунченко [11].

Приведенные модели и методы пригодны для описания общих тенденций динамики информационных процессов, однако, проблема прогнозирования остается открытой. По-видимому, реалистичные модели могут быть получены с

учетом дополнительного набора факторов, большинство которых не воспроизводятся во времени. Вместе с тем, структура правил, лежащих в основе функционирования большинства из доступных моделей, позволяет вносить соответствующие коррективы, например, искусственно моделировать случайные отклонения. Отметим, что воспроизведение результатов во времени является серьезной проблемой при моделировании информационных процессов, составляет основу научной методологии. В настоящее время только ретроспективный анализ уже реализованных информационных операций остается относительно надежным способом их верификации.

Список литературы

1. Kleinberg J. Temporal dynamics of on-line information streams // *Data Stream Management: Processing High-Speed Data Streams*. – Springer, 2006.
2. Lande D., Braichevski S, Busch D. Informationsfluesse im Internet // *IWP - Information Wissenschaft & Praxis*, 59(2007), Heft 5. – S. 277-284.
3. Information operations roadmap. – DoD US. – Washington, D.C.: GPO, 2003. – 78 p.
4. Горбулін В.П., Додонов О.Г., Ланде Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія. - К.: Інтертехнологія, 2009. - 164 с.
5. Ландэ Д.В., Снарский А.А., Брайчевский С.М., Дармохвал А.Т. Моделирование динамики новостных текстовых потоков // *Интернет-математика 2007: Сборник работ участников конкурса*. – Екатеринбург: Изд-во Урал. ун-та, 2007. – С. 98-107.
6. Астафьева Н.М. Вейвлет-анализ: основы теории и примеры применения // *Успехи физических наук*, 1996. – 166. – № 11. – P. 1145-1170 (http://www.isuct.ru/~artcol/articles/Uspekhi_Fiz_Nauk/wavelet-analys.pdf).
7. Lande D.V., Snarskii A.A. Diagram of measurement series elements deviation from local linear approximations // *Preprint Arxiv: 0903.3328*, 2009.
8. Григорьев А.Н., Ландэ Д.В. и др. Мониторинг новостей из Интернет: технология, система, сервис: научно-методическое пособие. – К.: ООО «Старт-98», 2007. – 40 с.
9. Хорошевский В.Ф. Семантические технологии: ожидания и тренды // *Открытые семантические технологии проектирования интеллектуальных систем – Open Semantic Technologies for Intelligent Systems (OSTIS-2012): материалы II Междунар. научн.-техн. конф. (Минск, 16-18 февраля 2012 г.)*. – Минск : БГУИР, 2012. – С. 143-158.
10. Додонов О.Г., Ланде Д.В., Путятін В.Г. Інформаційні потоки в глобальних комп'ютерних мережах. – К: Наукова думка, 2009. – 295 с.
11. Чертов О.Р. Поліноми Кунченка для розпізнавання образів // *Вісник НТУУ «КПІ» Інформатика, управління та обчислювальна техніка*, 2009. – № 50. – С. 105-110.