HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

M. Zhokhov, A. Galeev

REFORMING INTERNATIONAL TRADE POLICY: NLP-BASED CONTENT ANALYSIS OF THE RECENT HISTORY OF DEBATES

Working Paper WP11/2022/01
Series WP11
Economic Reforms of the Late 20th Century:
Experience and Lessons
from Contemporary History

Moscow
2022

UDC 339.5.01
LBK 65.02
Z10

Editors of the Series WP11
"Economic Reforms of the Late 20th Century:
Experience and Lessons from Contemporary History"
Vladimir Avtonomov, Oleg Ananyin

Zhokhov, M., Galeev, A.

Reforming international trade policy: NLP-based content analysis of the recent history of debates [Electronic resource] : Working paper WP11/2022/01 / M. Zhokhov, A. Galeev ; National Research University Higher School of Economics. – Electronic text data (340 Kb). – Moscow: HSE Publishing House, 2022. – (Series WP11 "Economic Reforms of the Late 20th Century: Experience and Lessons from Contemporary History"). – 14 p.

The paper analyses the development of scientific discourse in the field of international trade reforms and its impact on the implemented policy. We construct a database of articles in international trade indexed in the Scopus scientific citation database and compile an index reflecting the article's semantic orientation in terms of adherence to either economic liberalism or economic nationalism using NLP algorithms. The comparison of the index dynamics with the dynamics of the aggregated number of legislative acts in the Trade Monitoring Database suggests a direct relationship between political measures and the first lag of the "liberalism-protectionism" index from 2008 to 2021.

UDC 339.5.01
LBK 65.02

Keywords: NLP, content analysis, discourse analysis, international trade, economic liberalism, economic nationalism.

Mikhail Zhokhov – Research Assistant at the Centre for the History and Methodology of Economic Science, HSE University, New Economic School.

Anton Galeev – Research Fellow of the Centre for the History and Methodology of Economic Science, HSE University.

1. Introduction

More than 30 years have passed since the creation of the Washington Consensus, a set of policy prescriptions developed by the IMF and the World Bank. The prescriptions were designed to ensure economic growth in developing countries. The Consensus comprised policies promoting free-market doctrine and included the liberalisation of international trade [Stiglitz, Schoenfelder, 2003]. However, no real consensus has been reached on the issue of international trade policy yet [McDermott, 1999]. The discourse on this topic, both in academic literature and among politicians, represents a confrontation between the doctrines of protectionism (economic nationalism) and economic liberalism (globalism).

This paper analyses the development of scientific discourse in the field of international trade and its impact on the implemented policy. Firstly, we study the clash of the liberal and protectionist doctrines in academic articles. Secondly, we analyse the relationship between the discourse on international trade (theory) and trade policies (practice).

The contemporary works that study discussions between supporters of protectionism and globalism focus on the qualitative analysis of arguments and theories. Such an approach significantly constrains the discourse analysis over a lengthy period. Alternatively, the quantitative methodology allows analysis of aggregated information about scientific discourse at a "high level". In this paper, a quantitative approach is applied to content analysis [Gentzkow, Kelly, Taddy, 2019]. Nowadays, content analysis is used to solve both practical (for example, building the sentiment index of perception of Fed statements) [Banerjee et. al., 2021] and theoretical problems (for example, history of Islamic economic thought) [Handoko, 2020].

We perform content analysis of articles in the field of international trade. For this purpose, we prepare a database, which includes metadata of articles from the Scopus scientific citation database. Based on the compiled database, the index reflecting the article's adherence to one of the economic doctrines is compiled for each abstract using Natural Language Processing[1] tools. We focus only on analysis of abstracts since they contain the key information on the paper. Thus, we can infer a paper's adherence to the particular doctrine by studying its abstract alone. This approach allows us to draw conclusions about the global dynamics of the predominance of a particular doctrine in the academic community. Finally, we trace the relationship between the scientific discourse and the trade policy implementation using the data on the number of protectionist and liberal measures presented in the Trade Monitoring Database [WTO, 2022].

The paper has the following structure: section 2 describes the process of data collection and database preparation, section 3 demonstrates the construction of the semantic orientation index, section 4 studies the relation between the implemented policies and scientific discourse, section 5 contains several concluding remarks.

2. Data collection

The first stage of our work implied the construction of a database containing metadata of academic articles on international trade. For this purpose, we chose the Scopus bibliographic database

[1] Natural Language Processing (NLP) — a interdisciplinary approach in computer science, aimed at making computers "understand" the contents of nature language data (meaning not artificial texts: e.g. texts in English, Russian, etc.) [Bird, Loper, Klein, 2009, p. ix]

system. Scopus is associated with high-quality papers and includes a wide range of indexed journals (more than 25,100 titles) from all over the world [Elseveir, 2020]. Unlike Web of Science, Scopus provides a user-friendly API for data collection and is more efficient in article aggregation.

The article selection process consisted of multi-level filtering. Since we focus on the discourse in international trade, we had to pick the articles directly related to the field. To do so, firstly, we requested a search query in Scopus by the word "trade". Secondly, we limited the search results by the list of keywords characterising the article's research field as international trade (see Table 1).

Table 1. Keywords for data collection

1. ASEAN
2. Global Value Chains
3. OECD
4. Liberalisation
5. Bilateral Agreement
6. Regional Trade
7. Trade Union
8. Neoliberalism
9. Protectionism
10. Bilateral Agreement
11. Trade Openness
12. International Cooperation
13. Gravity Model
14. WTO
15. Tariff Structure
16. Economic Integration
17. Free Trade
18. Foreign Trade
19. Import
20. Trade Agreement
21. World Trade Organisation
22. Trade Flow
23. