

ДИНАМИКА ЧАСТОТЫ УПОМИНАНИЯ СЛОВ «МОДЕРНИЗАЦИЯ», «ИННОВАЦИИ» И «СКОЛКОВО» В РУССКОЯЗЫЧНОЙ БЛОГОСФЕРЕ

Ключевые слова: блогосфера, модернизация, инновации, Сколково, системная социология

Введение

Блогосфера - «общественное мнение Интернета» [1], анализ которого позволяет оперативно в динамике выявить более эмоциональные и откровенные точки зрения блоггеров о проблемах модернизации и инновационного развития России, по сравнению с мнением респондентов в традиционных социологических опросах общественного мнения о модернизации России [2-3], изучать динамику интереса блоггеров к данным темам во времени. Автор [4] уже содержательно исследовал мнения блоггеров о модернизации России, о «драйверах» модернизации и инновационного развития России, об инновационных инициативах Президента РФ Д.А.Медведева, в частности, об инновационном проекте «Сколково».

В данном исследовании автором была поставлена следующая исследовательская задача:

Выявить системные закономерности динамики частоты упоминания слов «Модернизация», «Инновации» и «Сколково» в русскоязычной блогосфере за период 2001-2010 гг.

Методология

Решение поставленной исследовательской задачи осуществлялось в рамках естественнонаучной и социально-инженерной методологических парадигм системной социологии [5-6]. В частности, исследование проводилось в рамках Социологии 2.0 (Интернет-социологии) [7] и Computational Sociology (вычислительной социологии) [6] - разделов системной социологии.

Методика

Ежегодная, помесечная и ежедневная динамика частоты упоминания слов «Модернизация», «Инновации» и «Сколково» в русскоязычной блогосфере за период 2001-2010 гг. была получена автором с помощью онлайн аналитического сервиса «Пульс блогосферы» Яндекса [8].

Существует более 200-х методов анализа и моделирования динамики, практически их всех разделов современной математики [цит. по 4], например Spectral Analysis, Fractal Analysis, ARIMA, Multivariate Wavelet Analysis, Singular Spectrum Analysis, Vector Autoregression Analysis, Dynamic Factor Analysis, Chronological Clustering, «нейронные» сети, клеточные автоматы, модели детерминированного хаоса, сигналов и т.д., которые реализованы в многочисленных пакетах для анализа временных рядов.

В соответствии с принятой методологией Computational Sociology (вычислительной социологии) [6] в данном исследовании использовалось множество методов анализа временных рядов, чтобы проанализировать различные системные свойства динамики. В частности, в данном исследовании использовался пакет CaterpillarSSA 3.40 [9], в котором реализован Singular Spectrum Analysis (SSA), предназначенный для автоматического анализа и прогнозирования многомерных временных рядов. SSA основан на Singular value decomposition (SVD) of the trajectory matrix, Covariance Principal Component Analysis (PCA). Более подробно с методом Singular Spectrum Analysis (SSA) заинтересованный читатель может ознакомиться на сайте Gistatgroup [10]. Использовался пакет Gretl (v.1.9.1) [11], в котором реализован Vector Autoregression Analysis (VAR) – классический эконометрический анализ, предназначенный для автоматического анализа и прогнозирования многомерных временных рядов. VAR анализирует зависимость нескольких эндогенных переменных (вектор зависимых переменных), которые зависят от собственных лагов (запаздываний) и от лагов других экзогенных переменных (вектор независимых переменных). Пакет TableCurve 2D (v.5.01) [12], предназначенный для автоматической аппроксимации математических функций, пакет FracLab Toolbox [13], предназначенный для фрактального анализа [14] временных рядов, пакет STATISTICA, в котором реализованы методы анализа временных рядов, пакет Wavelet Toolbox MATLAB, предназначенный для вейвлет-анализа временных рядов.

Пропущенные (отсутствующие) значения в данных заполнялись с помощью стандартной процедуры, принятой в анализе данных, а именно, добавлялось число $\varepsilon = 0.000001$.

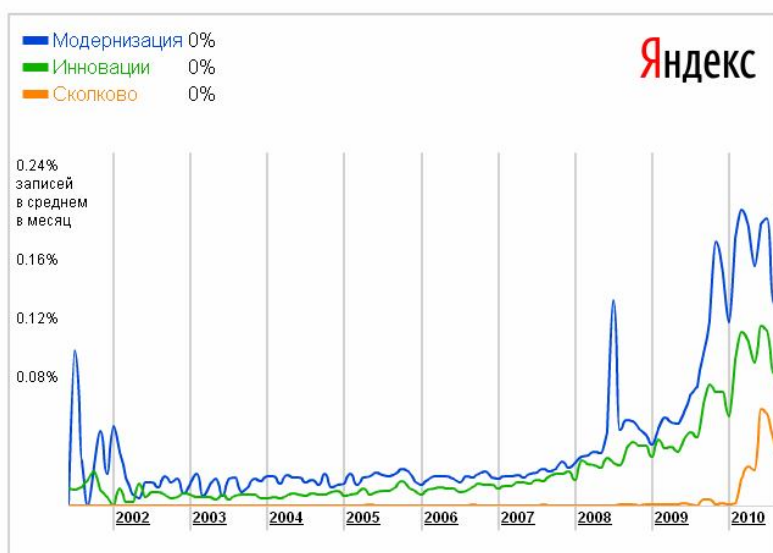
Полученные результаты

На рис. 1-3 представлена динамика частоты упоминания слов «Модернизация», «Инновации» и «Сколково» в русскоязычной блогосфере за период 2001-2010 гг. по годам, месяцам и дням.

Рис.1

Годовая динамика частоты упоминания слов «Модернизация», «Инновации» и «Сколково» в русскоязычной блогосфере за 2001-2010 гг.

Пульс блогосферы — [Модернизация](#) , [Инновации](#) и [Сколково](#)



Помесячная динамика частоты упоминания слов «Модернизация», «Инновации» и «Сколково» в русскоязычной блогосфере за 2009-2010 гг.

Пульт блогосферы — **Модернизация** , **Инновации** и **Сколково**

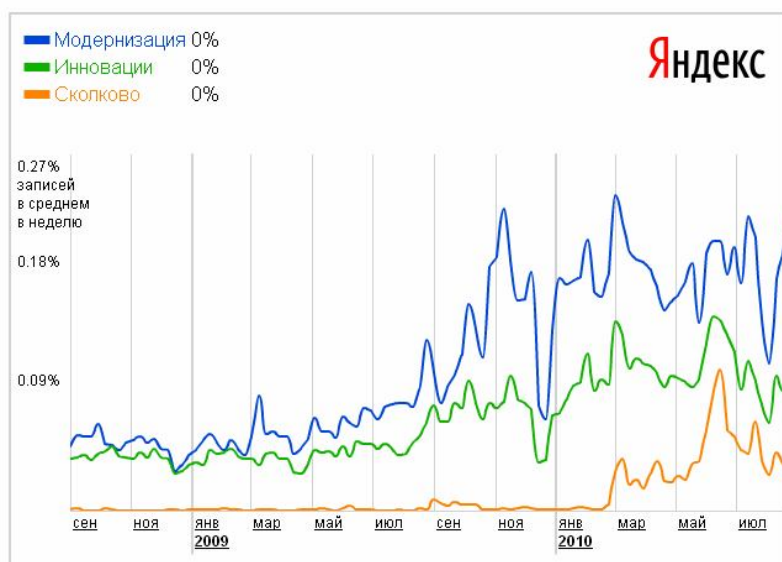
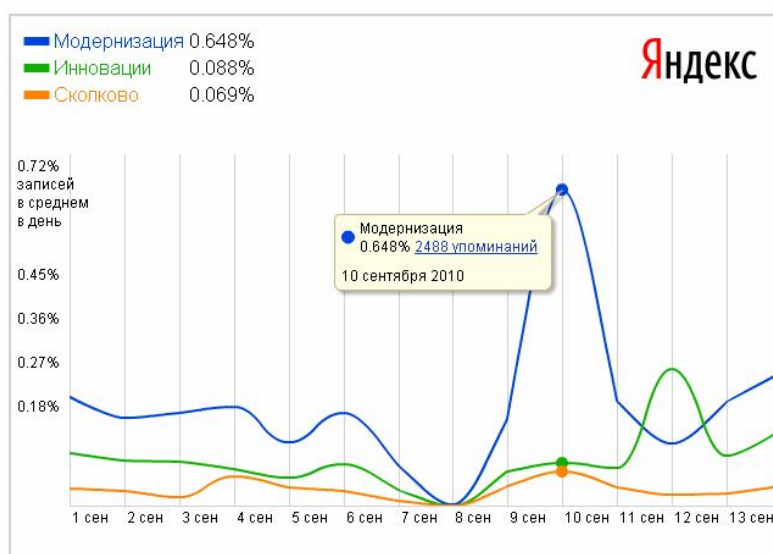


Рис.3

Ежедневная динамика частоты упоминания слов «Модернизация», «Инновации» и «Сколково» в русскоязычной блогосфере за сентябрь 2010 года

Пульт блогосферы — **Модернизация** , **Инновации** и **Сколково**



За период 2001 – 2010 гг. слово «Модернизация» упоминалось, в среднем за месяц - 2087 раз, слово «Инновации» упоминалось, в среднем за месяц – 1243 раза, слово «Сколково» - 209 раз. В частности, за период 14 июля – 14 сентября

2010 года, слово «Модернизация» упоминалось, в среднем за день - 557 раз, слово «Инновации» упоминалось, в среднем за день – 286 раз, слово «Сколково» - 128 раз.

На рис.4 представлена помесечная динамика совместной встречаемости (в одной записи блоггера) слов «Модернизация», «Инновации» и «Сколково» за период 1 января – 14 сентября 2010 года.

Рис.4

Помесечная динамика совместной частоты упоминания слов «Модернизация», «Инновации» и «Сколково» в русскоязычной блогосфере за период 1 января – 14 сентября 2010 года

Пульс блогосферы — [Модернизация](#), [Инновации](#), [Сколково](#)



Примечание: Наблюдаемый «всплеск» совместной частоты встречаемости слов «Модернизация», «Инновации», «Сколково», наблюдаемый за период 21 июня – 28 июня 2010 года, был обусловлен, в значительной мере, комментариями на выступления участников на Петербургском Экономическом Форуме (Санкт-Петербург, 17-19 июня 2010 г.)

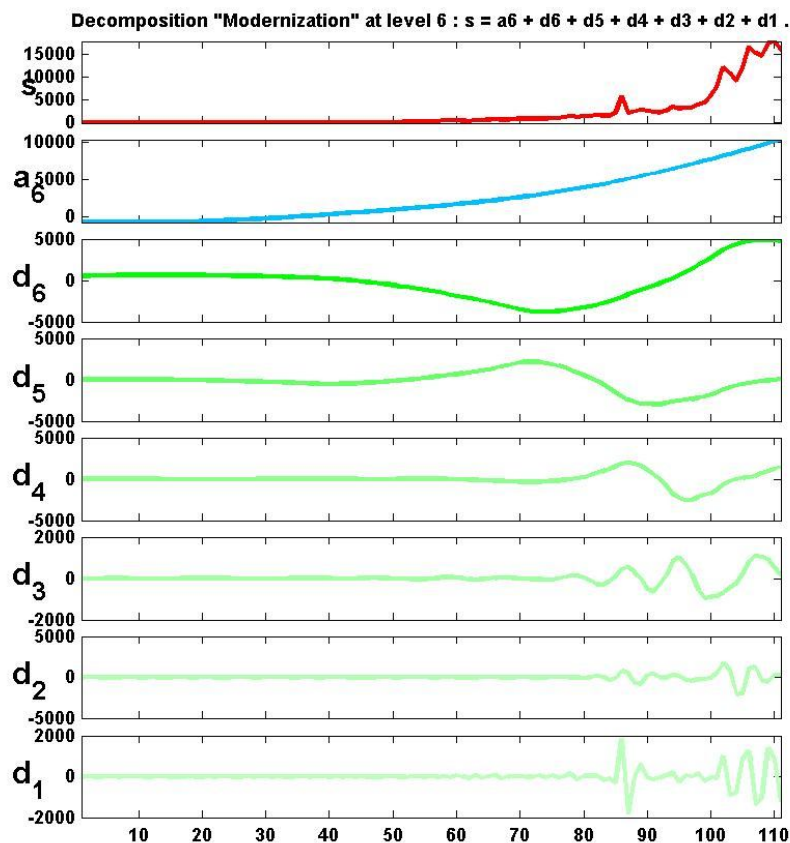
Ввиду ограниченного периода наблюдений и незначительного количества совместной частоты упоминания слов «Модернизация», «Инновации» и «Сколково», в настоящем исследовании данные, представленные на рис.4, далее не анализировались. В данном исследовании автор выбрал для дальнейшего анализа помесечные данные (111 месяцев за период июнь 2001 – август 2010 гг.) частоты упоминания слов «Модернизация», «Инновации» и «Сколково».

На рис. 5-7 представлены результаты вейвлет-анализа [14] помесечной динамики частоты упоминания слов «Модернизация», «Инновации» и «Сколково».

Для декомпозиции динамики на 6-ти уровнях использовался вейвлет Добеши с параметром порядка 4.

