

УДК 004.5

**С. В. Прищеп**

Институт проблем регистрации информации НАН Украины  
ул. Н. Шпака, 2, 03113 Киев, Украина

## **Методика выявления событийной основы информационных операций**

*Описана методология автоматического выявления событийной основы информационных операций, находящих отражение в тематических информационных потоках. Представленная методология базируется на технологиях выявления информационных операций, формирования терминологической основы предметной области, применения кластерного анализа с центроидами кластеров, определяемыми путем анализа терминологии информационного потока. Сформированные таким образом кластеры отражают основные события, происходящие во время проведения информационных операций, раскрывают технику их реализации.*

**Ключевые слова:** *информационные операции, информационные потоки, выявление событий, кластерный анализ, модель предметной области.*

### **Введение**

В настоящее время интернет-пространство становится полем боя, на котором происходят многочисленные информационные войны, отдельные информационные операции. Информационные операции определяются как «акции, направленные на воздействие на информацию и информационные системы противника, и защиту собственной информации и информационных систем» [1]. Информационные операции являются компонентами и сопровождением более общих процессов. Вместе с тем, ареной информационных операций является информационное пространство, которое, с одной стороны, является местом информационных сражений, а с другой — средой отображения реальных боевых действий [2]. При этом информационные операции на практике подкрепляются многочисленными событиями, процессами, акциями (под событиями в рамках данной работы будем понимать значительное происшествие, явление или проявление иной деятельности как факта общественной или личной жизни). Анализ отражения событий в интер-

© С. В. Прищеп

нет-пространстве дает возможность выявить участников информационных противоборств, методы информационного воздействия, раскрыть технику реализации информационных операций.

### Цель работы

Целью данной работы является создание и обоснование методики выявления событийной основы информационных операций. Реализация этой методики позволит определить временные рамки информационной операции, выявить основные события, которые сопровождают информационную операцию, увидеть приемы информационного воздействия.

При изучении информационных операций необходимо определить объективные критерии, и в качестве одного из таких, можно рассматривать динамику распространения информационных сюжетов в соответствующем фрагменте информационного пространства. Исследованию динамики информационных потоков посвящены многочисленные научные работы [3–6], в которых показано, что в типовых ситуациях динамике распространения новостей, информационного сюжета присущ характер «всплеска», волны с явным периодом возрастания его влияния и плавным спадом. Вместе с тем, вопрос определения событийной основы информационных операций остается открытым.

### Кривая динамики информационной операции

Предполагается, что системное нарушение типовой динамики некоторых тематических информационных потоков в открытом информационном пространстве может свидетельствовать об информационных операциях [7]. При исследовании информационных операций также большое внимание уделяется анализу динамики информационных потоков. В работе [8] представлен типовой шаблон информационной операции (рис. 1), который позволяет для их выявления пользоваться доступными аналитическими средствами, например, корреляционным анализом [9].

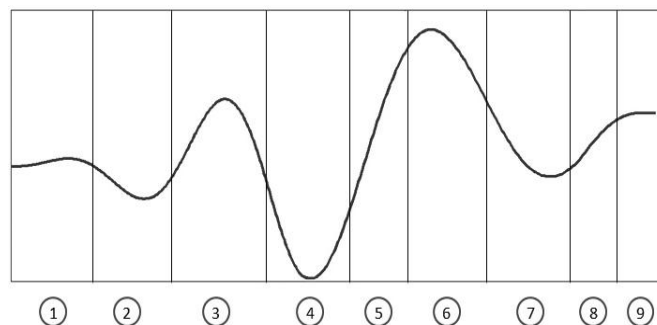


Рис. 1. Обобщенная диаграмма, соответствующая всем этапам жизненного цикла информационных операций: 1 — фон; 2 — затишье; 3 — «артподготовка»; 4 — затишье; 5 — атака/триггер роста; 6 — пик завышенных ожиданий; 7 — утрата иллюзий; 8 — общественное осознание; 9 — продуктивность/фон

## Использование системы контент-мониторинга

Для получения динамики тематического потока по определенной тематике можно воспользоваться системой контент-мониторинга, в качестве которой была выбрана система InfoStream, охватывающая в настоящее время 10 тыс. источников информации на русском, украинском и английском языках. В базы данных системы ежедневно поступает свыше 100 тыс. документов. Система InfoStream обеспечивает поиск, а также просмотр списка и полных текстов релевантных документов.

