

УДК 338.27

doi:10.18720/SPBPU/2/id24-143

Заграновская Анна Васильевна,
доцент, канд. экон. наук, доцент

**ПОСТРОЕНИЕ МОДЕЛИ ПРОГНОЗИРОВАНИЯ
СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ
НА ОСНОВЕ КОНЦЕПТОВ, ВЫЯВЛЕННЫХ ПО ДАННЫМ СМИ**

Россия, Санкт-Петербург,
Санкт-Петербургский государственный экономический университет,
zagranet@rambler.ru

Аннотация. В статье проверяется гипотеза о влиянии публикаций в СМИ на важнейшие социально-экономические показатели. С этой целью методами тематического моделирования выявляются 9 тем, обсуждаемых в российских СМИ в течение 20 лет, с 2002 по 2021 гг. Они представляются в числовом виде с помощью модели Doc2Vec. Далее проверяется возможная причинно-следственная связь между выявленными темами и социально-экономическими показателями. По результатам строятся регрессионные модели зависимости социально-экономических показателей от обсуждаемых в СМИ тем.

Ключевые слова: тематическое моделирование, модель Doc2Vec, выявление причинно-следственной связи, регрессионный анализ, системный анализ.

Anna V. Zagranovskaia,
Candidate of Economic Sciences (PhD), Associate Professor

BUILDING A MODEL FOR FORECASTING SOCIO-ECONOMIC INDICATORS BASED ON CONCEPTS IDENTIFIED FROM MEDIA DATA

St. Petersburg State University of Economics, St. Petersburg, Russia,
zagranet@rambler.ru

Abstract. The article tests the hypothesis about the influence of media publications on the most important socio-economic indicators. For this purpose, topic modeling methods are used to identify 9 topics discussed in the Russian media over 20 years, from 2002 to 2021. They are represented numerically using the Doc2Vec model. Next, the possible cause-and-effect relationship between the identified themes and socio-economic indicators is tested. Based on the results, regression models of the dependence of socio-economic indicators on topics discussed in the media are built.

Key words: topic modeling, Doc2Vec model, causality identification, regression analysis, system analysis.

Введение

Средства массовой информации являются индикатором волнующих общество проблем и в то же время катализатором происходящих изменений. Это всем кажется очевидным, тем не менее, до сих пор формулируется, как гипотеза, которую нужно проверить на фактах. Особенно полезной может быть формализация влияния СМИ на социально-экономические показатели, что позволит строить прогнозы, на основе которых будет возможность управлять ситуацией.

В данной статье проверяется гипотеза о влиянии публикаций в СМИ на важнейшие социально-экономические показатели и по результатам строятся соответствующие прогнозные модели.

1. Данные для построения модели

В статье [2] в ходе применения динамической тематической модели за период с 2002 по 2021 гг. были выявлены следующие ключевые темы в области стратегического управления в России:

- Тема 0 — Проекты России.
- Тема 1 — Развитие России: вопросы сотрудничества и управления.
- Тема 2 — Россия и другие страны: региональные проблемы.
- Тема 3 — Негативное мнение населения по поводу недостаточного использования страной своих возможностей.

- Тема 4 — Изменение территории и СМИ.
- Тема 5 — Расчеты в рублях.
- Тема 6 — Рынок и образование.
- Тема 7 — Проблемы российских компаний.
- Тема 8 — Инвестиции в города, повышение доходов населения.

Там же показано, что на основе модели Doc2Vec выявленные темы можно представить в виде векторов заданной пользователем размерности [6], что позволяет применить к ним статистические методы анализа. Так каждая из 9 тем была представлена 20 числами, каждое из которых соответствует временному интервалу с 2002 по 2021 гг. (рис. 1).

Выявленные темы можно рассматривать, как концепты, отражающие сущность системы стратегического управления в России.

Здравый смысл подсказывает, что СМИ обладают властью, т. к. широко освещают наиболее важные для общества вопросы, тем самым, они задают вектор развития экономической системы, что проявляется в изменении социально-экономических показателей. Проверим эту гипотезу.

| | Тема_0 | Тема_1 | Тема_2 | Тема_3 | Тема_4 | Тема_5 | Тема_6 | Тема_7 | Тема_8 |
|------|----------|-----------|----------|----------|-----------|-----------|-----------|-----------|-----------|
| Годы | | | | | | | | | |
| 2002 | 0.210630 | -0.050016 | 0.428713 | 0.496185 | 0.253768 | -0.351345 | 0.182328 | -0.179534 | 0.012426 |
| 2003 | 0.124345 | -0.050657 | 0.077547 | 0.455622 | -0.130650 | -0.365624 | 0.463357 | 0.141498 | 0.046001 |
| 2004 | 0.303220 | 0.541048 | 0.242314 | 0.322826 | 0.050103 | 0.049970 | -0.060376 | -0.139278 | 0.032269 |
| 2005 | 0.227605 | 0.586174 | 0.327464 | 0.014547 | -0.060888 | -0.109682 | 0.211177 | -0.251125 | 0.292946 |
| 2006 | 0.565804 | -0.016194 | 0.073437 | 0.512683 | 0.419634 | -0.267247 | 0.042331 | -0.206973 | -0.108675 |

Рис. 1. Первые 5 строк скалярного представления тем динамической тематической модели

Из Российского статистического ежегодника [5] были отобраны следующие социально-экономические показатели России за период с 2002 по 2021 гг.:

- ИНВ — инвестиции в основной капитал в фактически действовавших ценах, млн руб.
- ОПЖ — ожидаемая продолжительность жизни при рождении, лет.
- ЗП — среднедушевые денежные доходы населения (в месяц), руб.

- ВВП — валовой внутренний продукт, млрд руб.
- ГРУЗЫ — перевозки грузов всеми видами транспорта, млн т.

Выбор такого набора показателей объясняется желанием осветить, по возможности, все системные элементы из матрицы системных характеристик (вход, выход, компоненты процессора: оснащение, последовательность, субъект труда и катализатор), которая выступает конструктивным определением системы и обеспечивает охват всех ее сфер деятельности [3].

2. Выявление причинно-следственной связи между темами и социально-экономическими показателями

При построении модели регрессии желательно убедиться в том, что распределение резульативного признака несущественно отличается от нормального. Поэтому ко всем социально-экономическим показателям был применен тест Шапиро-Уилка, который не обнаружил статистически значимых (на уровне 0.05) различий распределений признаков от нормального вида (рис. 2).

| | W | pval | normal |
|--------------|----------|----------|--------|
| ОПЖ | 0.932853 | 0.175196 | True |
| ИНВ | 0.961514 | 0.574522 | True |
| ЗП | 0.947207 | 0.326680 | True |
| ВВП | 0.950772 | 0.378955 | True |
| ГРУЗЫ | 0.939822 | 0.237934 | True |

Рис. 2. Проверка нормальности распределений социально-экономических показателей на основе критерия Шапиро-Уилка

Стоит также отметить, что распределения тем тоже не отличаются от нормального.

Предположим, что влияние тем на социально-экономические показатели, если оно наблюдается, может происходить с лагом во времени. Это проверяется с помощью теста Грэнджера [4].

Прежде всего, определим максимальный лаг для показателей. Это 1/4 от количества наблюдений, т. е. 5 [1]. Далее применим тест Грэнджера. Осознав его результаты, можно сделать выводы, представленные на рисунке 3.