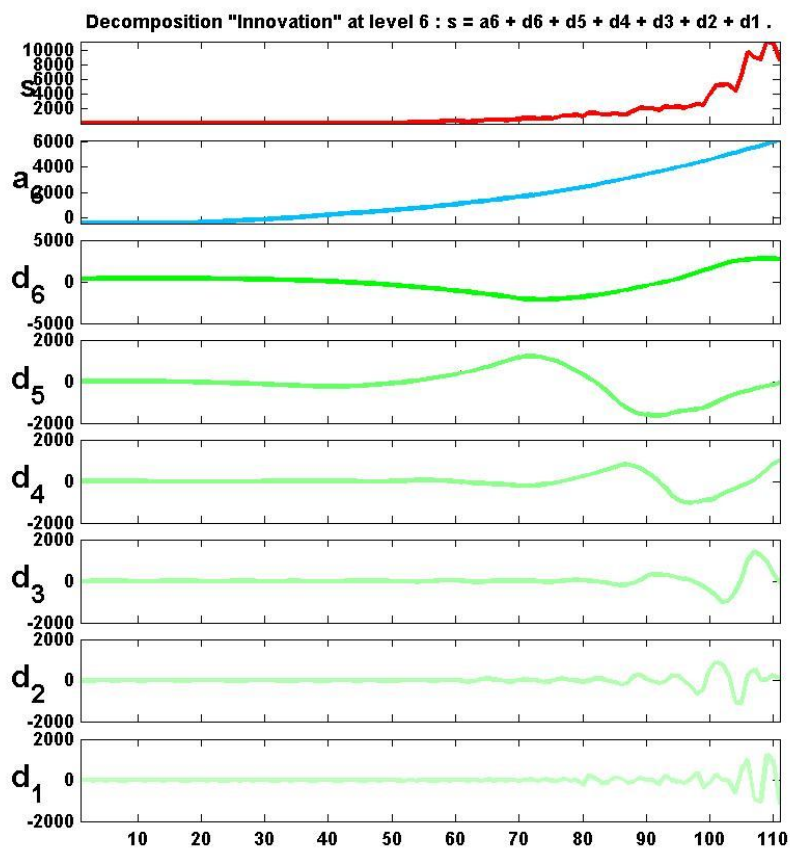
Рис.5

Вейвлет-декомпозиция динамики частоты упоминания слова «Модернизация»



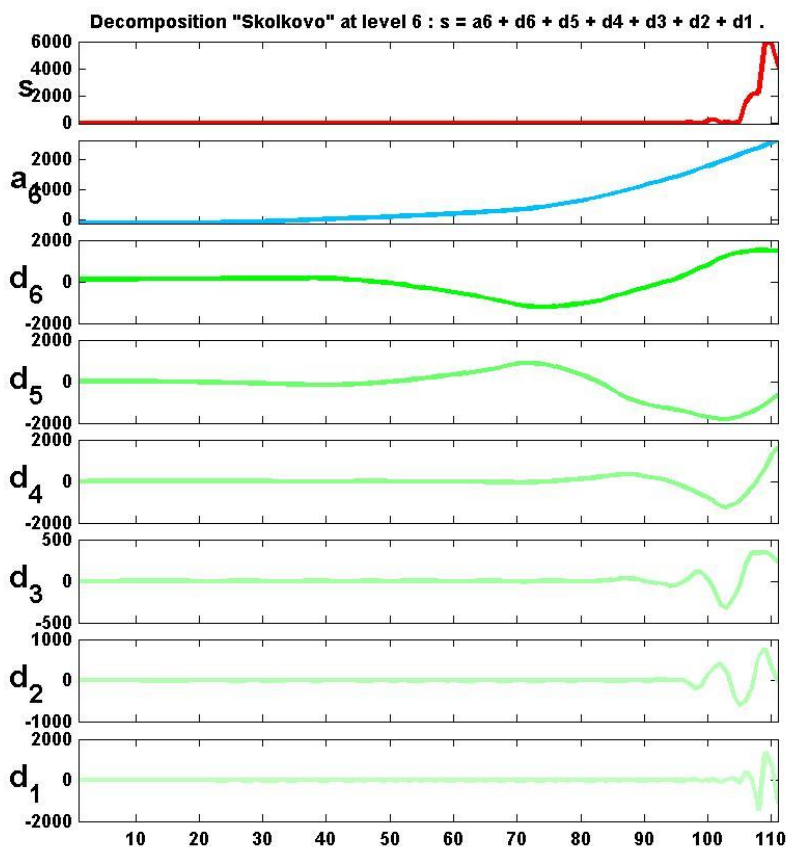
Примечание: s - исходный временной ряд за 111 месяцев, a_6 - тренд (долговременная составляющая) временного ряда, $d_1 - d_6$ - периодические составляющие временного ряда.

Вейвлет-декомпозиция динамики частоты упоминания слова «Инновации»



Примечание: s - исходный временной ряд за 111 месяцев, a_6 - тренд (долговременная составляющая) временного ряда, $d_1 - d_6$ - периодические составляющие временного ряда.

Вейвлет-декомпозиция динамики частоты упоминания слова «Сколково»



Примечание: s - исходный временной ряд за 111 месяцев, a_6 - тренд (долговременная составляющая) временного ряда, $d_1 - d_6$ - периодические составляющие временного ряда.

В таблице 1 представлены значения фрактальной размерности (меры самоподобия, сложности) [15] динамики частоты упоминания слов «Модернизация», «Инновации» и «Сколково», вычисленные с помощью метода регуляризации в пакете FracLab Toolbox MATLAB [13].

Таблица 1

Фрактальная размерность динамики частоты упоминания слов
«Модернизация», «Инновации» и «Сколково»

Динамика частоты упоминания слов	Значение фрактальной размерности
Модернизация	1.2065
Инновации	1.2241
Сколково	1.3293

Из результатов, представленных в таб.1 следует, что по значению фрактальной размерности, динамика частоты упоминания слов «Модернизация» и «Инновации» «ближе» друг к другу, чем динамика упоминания слова «Сколково».

В таблице 2 представлены результаты кросскорреляционного и автокорреляционного анализов, а именно, наибольшие значения кросскорреляции и автокорреляции с лагами (запаздываниями) между динамикой частоты упоминания слов «Модернизация», «Инновации» и «Сколково».

Таблица 2

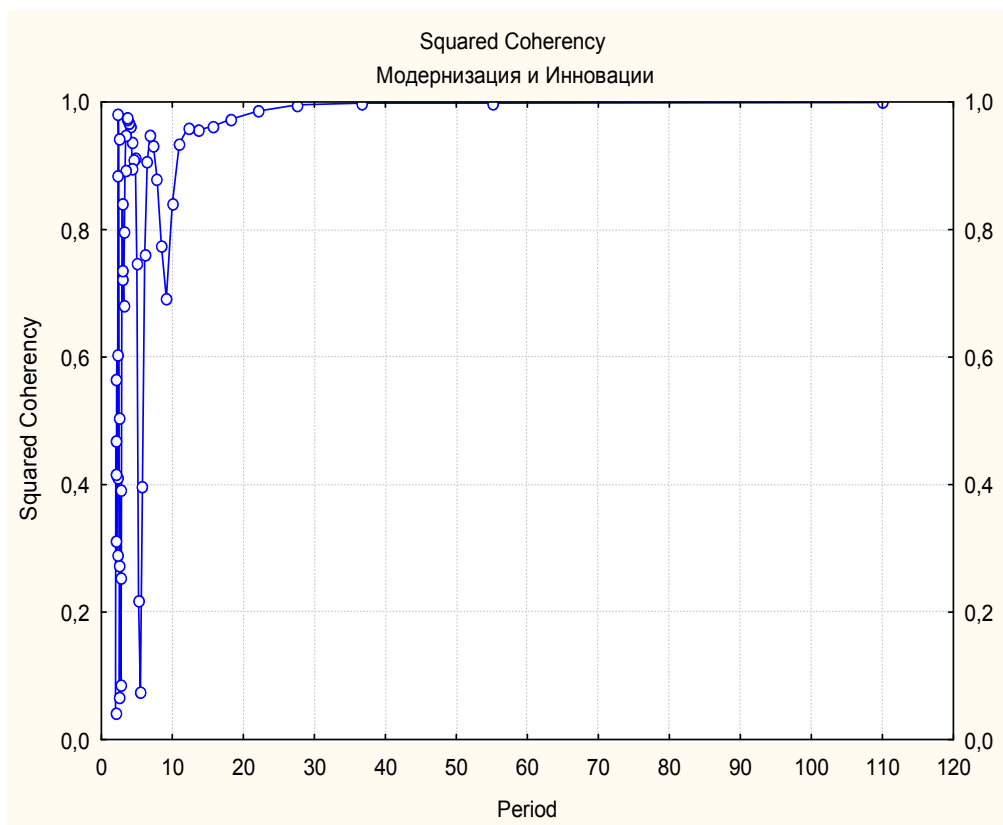
Кросскорреляция и автокорреляция с лагами (запаздываниями) между динамикой частоты упоминания слов «Модернизация», «Инновации» и «Сколково»

