

**Динамика информационных потоков при выявлении информационных операций**

*Исследованы линии трендов динамики публикаций в информационном пространстве, соответствующие информационным операциям. Приведенные зависимости могут использоваться как шаблоны для выявления информационных операций с помощью технологий распознавания образов.*

*Информационные операции, контент-мониторинг, тренды, информационное пространство, распознавание образов.*

Информационные операции определяются как «акции, направленные на воздействие на информацию и информационные системы противника и защиту собственной информации и информационных систем» [1, 2]. Проявления информационных операций встречаются во многих сферах – военной, социальной, экономической. Информационные операции в настоящее время непосредственно связаны с воздействием на людей, манипулированием.

Динамика публикации документов в информационном пространстве, в том числе, имеющих непосредственное отношение к информационным операциям, образуют временные ряды, изучению общих тенденций которых (трендов), посвящена данная работа. Целью настоящей работы является определение типичных трендов динамики публикаций в информационном пространстве, соответствующих информационным операциям.

Веб-пространство представляет собой динамическую систему из связанных по смыслу элементов (документов), образующих в динамике своей эволюции информационные потоки [3-7].

Основным объектом моделирования информационных потоков [8] сегодня являются тематические сюжеты новостей, последовательности сообщений, соответствующих определенной тематике. Тематическим сюжетам новостей можно ставить в соответствие временные ряды, для решения задач анализа которых все чаще применяются дисперсионный, фрактальный и вейвлет-анализ [9-11].

Современное информационное пространство представляет собой уникальную возможность получения любой информации по выбранному вопросу, при условии наличия соответствующего инструментария, применение которого позволяет анализировать взаимосвязь возможных событий или событий, которые уже происходят, с информационной активностью определенного круга источников информации. С другой стороны, при ретроспективном анализе любого процесса или явления интерес представляют определенные характеристики их развития, а именно:

- количественная динамика, присущая процессу или явлению, например, количество событий в единицу времени, или количество опубликованных сообщений, имеющих отношение к нему;

- определение критических, пороговых точек, соответствующих количественной динамике явления;

- определение проявлений в критических точках, например, выявление основных сюжетов публикаций в информационном пространстве, соответствующих выбранному процессу или явлению;

- после выявления основных проявлений явления в критических точках они ранжируются, и исследуется динамика развития отдельных определенных проявлений до и после этих критических точек;

- проводится статистический, корреляционный и фрактальный анализ общей динамики и динамики отдельных проявлений, на основе которых осуществляются попытки прогнозирования развития явления.

Многочисленные факты свидетельствуют о том, что динамика тематических информационных потоков определяется комплексом внутренних нелинейных механизмов, которые лишь частично коррелируют с реальностью. Например, сюжет может быть полностью вымышленным, с целью манипуляции общественным мнением, а с другой стороны, он может быть связан со многими событиями. Следует признать, что поведение тематических сюжетов не всегда напрямую связано с динамикой событий, хотя с другой стороны, их появление также может считаться событием.

Получить данные динамики тематических информационных потоков можно, например, ежедневно посещая сайты интеграторов новостей ([news.yandex.ru](http://news.yandex.ru), [webground.su](http://webground.su), [uaport.net](http://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. На рис. 1 приведен пример динамики количества публикаций, в которых встречались слова «Хрущев» и «Брежнев» с 1940 по 1990 год.

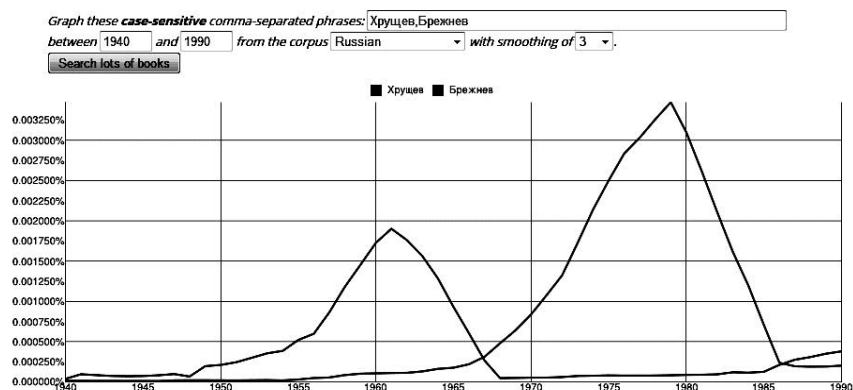


Рис. 1. Динамика публикаций (Google books Ngram Viewer), содержащих заданные слова

Сервис «Яндекс пульс блогосферы» (<http://blogs.yandex.ru/pulse/>) также позволяет отображать динамику публикаций в блогах, содержащих заданные пользователем ключевые слова. На рис. 2 приведена динамика сообщений, соответствующих запросам «Олланд» и «Саркози» за период с ноября 2011 по май 2012 года.