показатель Тема_0 влияет на показатель ОПЖ
 показатель Тема_2 влияет на показатель ОПЖ
 показатель Тема_6 влияет на показатель ОПЖ
 показатель Тема_8 влияет на показатель ОПЖ
 показатель ВВП влияет на показатель ИНВ
 показатель ГРУЗЫ влияет на показатель ИНВ
 показатель Тема_0 влияет на показатель ИНВ
 показатель Тема_2 влияет на показатель ИНВ
 показатель Тема_3 влияет на показатель ИНВ
 показатель Тема_6 влияет на показатель ИНВ
 показатель Тема_7 влияет на показатель ИНВ
 показатель ГРУЗЫ влияет на показатель ЗП
 показатель Тема_0 влияет на показатель ЗП
 показатель Тема_7 влияет на показатель ЗП
 показатель Тема_8 влияет на показатель ЗП
 показатель ЗП влияет на показатель ВВП
 показатель Тема_3 влияет на показатель ВВП
 показатель Тема_7 влияет на показатель ВВП
 показатель Тема_2 влияет на показатель ГРУЗЫ
 показатель Тема_3 влияет на показатель Тема_0
 показатель ИНВ влияет на показатель Тема_1
 показатель ВВП влияет на показатель Тема_1
 показатель ГРУЗЫ влияет на показатель Тема_1
 показатель Тема_0 влияет на показатель Тема_1
 показатель Тема_3 влияет на показатель Тема_1
 показатель Тема_0 влияет на показатель Тема_2
 показатель Тема_4 влияет на показатель Тема_2
 показатель Тема_8 влияет на показатель Тема_2
 показатель ЗП влияет на показатель Тема_3
 показатель Тема_2 влияет на показатель Тема_3
 показатель Тема_0 влияет на показатель Тема_4
 показатель Тема_0 влияет на показатель Тема_5
 показатель Тема_6 влияет на показатель Тема_5
 показатель Тема_1 влияет на показатель Тема_6
 показатель ГРУЗЫ влияет на показатель Тема_7
 показатель Тема_8 влияет на показатель Тема_7

Рис. 3. Выводы о возможной причинно-следственной связи между показателями на основе результатов теста Грэнджера

Из рисунка 3 видно, что наша гипотеза подтвердилась, а именно темы, освещаемые в СМИ (Тема_0, Тема_2, Тема_3, Тема_6, Тема_7), влияют на социально-экономические показатели (ОПЖ, ЗП, ИНВ, ВВП, ГРУЗЫ). Кроме того, видно, что социально-экономические показатели (ЗП, ИНВ, ВВП, ГРУЗЫ) влияют на темы (Тема_1, Тема_3, Тема_7) и темы оказывают друг на друга влияние, что подтверждает результаты предыдущих исследований [2].

3. Прогнозная модель

Построим набор моделей линейной регрессии, описывающих зависимость значений социально-экономических показателей от тем, освещаемых в СМИ. При этом будем опираться на результаты теста Грэнджера.

Начнем с показателя ожидаемой продолжительности жизни при рождении (ОПЖ). Наилучшей оказалась модель зависимости ОПЖ от содержимого темы 6 (Рынок и образование) в текущем и предыдущем периоде (Тема_6(t), Тема_6(t-1)), что подтверждается рисунком 4.

| OLS Regression Results | | | | | | |
|------------------------|------------------|------------------------------|----------|-------|--------|---------|
| Dep. Variable: | y | R-squared (uncentered): | 0.936 | | | |
| Model: | OLS | Adj. R-squared (uncentered): | 0.927 | | | |
| Method: | Least Squares | F-statistic: | 95.84 | | | |
| Date: | Wed, 06 Sep 2023 | Prob (F-statistic): | 1.65e-08 | | | |
| Time: | 14:25:39 | Log-Likelihood: | -64.469 | | | |
| No. Observations: | 15 | AIC: | 132.9 | | | |
| Df Residuals: | 13 | BIC: | 134.4 | | | |
| Df Model: | 2 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| x1 | 96.4122 | 19.521 | 4.939 | 0.000 | 54.240 | 138.584 |
| x2 | 76.9787 | 19.853 | 3.877 | 0.002 | 34.088 | 119.869 |
| Omnibus: | 0.327 | Durbin-Watson: | 1.304 | | | |
| Prob(Omnibus): | 0.849 | Jarque-Bera (JB): | 0.471 | | | |
| Skew: | 0.137 | Prob(JB): | 0.790 | | | |
| Kurtosis: | 2.176 | Cond. No. | 2.98 | | | |

Рис. 4. Сводка по модели линейной регрессии, описывающей зависимость ОПЖ от переменных Тема_6(t), Тема_6(t-1)

Статистика F позволяет проверить гипотезу о равенстве нулю коэффициентов при всех регрессорах, т. е. проверить значимость уравнения регрессии. Из рисунка 4 видно, что для F-statistic $P\text{-value} < 0.01$, что говорит о значимом отличии от нуля всех коэффициентов модели регрессии на уровне 0.01.

Коэффициент детерминации R-squared — это доля вариации Y, объясняемая с помощью модели. Он показывает качество подгонки регрессионной модели к наблюдаемым значениям. В нашем случае $R\text{-squared} = 0.936$, что свидетельствует о высоком качестве модели регрессии. t-статистика проверяет гипотезу о том, что соответствующий

коэффициент в регрессии равен нулю. В нашем случае коэффициенты при переменных Тема_6(t), Тема_6(t-1) значимо отличаются от нуля на уровне 0.01.

Далее рассмотрим показатель среднедушевых денежных доходов населения (ЗП). Наилучшей оказалась модель зависимости ЗП от содержимого темы 0 (Проекты России) с лагами 1–4 (Тема_0(t-1), Тема_0(t-2), Тема_0(t-3), Тема_0(t-4)), от содержимого темы 7 (Проблемы российских компаний) с лагом 1 (Тема_7(t-1)), а также от показателя перевозки грузов всеми видами транспорта с лагами 0, 1, 3, 5 (ГРУЗЫ, ГРУЗЫ(t-1), ГРУЗЫ(t-3), ГРУЗЫ(t-5)). Это подтверждается рисунком 5.

| OLS Regression Results | | | | | | |
|------------------------|------------------|------------------------------|----------|-------|----------|----------|
| ===== | | | | | | |
| Dep. Variable: | y | R-squared (uncentered): | 1.000 | | | |
| Model: | OLS | Adj. R-squared (uncentered): | 0.999 | | | |
| Method: | Least Squares | F-statistic: | 1614. | | | |
| Date: | Wed, 06 Sep 2023 | Prob (F-statistic): | 1.89e-09 | | | |
| Time: | 15:24:38 | Log-Likelihood: | -116.34 | | | |
| No. Observations: | 15 | AIC: | 250.7 | | | |
| Df Residuals: | 6 | BIC: | 257.1 | | | |
| Df Model: | 9 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| ===== | | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| ----- | | | | | | |
| x1 | 1.758e+04 | 1246.770 | 14.098 | 0.000 | 1.45e+04 | 2.06e+04 |
| x2 | 2.904e+04 | 2267.841 | 12.805 | 0.000 | 2.35e+04 | 3.46e+04 |
| x3 | 2.41e+04 | 1460.211 | 16.506 | 0.000 | 2.05e+04 | 2.77e+04 |
| x4 | 1.227e+04 | 1215.068 | 10.098 | 0.000 | 9296.574 | 1.52e+04 |
| x5 | 4087.8995 | 1738.635 | 2.351 | 0.057 | -166.388 | 8342.187 |
| x6 | -3.1225 | 0.845 | -3.694 | 0.010 | -5.191 | -1.054 |
| x7 | -3.2609 | 0.537 | -6.075 | 0.001 | -4.574 | -1.947 |
| x8 | 1.2482 | 0.367 | 3.398 | 0.015 | 0.349 | 2.147 |
| x9 | 4.4026 | 0.560 | 7.869 | 0.000 | 3.034 | 5.772 |
| ===== | | | | | | |
| Omnibus: | 1.779 | Durbin-Watson: | 2.018 | | | |
| Prob(Omnibus): | 0.411 | Jarque-Bera (JB): | 0.978 | | | |
| Skew: | 0.621 | Prob(JB): | 0.613 | | | |
| Kurtosis: | 2.852 | Cond. No. | 2.02e+05 | | | |
| ===== | | | | | | |

Рис. 5. Сводка по модели линейной регрессии, описывающей зависимость ЗП от переменных Тема_0(t-1), Тема_0(t-2), Тема_0(t-3), Тема_0(t-4), Тема_7(t-1), ГРУЗЫ, ГРУЗЫ(t-1), ГРУЗЫ(t-3), ГРУЗЫ(t-5)

Из рисунка 5 видно, что модель и коэффициенты значимы на уровне 0.05.