Динамика частоты упоминания слов	Значение коэффициента кросскорреляции	Лаг	Значение коэффициента автокорреляции	Лаг
Модернизация и Инновации	0.9875	0		
Модернизация и Сколково	0.7711	0		
Инновации и Сколково	0.8014	0		
Модернизация			0.922	1 месяц
Инновации			0.929	1 месяц
Сколково			0.814	1 месяц

Из результатов, представленных в таблице 2 следует, что ежемесячная динамика частоты упоминания слов «Модернизация», «Инновации» и «Сколково» сильно взаимосвязана, при этом изменения происходят без запаздываний (одновременно). На динамику частоты упоминания слов «Модернизация», «Инновации» и «Сколково» в наибольшей мере влияет частота упоминания слов «Модернизация», «Инновации» и «Сколково» за предыдущий месяц.

На рис. 8 представлен результат кросс-спектрального анализа когерентности динамики частоты упоминания слов «Модернизация» и «Инновации», вычисленный с помощью пакета STATISTICA.

Когерентность (согласованность) динамики частоты упоминаний слов
«Модернизация» и «Инновации» за период 2001 - 2010 гг.



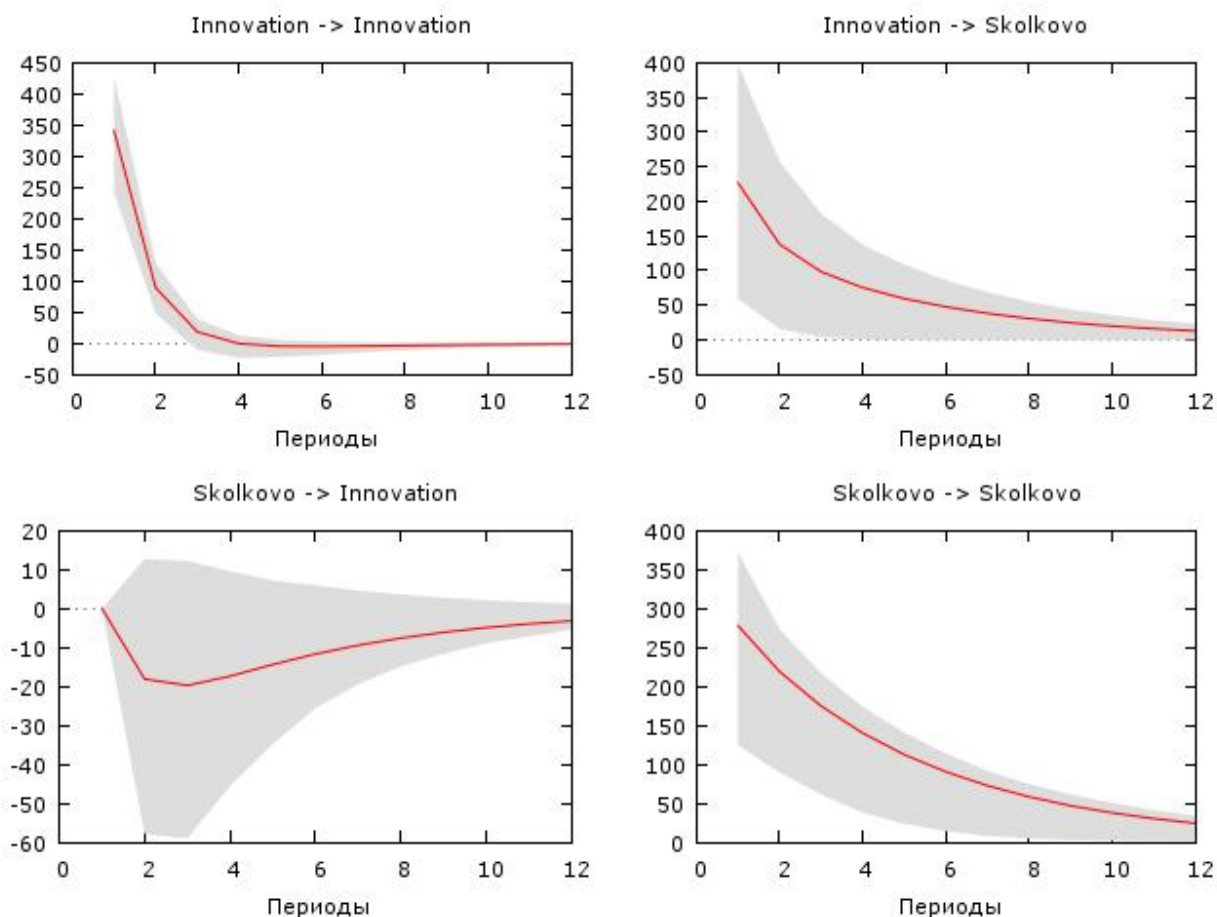
Из рис. 8 следует, что значение когерентности (согласованности) динамики частоты упоминаний слов «Модернизация» и «Инновации» за период 2001 - 2010 гг. является высоким, за исключением первых месяцев наблюдения в 2001 году, что обусловлено «всплесками» частоты упоминания слова «Модернизация» в данный ограниченный период времени (см. рис.1). Аналогичные закономерности высокого значения когерентности наблюдались также при кросс-спектральном анализе совместной динамики частоты упоминания слов «Модернизация» и «Сколково», «Инновации» и «Сколково», за период 2001 - 2010 гг.

Более детальную информацию о совместном изменении динамики частоты упоминания слов «Инновации» и «Сколково» в зависимости от изменения частоты упоминания слова «Модернизация» при запаздывании на n периодов времени, дают функции реакции на импульсы (IRF - impulse response function). Для выявления функций реакции на импульсы (IRF) была построена статистическая модель на основе Vector Autoregression Analysis (VAR) из пакета Gretl (v.1.9.1) [11]. В качестве эндогенных переменных выступала частота упоминания слов «Инновация» и «Сколково», а в качестве экзогенной переменной – частота упоминания слова «Модернизация». В Приложении представлены статистики для

построенной модели (мера аппроксимации модели $R^2 \geq 0.847$) с лагом (запаздыванием) в один месяц. Полученные результаты представлены на рис.9.