В приведенном на рис. 2 примере показан фрагмент интерфейса системы, через который обрабатывался запрос, относящийся к референдуму по выходу Великобритании из Европейского союза (сокр. Brexit от сочетания слов англ. Britain — Британия и англ. Exit — выход) во время 2016 года (период исследования июнь-июль 2016 года). В результате был сформирован тематический информационный массив, охватывающий 43697 документов. На рис. 3 приведен график динамики данного информационного потока, а также результат его сглаживания с окном в 7 дней.



Рис. 2. Фрагмент интерфейса системы контент-мониторинга

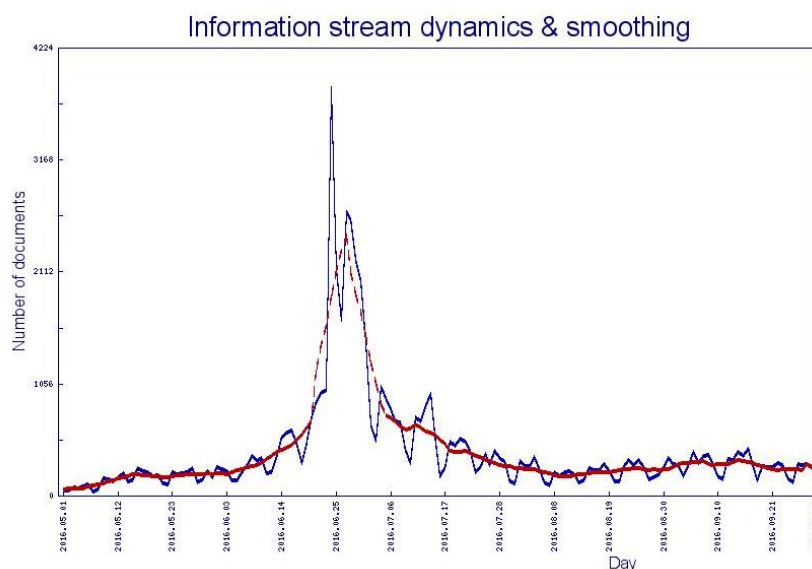


Рис. 3. Динамика тематического информационного потока (Brexit)

### Корреляция с шаблоном информационной операции

Для определения степени «близости» динамики тематического информационного потока к информационной операции использовался метод, предложенный в [9]. Идея заключается в том, чтобы сравнивать части ряда динамики информационного потока  $x_i, i = 1, \dots, N$ , с некоторым шаблоном на разных масштабах. Для этого вычисляется корреляция между частью временного ряда и некоторым шаблоном  $p$  (масштабированной частью графика, приведенного на рис. 1):

$$C(l, k) = \frac{\sum_{i=1}^k (x_{l+i} - \bar{x})(p_i - \bar{p})}{\sqrt{\sum_{i=1}^k (x_{l+i} - \bar{x})^2 \sum_{i=1}^k (p_i - \bar{p})^2}}$$

Полученный коэффициент  $C(l, k)$  зависит от значений  $x_{l+1}, \dots, x_{l+k}$ . То есть параметр  $l$  соответствует сдвигу шаблона, а параметр  $k$  — количеству точек в шаблоне и в рассматриваемом отрезке ряда. Параметр  $k$  в данном случае является аналогом масштаба. В данном случае для вычисления  $C(l, k)$  используются  $l$  точек ряда и шаблон длины  $k$ . При визуализации наибольшим значениям соответствуют наиболее светлые тона. На рис. 4 показана кореллограмма временного ряда, соответствующего динамике информационного потока по теме Brexit и шаблона, приведенного на рис. 1.