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. На рис. 3 представлена динамика публикаций, соответствующих запросам «шовинизм» и «космополитизм» за период с 1820 по 2010 год.

### Пульс блогосферы — опланд и саркози

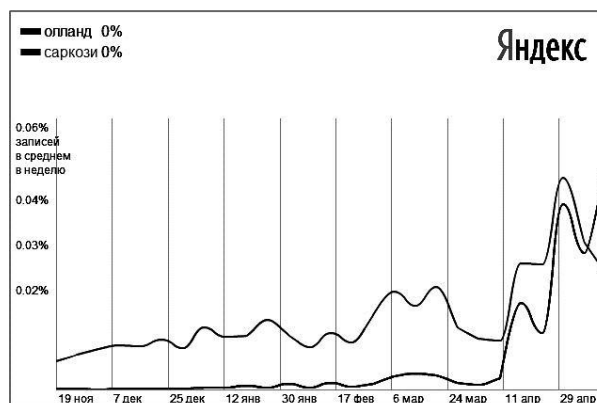


Рис.2. Динамика блогов, содержащих заданные слова

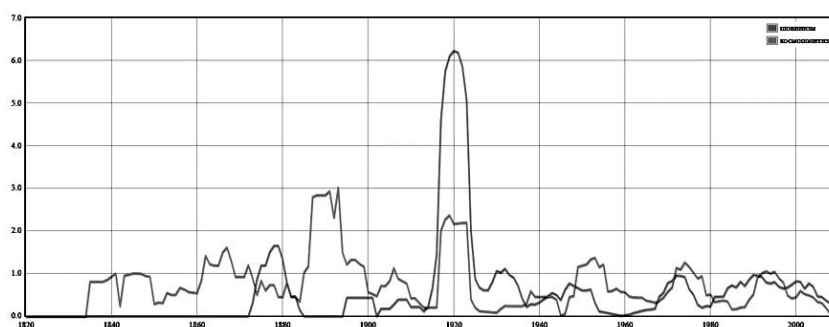


Рис.3. Динамика публикаций (Национальный корпус русского языка), содержащих заданные слова

Конечно, в лучшем положении пользователи профессиональных систем контент-мониторинга. 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 [12], реализующая данную функциональность.

При изучении трендов информационных операций в качестве временных рядов рассматриваются именно ряды по количеству тематических публикаций за определенный промежуток времени (чаще всего – за сутки), соответствующие этим информационным операциям. Поэтому для выявления трендов исследуются информационные потоки, соответствующие тематикам информационных операций – тематические информационные потоки.

Приведенные в [13] тренды сообщений, соответствующих этапам информационной операции приведены на рис. 4. 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 распространения электромагнитных волн:

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

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

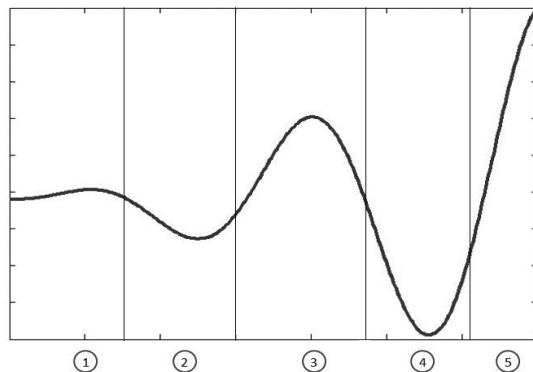


Рис. 4 – Динамика количества тематических сообщений во время проведения информационной операции: 1 – фон; 2 – затишье; 3 – «артподготовка»; 4 – затишье; 5 – атака/триггер роста

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 [14]. Вместе с тем, внедрение инноваций также можно считать информационными операциями. Поэтому обратимся к результатам соответствующих исследований. На рис. 5 приведена обоснованная в [15] диаграмма количества публикаций, соответствующая тренду инновационной деятельности.

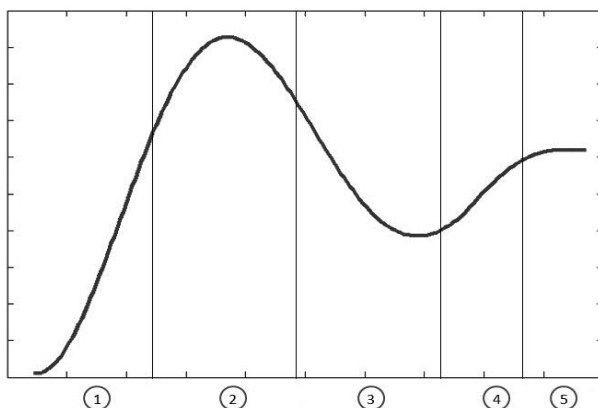


Рис. 5 – Диаграмма количества публикаций, соответствующих тренду инновационной деятельности: 1 – атака/триггер роста; 2 – пик завышенных ожиданий; 3 – утрата иллюзий; 4 – общественное осознание; 5 – продуктивность/фон

Объединяя графики, соответствующие началу информационной операции (рис. 4) и тренду инновационной деятельности (рис. 5), можно получить полный график, соответствующий отображению информационных операций в информационном пространстве (рис. 6).