Рис.9

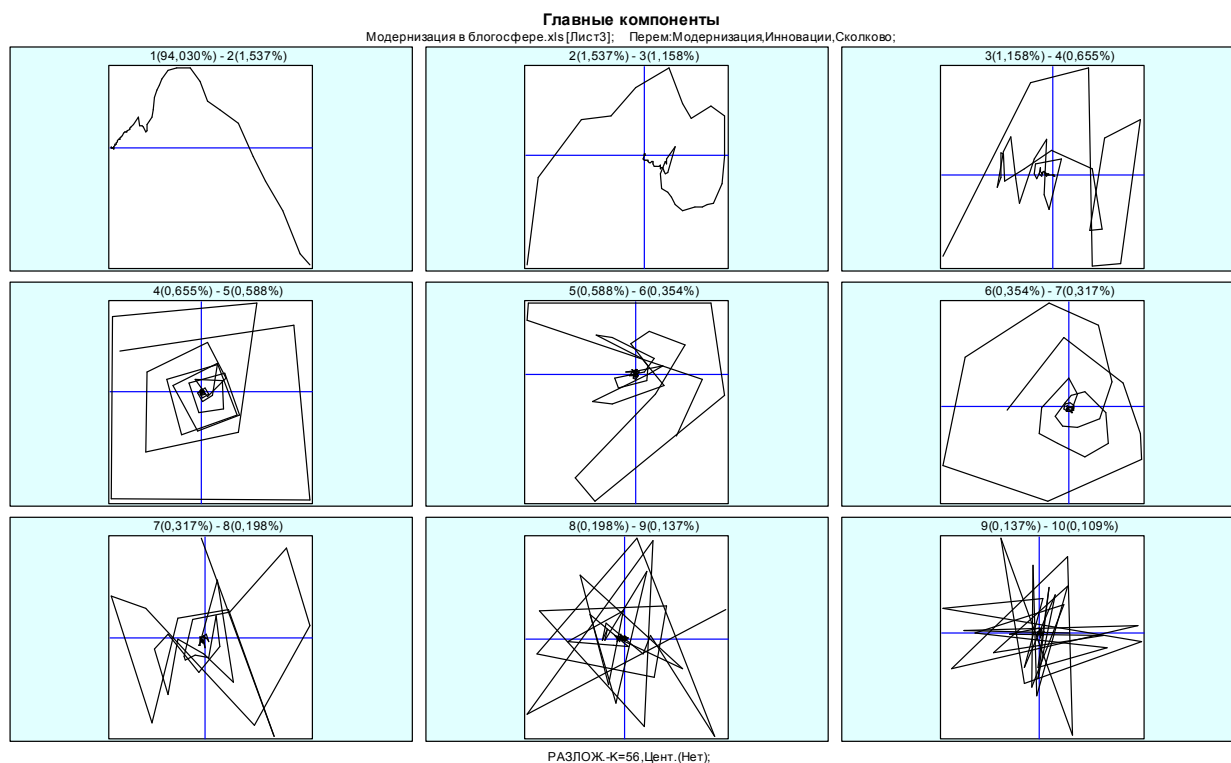
Функции реакции на импульсы для частот упоминания слов «Инновации» и «Сколково», в зависимости от изменения частоты упоминания слова «Модернизация»



Примечание: Серым цветом показан бутстреп доверительный интервал. Величина периода равна запаздыванию на 0,1,...,12 месяцев

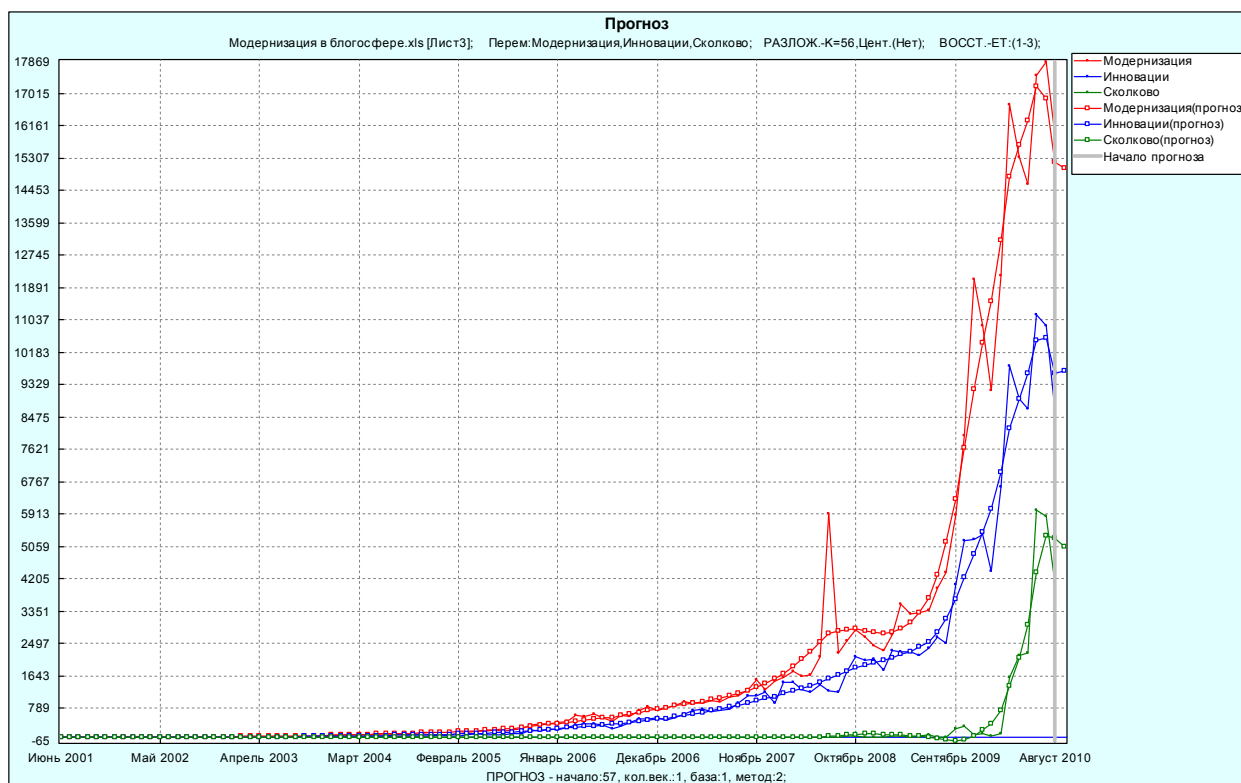
На рис. 10-11 представлены результаты Singular Spectrum Analysis (SSA), полученные с помощью пакета CaterpillarSSA 3.40 [9], для месячной динамики частоты упоминания слов «Модернизация», «Инновации», «Сколково».