Анализ приведенной кореллограммы позволяет сужать временные рамки при исследовании тематического информационного потока.

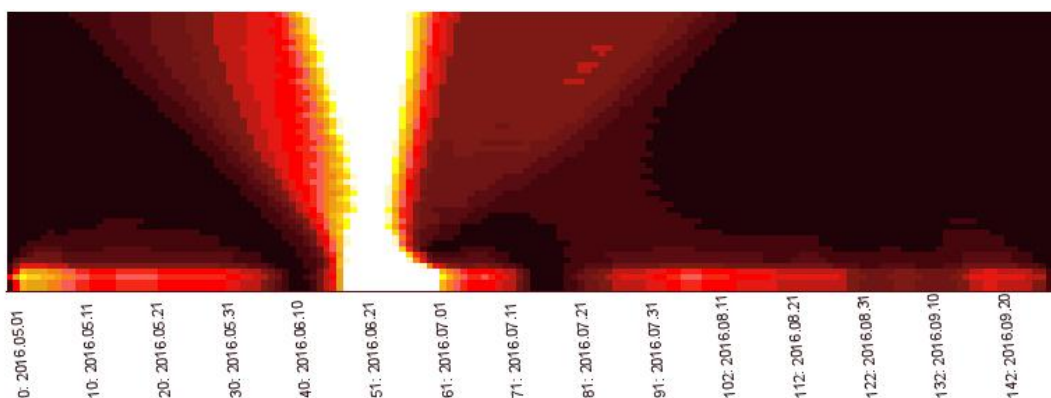


Рис. 4. Корреляционные коэффициенты  $C(l, k)$  тематического информационного потока с шаблонами информационных операций

### Терминологическая основа предметной области

Для определения событийной основы информационных операций из документов информационного потока экстрагируются опорные слова, для чего может использоваться несколько алгоритмов [10]. В частности, использовался алгоритм TF-IDF, реализованный в системе InfoStream. Значимые для информационного потока опорные слова затем автоматически сравнивались с разработанным словарем слов, которыми описываются события. В результате анализа информационного потока по тематике Brexit были выявлены такие опорные слова: PROTEST, REFERENDUM, PETITION, SIGNATURES, DEMONSTRATION, TERRORIST... (рис. 5).

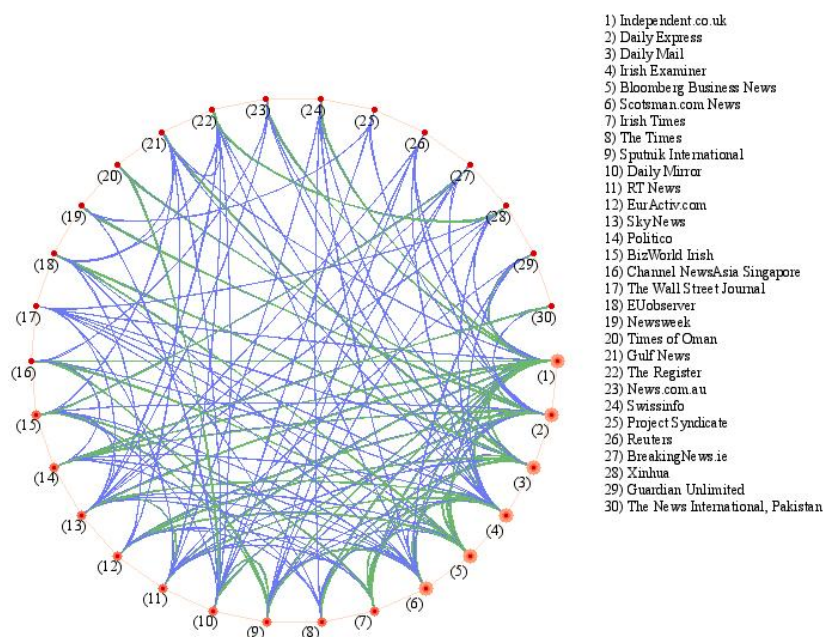


Рис. 5. Фрагмент графа горизонтальной видимости, отражающего связи источников информации по заданной теме (Brexit)