Теперь обратимся к показателю инвестиций в основной капитал в фактически действовавших ценах (ИНВ). Наилучшей оказалась модель зависимости ИНВ от показателей ВВП и ГРУЗЫ (рис. 6).

| OLS Regression Results | | | | | | |
|------------------------|------------------|------------------------------|----------|-------|---------|---------|
| Dep. Variable: | y | R-squared (uncentered): | 0.999 | | | |
| Model: | OLS | Adj. R-squared (uncentered): | 0.998 | | | |
| Method: | Least Squares | F-statistic: | 4631. | | | |
| Date: | Wed, 06 Sep 2023 | Prob (F-statistic): | 2.84e-19 | | | |
| Time: | 15:11:26 | Log-Likelihood: | -219.51 | | | |
| No. Observations: | 15 | AIC: | 443.0 | | | |
| Df Residuals: | 13 | BIC: | 444.4 | | | |
| Df Model: | 2 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| x1 | 163.1010 | 5.058 | 32.245 | 0.000 | 152.174 | 174.028 |
| x2 | 166.8447 | 49.970 | 3.339 | 0.005 | 58.890 | 274.799 |
| Omnibus: | 5.448 | Durbin-Watson: | 1.266 | | | |
| Prob(Omnibus): | 0.066 | Jarque-Bera (JB): | 2.534 | | | |
| Skew: | 0.807 | Prob(JB): | 0.282 | | | |
| Kurtosis: | 4.205 | Cond. No. | 27.1 | | | |

Рис. 6. Сводка по модели линейной регрессии, описывающей зависимость ИНВ от переменных ВВП и ГРУЗЫ

Из рисунка 6 видно, что модель и коэффициенты значимы на уровне 0.01.

Зависимость ИНВ от тем (Тема_0, Тема_2, Тема_3, Тема_6, Тема_7) с различными лагами прослеживается, но коэффициенты детерминации в тех моделях меньше, чем в представленной на рисунке 6 модели.

Посмотрим на важнейший экономический показатель валового внутреннего продукта (ВВП). Лучшей оказалась модель зависимости ВВП от содержимого темы 3 (Негативное мнение населения по поводу недостаточного использования страной своих возможностей) с лагом 5 (Тема_3(t-5)) и от значения следующих показателей: ОПЖ(t-5), ИНВ, ИНВ(t-4), ЗП, ЗП(t-1). Это подтверждается рисунком 7.

Из рисунка 7 видно, что модель и коэффициенты значимы на уровне 0.05.

Наконец, рассмотрим показатель перевозки грузов всеми видами транспорта (ГРУЗЫ). Наилучшей оказалась модель зависимости показателя ГРУЗЫ от содержимого темы 2 (Россия и другие страны: региональные проблемы) в текущем периоде и с лагом 3 (Тема_2, Тема_2(t-3)), что подтверждается рисунком 8.

Из рисунка 8 видно, что модель и коэффициенты значимы на уровне 0.01.

OLS Regression Results

```

=====
Dep. Variable:          y      R-squared (uncentered):      1.000
Model:                 OLS    Adj. R-squared (uncentered):  1.000
Method:                Least Squares  F-statistic:                 7692.
Date:                  Wed, 06 Sep 2023  Prob (F-statistic):         3.61e-16
Time:                  15:46:15   Log-Likelihood:             -126.90
No. Observations:     15      AIC:                        265.8
Df Residuals:         9       BIC:                        270.1
Df Model:              6
Covariance Type:      nonrobust
=====

```

| | coef | std err | t | P> t | [0.025 | 0.975] |
|----|-----------|----------|--------|-------|----------|----------|
| x1 | 6765.2096 | 1915.566 | 3.532 | 0.006 | 2431.899 | 1.11e+04 |
| x2 | -176.0107 | 53.861 | -3.268 | 0.010 | -297.853 | -54.168 |
| x3 | 0.0032 | 0.000 | 7.619 | 0.000 | 0.002 | 0.004 |
| x4 | 0.0017 | 0.001 | 2.640 | 0.027 | 0.000 | 0.003 |
| x5 | 3.1049 | 0.605 | 5.129 | 0.001 | 1.735 | 4.474 |
| x6 | -2.2656 | 0.568 | -3.986 | 0.003 | -3.552 | -0.980 |

```

=====
Omnibus:                0.447   Durbin-Watson:            2.034
Prob(Omnibus):          0.800   Jarque-Bera (JB):         0.526
Skew:                   -0.310  Prob(JB):                 0.769
Kurtosis:               2.325   Cond. No.                 9.03e+07
=====

```

Рис. 7. Сводка по модели линейной регрессии, описывающей зависимость ВВП от переменных Тема_3(t-5), ОПЖ(t-5), ИНВ, ИНВ(t-4), ЗП, ЗП(t-1)