Главные компоненты матрицы траекторий динамики частот упоминания слов
«Модернизация», «Инновации» и «Сколково»



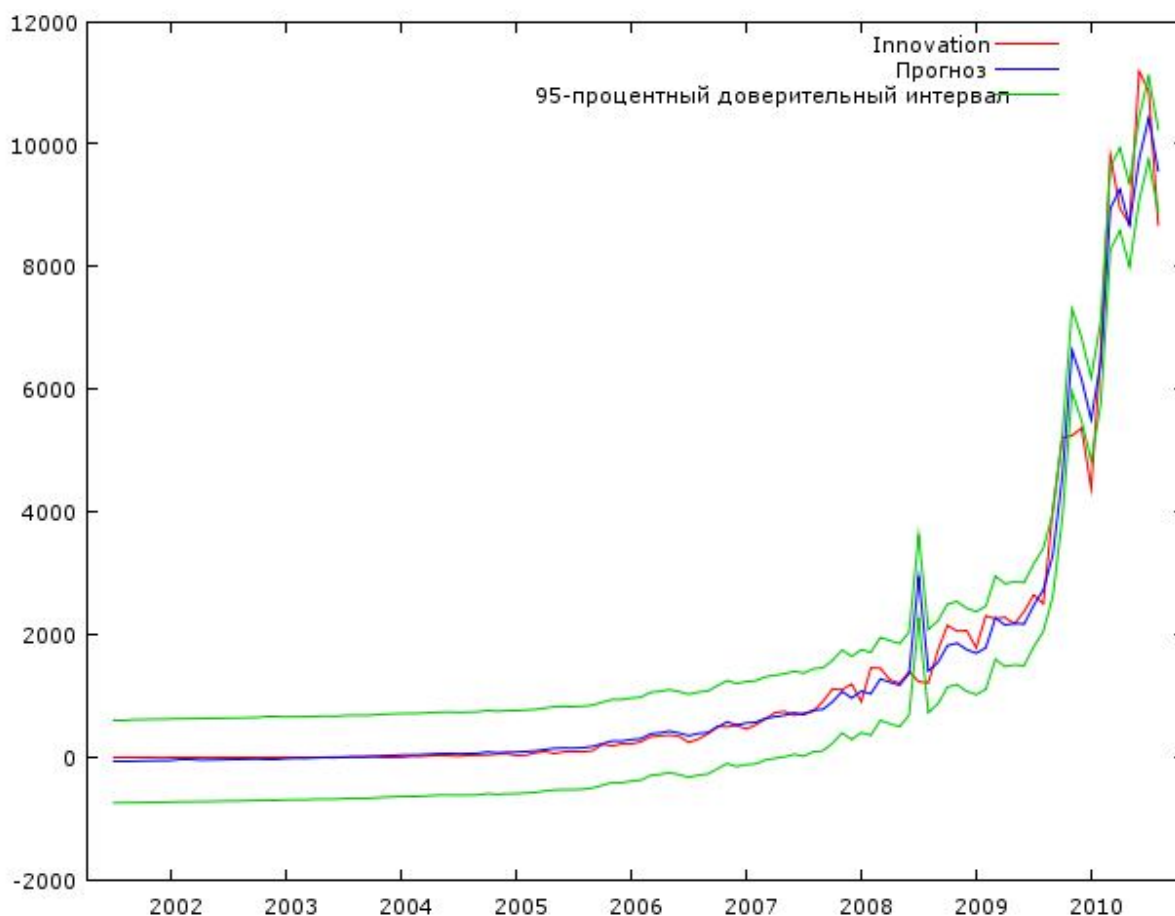
Из рис. 10 следует, что первая главная компонента объясняет 94.030% месячной динамики частот упоминания слов «Модернизация», «Инновации» и «Сколково». Проведенный содержательный анализ комментариев блоггеров, по словам «Модернизация», «Инновации» и «Сколково» в различные моменты времени (см. рис.1-3), показал, что первая главная компонента отражает интерес блоггеров к развитию России. Таким образом, исходя из контекста употребления и содержания слов «Модернизация», «Инновации» и «Сколково» в записях блоггеров, первую главную компоненту можно условно назвать «Развитие России».

Аппроксимация динамики частоты упоминания слов «Модернизация», «Инновации» и «Сколково»



Для проверки надежности результатов, представленных на рис.11, была использована построенная модель VAR (см. Приложение). В качестве иллюстрации, на рис. 12 представлена аппроксимация динамики частоты упоминания слова «Инновации», восстановленная с помощью модели VAR.

Аппроксимация динамики частоты упоминания слова «Инновации» с помощью построенной модели VAR



В целом, результаты анализа, представленные на рис. 5-8,11-12, полученные с помощью различных методов анализа динамики, свидетельствуют в пользу хорошей согласованности полученных результатов, в частности, о динамике тренда (долговременной составляющей временного ряда).

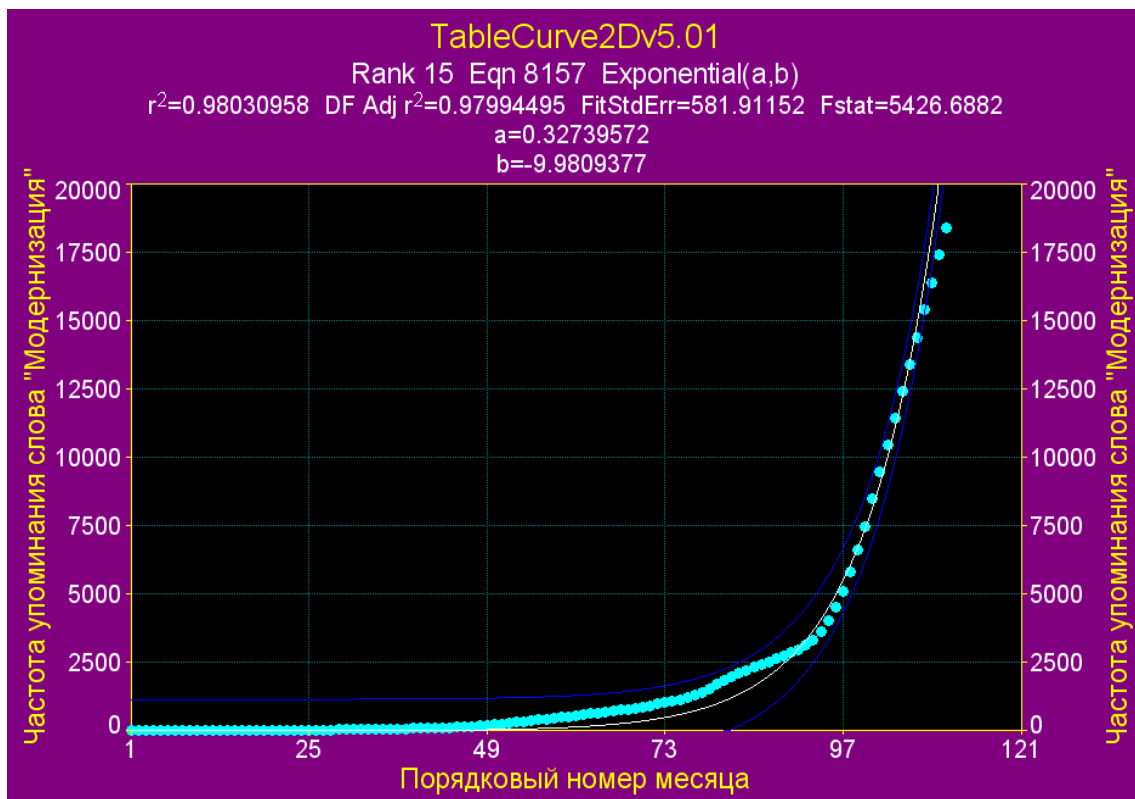
Обсуждение полученных результатов

Из полученных результатов, представленных на рис. 1-3,5-8,11-12 следует, что с течением времени наблюдается рост частоты упоминания слов «Модернизация», «Инновации» и «Сколково» в русскоязычной блогосфере. Проанализируем тренды (долговременные составляющие временного ряда), представленные на рис. 1-3,5-8, 11-12 с помощью пакета TableCurve 2D (v.5.01) [12]. Фильтрация (автоматическое сглаживание) «всплесков» (резкий рост количества упоминаний слов в короткий период времени) в динамике осуществлялось с помощью алгоритма Loess. Величина сглаживания составляла

до 20.8%. Проведенный содержательный анализ комментариев блоггеров, по словам «Модернизация», «Инновации» и «Сколково» в моменты «всплесков», что позволяет онлайн аналитический сервис «Пульс блогосферы» Яндекса [8], показал, что данные «всплески» обусловлены заявлениями и решениями Президента РФ Д.А.Медведева о модернизации и инновационном развитии России, действиями органов государственного и местного управления, спонтанно возникающими дискуссиями в сообществах блоггеров и другими событиями. На рис. 13-15 представлены полученные результаты сглаживания с 95% доверительным интервалом.