Базовый запрос для формирования информационного потока по заданной к системе InfoStream был дополнен этими словами, что позволило ограничить количество документов, оставив только те, которые описывают конкретные события. Это позволило также определить перечень информационных источников, публикующих событийную информацию и построить сеть их взаимосвязей.

### Кластеризация предметной области

Для выявления кластеров основных событий полученный информационный поток анализируется методом *k-means* (*k*-средних) группировки документов  $\{d^{(1)}, \dots, d^{(N)}\}$  в фиксированное количество кластеров, состоящий в следующем: некоторым образом выбирается  $k$  документов, которые определяются как центроиды (наиболее типичные представители) кластеров. То есть каждый кластер  $C_j (j = 1, \dots, k)$  представлен соответствующим центроидом. Близость документов к центроиду может определяться разными способами, с помощью некоторой меры  $Sim(d, C_j)$ .

После этого документ приписывается к тому кластеру, значение  $Sim(d, C_j)$  для которого оказывается самым большим. Затем для каждого из новых кластеров заново определяется центроид  $C_j (j = 1, \dots, k)$ , например, как набор самых весомых (по какому-нибудь выбранному критерию) слов из документов, входящих в данный кластер.

После этого снова происходит процесс наполнения кластеров, затем расчет новых центроидов и т.д., пока процесс формирования кластеров не стабилизируется.

Известно, что алгоритм *k-means* максимизирует функцию качества кластеризации  $Q$ :

$$Q(C_1, \dots, C_k) = \sum_{j=1}^k \sum_{d \in C_j} Sim(d, C_j).$$

Данный метод обладает высоким быстродействием (сложность порядка  $O(k, N)$ , где  $k$  — необходимое число кластеров,  $N$  — количество документов). Вместе с тем, при использовании этого метода открытым вопросом всегда остается выбор стартовых документов для кластеризации — центроидов. Автором предложено в качестве центроидов использовать определенные ранее опорные слова, которые выражают суть событий. Предложенный подход позволил простым способом реализовать кластеризацию в рамках возможностей системы InfoStream (рис. 6).



brexit&(PROTEST|REFERENDUM|PETITION|SIGNATURES|DEMONSTRATION|TERAKT|TERRORIST)

Период: Другой ▼  
От: 201606 До: 201607

Обзор основных сюжетов

brexit & (PROTEST | REFERENDUM | PETITION | SIGNATURES | DEMONSTRATION | TERAKT | TERRORIST) ;  
документов - 3000, сюжетов - 1239