OLS Regression Results

```

=====
Dep. Variable:          y      R-squared (uncentered):      0.794
Model:                 OLS    Adj. R-squared (uncentered):  0.762
Method:                Least Squares  F-statistic:                 25.05
Date:                  Wed, 06 Sep 2023  Prob (F-statistic):         3.47e-05
Time:                  15:27:31   Log-Likelihood:             -144.77
No. Observations:     15      AIC:                        293.5
Df Residuals:         13      BIC:                        294.9
Df Model:              2
Covariance Type:      nonrobust
=====

```

| | coef | std err | t | P> t | [0.025 | 0.975] |
|----|-----------|----------|-------|-------|----------|----------|
| x1 | 9880.0919 | 3168.866 | 3.118 | 0.008 | 3034.173 | 1.67e+04 |
| x2 | 1.093e+04 | 3461.755 | 3.157 | 0.008 | 3450.110 | 1.84e+04 |

```

=====
Omnibus:                2.466   Durbin-Watson:            0.971
Prob(Omnibus):          0.291   Jarque-Bera (JB):         1.259
Skew:                   0.363   Prob(JB):                 0.533
Kurtosis:               1.780   Cond. No.                 2.04
=====

```

Рис. 8. Сводка по модели линейной регрессии, описывающей зависимость показателя ГРУЗЫ от переменных Тема_2, Тема_2(t-3)

Заключение

Проведенный анализ позволил подтвердить гипотезу о влиянии публикаций в СМИ на важнейшие социально-экономические показатели. В частности, были получены следующие статистически значимые модели с высоким уровнем коэффициента детерминации:

1. Ожидаемая продолжительность жизни при рождении (ОПЖ) зависит от содержимого темы 6 (Рынок и образование) в текущем и предыдущем периоде.

2. Среднедушевые денежные доходы населения (ЗП) зависят от содержимого темы 0 (Проекты России), взятой с лагами 1–4, от содержимого темы 7 (Проблемы российских компаний), взятой с лагом 1, а также от показателя перевозки грузов всеми видами транспорта (ГРУЗЫ) с лагами 0, 1, 3, 5.

3. Инвестиции в основной капитал в фактически действовавших ценах (ИНВ) зависят от показателей ВВП и ГРУЗЫ. Также была установлена менее точная зависимость показателя ИНВ от тем 0, 2, 3, 6, 7 с различными лагами.

4. Показатель валового внутреннего продукта (ВВП) зависит от содержимого темы 3 (Негативное мнение населения по поводу недостаточного использования страной своих возможностей) с лагом 5 и от значения следующих показателей: ОПЖ($t-5$), ИНВ, ИНВ($t-4$), ЗП, ЗП($t-1$).

5. Показатель перевозки грузов всеми видами транспорта (ГРУЗЫ) зависит от содержимого темы 2 (Россия и другие страны: региональные проблемы) в текущем периоде и с лагом 3.

Дальнейшее развитие данной темы представляется в проверке точности прогнозирования по выявленным моделям. Это потребует сбора и подготовки к моделированию новых данных, например, за 2022 год.

Список литературы

1. Дуброва Т.А. Статистические методы прогнозирования: учеб. пособие для вузов. – М.: ЮНИТИ-ДАНА, 2003. – 206 с.

2. Заграновская А. В. Стратегическое управление в России: автоматизированное построение причинно-следственной диаграммы и выявление ключевых тем по данным СМИ // Прикладная информатика. – 2023. – Т. 18. – № 2. – С. 73–84. – DOI: 10.37791/2687-0649-2023-18-2-73-84.

3. Заграновская А.В., Эйсснер Ю.Н. Теория систем и системный анализ в экономике: учеб. пособие для вузов. – М.: Издательство Юрайт, 2020. – 266 с.

4. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: учебник – 6-е изд., перераб. и доп. – М.: Дело, 2004. – 576 с.

5. Российский статистический ежегодник. 2022: Стат.сб./Росстат. – М., 2022 – 691 с.

6. Quoc Le, Tomas Mikolov. Distributed Representations of Sentences and Documents // Proceedings of the 31st International Conference on Machine Learning. PMLR. – 2014. – Vol. 32(2). – Pp. 1188–1196.