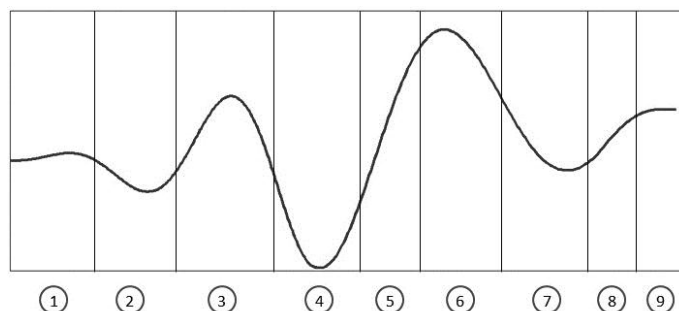


Рис. 6 – Обобщенная диаграмма количества публикаций, соответствующая всем этапам жизненного цикла информационных операций: 1 – фон; 2 – затишье; 3 – «артподготовка»; 4 – затишье; 5 – атака/триггер роста; 6 – пик завышенных ожиданий; 7 – утрата иллюзий; 8 – общественное осознание; 9 – продуктивность/фон

Предложенные модели полностью соответствуют реальным данным, которые экстрагируются системами контент-мониторинга [16, 17]. 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. 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 [18].

В качестве примера, на рис. 7 показана динамика публикаций в RUNet (тематических информационных потоков) по запросам «Банки, Кипр», «Офшор», «Вирджинские острова» за март-апрель 2013 года, в период известных кризисных событий, полученная с помощью системы InfoStream. Как видно по рис. 7, пик публикаций, связанных с банковским кризисом на Кипре приходится на 17-18 марта 2013 года, в то время, как большинство публикаций по Вирджинским островам пришелся на 4-5 апреля, когда там, со значительно меньшими масштабами, стали проявляться события, подобные кипрским. При этом следует отметить слабую коррелированность динамики информационных потоков, связанных с Кипром и Вирджинскими островами. В этом случае коэффициент взаимной корреляции соответствующих числовых рядов составил всего 0,3. При этом отмечается высокий уровень взаимной корреляции рядов соответствующих тематикам «Офшор» и «Банки Кипра» (0,73), а также «Офшор» и «Вирджинские острова» (0,77).

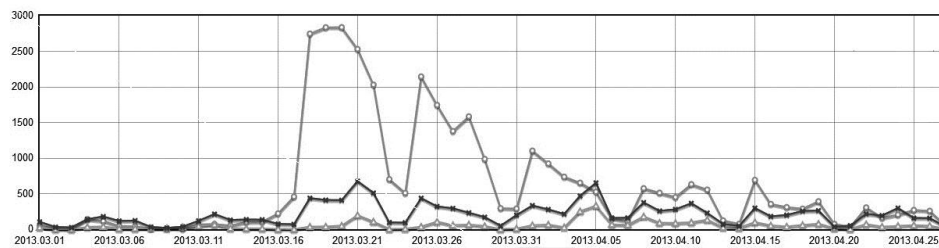


Рис. 7 – Диаграмма динамики тематических информационных потоков по запросам: о – «Банки Кипра»; Δ – «Вирджинские острова»; х – «Офшор»

По-видимому, проявления информационных операций в области офшорных банков в данном случае лучше всего увидеть при анализе более общей тематики – «Офшоры». На графике соответствующего числового ряда четко видны две области локальных экстремумов, соответствующих кризисным ситуациям на Кипре и на Вирджинских островах, а также фазы, соответствующие «затишьям» и «артподготовкам». Можно высказать предположение, что если динамика частного информационного потока в какой-то момент начинает существенно отличаться от динамики потока, соответствующего более общей тематике (как в рассматриваемом случае, «Банки Кипра» и «Офшор»), то возможно проявление признаков начала информационной операции, относящейся узкой тематике.

При проведении вейвлет-анализа (рис. 8) было принято решение использования вейвлета «Мексиканская шляпа» [9], как наиболее близкого по форме к диаграмме, приведенной на рис. 6.