- Scotland is determined to retain its EU status**  
by Iain Robertson Special To The Japan Times GLASGOW, SCOTLAND - The result of the United Kingdom's referendum on its continued membership of the European Union has caused a constitutional crisis that has reignited calls for Scottish independence. June's plebiscite on EU membership saw 53 percent of the English electorate vote for Brexit.  
[Сюжет полностью \(141\)](#)
- Asia markets soar again, led by Tokyo**  
Japanese stocks led a broad rally on Asian markets on Thursday (Jul 14), posting a fourth straight gain as investors took heart from another record close on Wall Street, while the pound rose ahead of a key central bank meeting. A pedestrian looks at a screen showing movements for the Tokyo Stock Exchange in Tokyo.  
[Сюжет полностью \(79\)](#)
- Scotland, Wales, Northern Ireland gov'ts should OK Brexit deal: Welsh first minister**  
Baku-APA Britain's deal to exit the European Union (EU) should first be ratified by the country's four parliaments, the First Minister of Wales said Friday, APA reports quoting Xinhua. Carwyn Jones was speaking after a meeting in Cardiff Friday of the British Irish Council.  
[Сюжет полностью \(74\)](#)
- Key survey sees UK economy shrinking after Brexit vote**  
British business activity sank in July to its lowest level since April 2009, a survey showed Friday. People hold up pro-Europe placards as thousands of protesters take part in a March for Europe, through the centre of London on July 2, 2016, to protest against Britain's vote to leave the EU.  
[Сюжет полностью \(35\)](#)
- Boris Johnson shrugs off Europe's shock at top diplomatic job**  
By Alasdair Sandford | With REUTERSFollow @SandforAlasdair 14/07 16:27 CET | updated xx mn ago | updated at xx 1468520284 1468506423 "It is inevitable there is going to be a certain amount of plaster coming off the ceiling in the chancelleries of Europe."  
[Сюжет полностью \(96\)](#)

Рис. 6. Основные сюжеты (кластеры) тематического информационного потока

## Выводы

Предлагаемая методика выявления событийной основы информационных операций охватывает следующие этапы исследования, результат каждого из которых может рассматриваться как самостоятельный информационный продукт:

- 1) формирование тематического информационного потока по исследуемой тематике;
- 2) исследование динамики полученного потока, выявление признаков информационной операции;
- 3) сужение временных рамок информационного потока, получение более репрезентативной выборки документов;
- 4) определение терминологической основы описания событий в рамках исследуемой предметной области. Определение источников информации о событиях (при необходимости);
- 5) кластеризация, выявления основных событий, сопровождающих информационную операцию

Определяемые в результате предложенной методики кластеры документов соответствуют основным событиям, сопровождающим информационную операцию. Дальнейший содержательный анализ этих кластеров позволяет выявить основных действующих лиц информационной операции, раскрыть технику ее реализации.

1. Information operations roadmap. DoD US. Washington, D.C.: GPO, 2003.

2. Додонов А.Г., Ландэ Д.В., Цыганок В.В., Андрейчук О.В., Каденко С.В., Грайворонская А.Н. Распознавание информационных операций. Киев: ООО «Инжиниринг», 2017. 282 с.
3. Del Corso G.M., Gulli A., Romani F. Ranking a stream of news. In Proceedings of 14th International World Wide Web Conference. Chiba, Japan, 2005. P. 97–106,
4. Kleinberg J. Temporal dynamics of on-line information streams. Data Stream Management: Processing High-Speed Data Streams. Springer, 2006.
5. Ландэ Д.В., Фурашев В.Н., Брайчевский С.М., Григорьев А.Н. Основы моделирования и оценки электронных информационных потоков: моногр. Киев: Инжиниринг, 2006. 176 с.
6. Rakesh V., Singh D., Vinzamuri B., Reddy C.K. Personalized Recommendation of Twitter Lists Using Content and Network Information . Association for the Advancement of Artificial Intelligence (Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, 2014.
7. Горбулін В.П., Додонов О.Г., Ландэ Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: моногр. Київ: Інтертехнологія, 2009. 164 с.
8. Додонов А.Г., Ландэ Д.В. Мультиагентная модель поведения тематических информационных потоков. Материалы VI Всероссийской мультikonференции по проблемам управления (30 сен. – 5 окт. 2013 г.). 2013. Т. 4. С. 102-107.
9. Грайворонская А.Н., Ландэ Д.В. Элементы нелинейного анализа информационных потоков. *Реєстрація, зберігання і оброб. даних*. 2016. Т. 19. № 3. С. 13–33.
10. Ландэ Д.В., Снарский А.А. Подход к созданию терминологических онтологий. *Онтология проектирования*. 2014. № 2(12). С. 83–91.

Поступила в редакцию 12.06.2018