Рис.13

Тренд частоты упоминания слова «Модернизация»



Тренд частоты упоминания слова «Инновации»

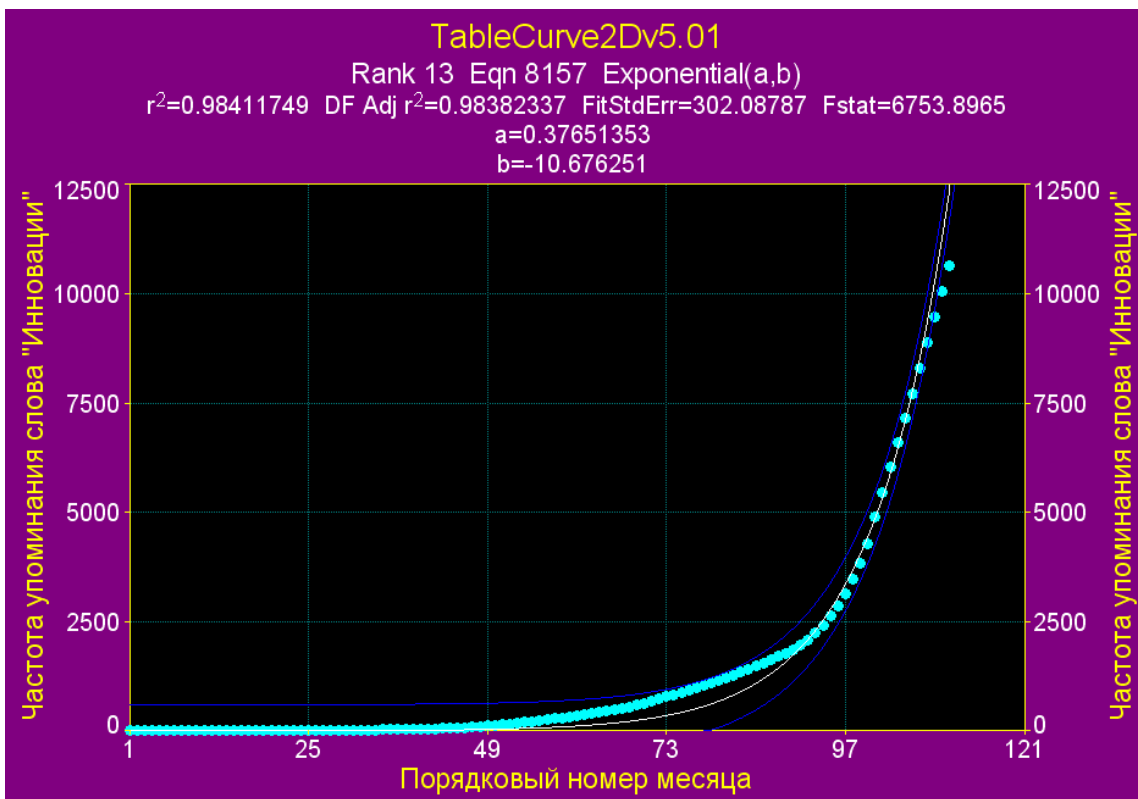
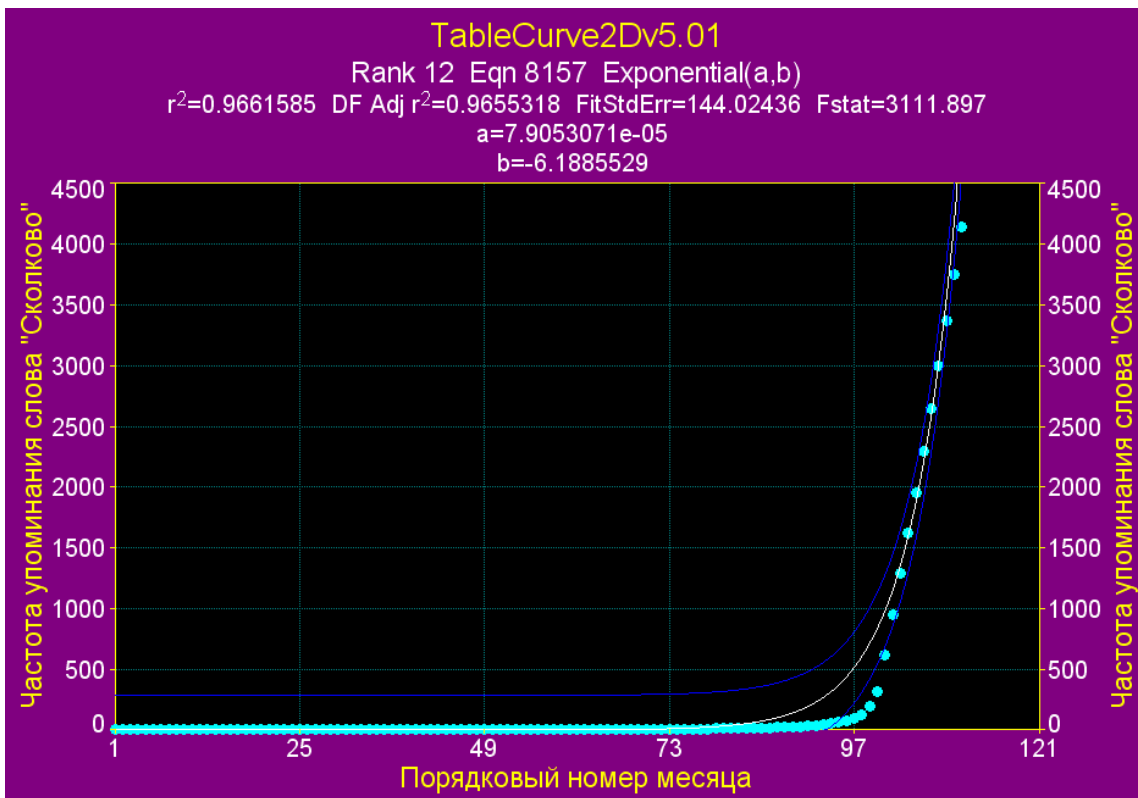


Рис.15

Тренд частоты упоминания слова «Сколково»



Из рис. 13-15 следует, что тренды можно хорошо аппроксимировать (мера аппроксимации $R^2 \geq 0.966$) экспоненциальной функцией. Таким образом, тренды динамики частоты упоминания слов «Модернизация», «Инновации» и «Сколково» можно описать известным дифференциальным уравнением (1).

$$\frac{dy}{dt} = ky, \quad (1)$$

где k - коэффициент пропорциональности

Решением дифференциального уравнения (1) является экспоненциальная функция, которая хорошо аппроксимирует тренды, представленные на рис. 13-15. Экспоненциальная функция описывает известный общесистемный закон естественного роста, когда скорость изменения величины y (1) прямо пропорциональна самой величине, т.е. в геометрической прогрессии. Коэффициент пропорциональности k (1) можно содержательно интерпретировать как среднее количество слов в комментариях блоггеров, инициированных упоминанием одного слова. В этой связи напомним, что экспоненциальный рост является одним из системных «архетипов» динамики (положительная обратная связь) в моделях системной динамики (SD) [цит. по 5], который реализован в многочисленных пакетах для системного моделирования.