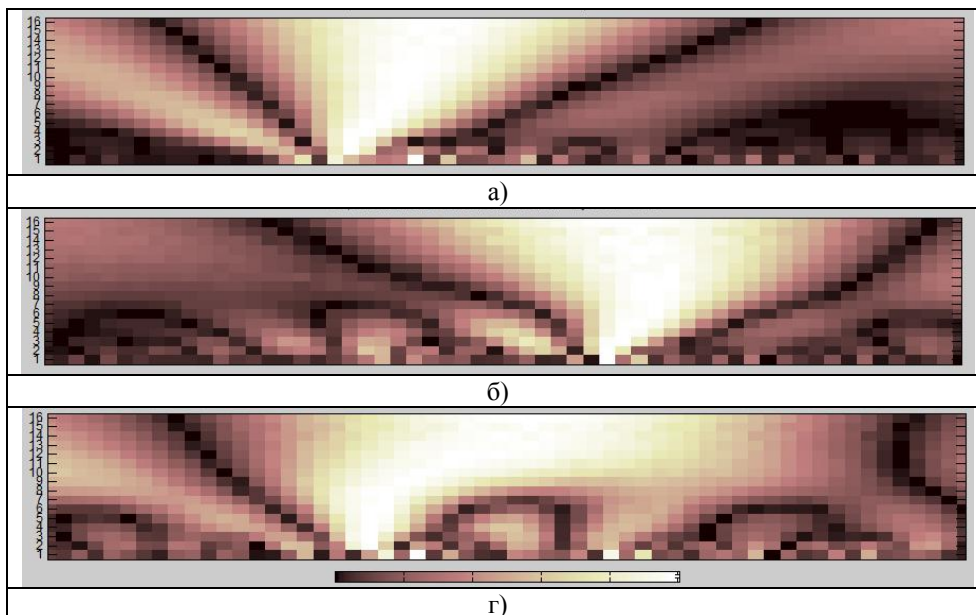


Рис. 8 – Вейвлет-спектрограммы, соответствующие динамике тематических информационных потоков по запросам: а – «Банки Кипра»; б – «Вирджинские острова»; в – «Офшор»

Рассматриваемые процессы четко просматриваются как на вейвлет-спектрограммах, так и на соответствующих им скелетонах (графиках линий экстремумов).

Приведенные модели и методы пригодны для описания общих тенденций динамики информационных процессов, однако, проблема прогнозирования остается открытой. По-видимому, более 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.

Авторы благодарны коллегам и соавторам по другим работам А.А. Снарскому, С.М. Брайчевскому, А.Т. Дармохвалу и В.Н. Фурашеву за внимание, поддержку и интерес, проявляемые при обсуждении рассматриваемых подходов.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Information operations roadmap. – DoD US. – Washington, D.C.: GPO, 2003. – 78 p.
2. Додонов А.Г., Ландэ Д.В. Живучесть информационных систем. – К.: Наук. думка, 2011. – 256 с.
3. Kleinberg J. Temporal dynamics of on-line information streams // Data Stream Management: Processing High-Speed Data Streams. – Springer, 2006.
4. Del Corso G.M., Gulli A., Romani F. Ranking a Stream of News. In Processing of the 14th International World Wide Web Conference, 2005.
5. Atkinson M., Van der Goot E. Near real time information mining in multilingual news // in WWW '09: Proceedings of the 18th international conference on World Wide Web. ACM, 2009. – P. 1153–1154.
6. Lande D.V., Braichevskii S.M. Dynamics of thematic information flows// arXiv:0805.4081, 2005.
7. Lande D., Braichevski S, Busch D. Informationsfluesse im Internet // IWP - Information Wissenschaft & Praxis, 59(2007), Heft 5. – S. 277-284.
8. 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.
9. 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](http://www.isuct.ru/~artcol/articles/Uspekhi_Fiz_Nauk/wavelet-analys.pdf)).
10. Buckheit J., Donoho D. Wavelab and reproducible research // Stanford University Technical Report 474: Wavelets and Statistics Lecture Notes, 1995. – 27 p.
11. Lande D.V., Snarskii A.A. Diagram of measurement series elements deviation from local linear approximations // Preprint Arxiv: 0903.3328, 2009.
12. 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.
13. Горбулін В.П., Додонов О.Г., Ландэ Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія. - К.: Інтертехнологія, 2009. - 164 с.
14. Bhargava S.C., Kumar A., Mukherjee A. A stochastic cellular automata model of innovation diffusion // Technological forecasting and social change, 1993. – Vol. 44. – № 1. – P. 87-97.

15. *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.
16. *Додонов О.Г., Ланде Д.В., Путятин В.Г.* Інформаційні потоки в глобальних комп'ютерних мережах. – К: Наукова думка, 2009. – 295 с.
17. *Ланде Д.В., Фурашев В.М.* Основи інформаційного і соціально-правового моделювання: монографія. – К.: ТОВ "ПанТот", 2012. – 144 с.
18. *Чертов О.Р.* Поліноми Кунченка для розпізнавання образів // Вісник НТУУ «КПІ» Інформатика, управління та обчислювальна техніка, 2009. – № 50. – С. 105-110.