Проведенный содержательный анализ комментариев блоггеров, по словам «Модернизация», «Инновации» и «Сколково» в различные моменты времени (см. рис.1-3), показал, что закон естественного роста отражает рост интереса блоггеров к развитию России, в частности, актуальности модернизации и инновационного развития России, проблемам модернизации различных подсистем российского общества (МВД, здравоохранение, пожарная служба и т.д.), актуальности реализации инновационного проекта «Сколково».

В целом, закон естественного роста для месячной динамики частоты упоминания слов «Модернизация», «Инновации» и «Сколково», хорошо объясняет полученные результаты, представленные в таблицах 1-2 и на рис. 1-15.

Выводы

Полученные результаты позволяют сделать следующий вывод:

В русскоязычной блогосфере за период 2001-2010 гг. месячная динамика частоты упоминания слов «Модернизация», «Инновации» и «Сколково» соответствует общесистемному закону естественного роста.

СПИСОК ЛИТЕРАТУРЫ

1. Давыдов А.А. Социология изучает блогосферу//Социолог. исслед. 2008, № 11, С. С.92-101. (<http://www.ecsocman.edu.ru/text/19144318>)
2. Аналитический доклад «Готово ли российское общество к модернизации». М.: Институт социологии РАН, 2010. (http://www.isras.ru/index.php?page_id=1413)
3. Аналитический доклад «Россияне о модернизации». М.: Исследовательская группа «ЦИРКОН», 2010. (<http://www.zircon.ru/>)
4. Давыдов А.А. Решения Президента РФ Д.Медведева о модернизации и инновационном развитии России: теория принятия решений. Официальный сайт РОС,2010. (http://www.ssa-rss.ru/index.php?page_id=22&id=53)
5. Давыдов А.А. Конкурентные преимущества системной социологии. (Электронное издание) М.: ИС РАН, 2008. (<http://www.isras.ru/publ.html?id=855>
<http://www.ecsocman.edu.ru/db/msg/324618.html>)
6. Давыдов А.А. Системная социология: введение в анализ динамики социума. М.: ЛКИ, 2007.
7. Обзорные научные статьи в блоге А.Давыдова на официальном сайте Института социологии РАН. (http://www.isras.ru/blog_Davydov.html)
8. «Пuls блогосферы» Яндекса. (<http://blogs.yandex.ru/pulse/>)
9. CaterpillarSSA 3.40. (<http://www.gistatgroup.com/gus/programs.html>)
- 10.Метод Singular Spectrum Analysis (SSA). (<http://www.gistatgroup.com/gus/books.html>)
11. Gretl (v.1.9.1). (<http://gretl.sourceforge.net/>)
12. TableCurve 2D (v.5.01). (<http://www.systat.com/products/tablecurve2d/>)
13. FracLab Toolbox (v.2.0). (<http://fraclab.saclay.inria.fr/homepage.html>)
14. Давыдов А.А. Вейвлет-анализ социальных процессов//Социолог.исслед. 2003, №11, С. 97-103. (<http://www.ecsocman.edu.ru/socis/msg/315095.html>)
15. Давыдов А.А. Системный подход в социологии: новые направления, теории и методы анализа социальных систем. М.: Эдиториал УРСС, 2005.

ПРИЛОЖЕНИЕ. Статистики для модели Vector Autoregression Analysis.

Эндогенные переменные – «Инновации» и «Сколково», Экзогенная переменная - «Модернизация»

VAR система, порядок лага 1

Метод оценки - МНК, наблюдения 2001:07-2010:08 (T = 110)

Лог. правдоподобие = 1454,6659

Определитель ковариационной матрицы = 6,5486477e-016

Крит. Акаике = -26,1212

Крит. Шварца = -25,6793

Крит. Хеннана-Куинна = -25,9420

Портмане-тест (Portmanteau): LB(27) = 439,025, Ст. свободы = 234 [0,0000]

Уравнение 1: Innovation (Инновации)

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	
const	-46,0047	77,2051	-0,5959	0,55255	
Innovation_1	0,502081	0,123588	4,0625	0,00009	***
Skolkovo_1	-0,102534	0,0816149	-1,2563	0,21182	
Modernization_1	-0,131423	0,0562962	-2,3345	0,02149	**
Modernization	0,428436	0,0404111	10,6019	<0,00001	***
time	1,59132	1,64821	0,9655	0,33654	

Среднее зав. перемен	1288,909	Ст. откл. зав. перемен	2411,561
Сумма кв. остатков	12163345	Ст. ошибка модели	341,9872
R-квадрат	0,980812	Испр. R-квадрат	0,979889
F(5, 104)	1063,210	P-значение (F)	1,48e-87
Параметр rho	-0,002342	Стат. Дарбина-Вотсона	1,953107

F-тесты для нулевых ограничений:

Все лаги для Innovation F(1, 104) = 16,504 [0,0001]

Все лаги для Skolkovo F(1, 104) = 1,5783 [0,2118]

Все лаги для Modernization F(1, 104) = 5,4499 [0,0215]

Уравнение 2: Skolkovo (Сколково)

	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	
const	92,2333	82,292	1,1208	0,26495	
Innovation_1	0,0234023	0,131731	0,1777	0,85934	
Skolkovo_1	0,760923	0,0869923	8,7470	<0,00001	***
Modernization_1	-0,0947066	0,0600054	-1,5783	0,11753	
Modernization	0,158256	0,0430737	3,6741	0,00038	***
time	-3,37523	1,7568	-1,9212	0,05744	*

Среднее зав. перемен	214,3455	Ст. откл. зав. перемен	933,2390
Сумма кв. остатков	13818982	Ст. ошибка модели	364,5200
R-квадрат	0,854433	Испр. R-квадрат	0,847434
F(5, 104)	122,0893	P-значение (F)	6,99e-42
Параметр rho	0,070467	Стат. Дарбина-Вотсона	1,778810

F-тесты для нулевых ограничений:

Все лаги для Innovation $F(1, 104) = 0,031561 [0,8593]$

Все лаги для Skolkovo $F(1, 104) = 76,51 [0,0000]$

Все лаги для Modernization $F(1, 104) = 2,491 [0,1175]$

Остатки модели Vector Autoregression Analysis (VAR)

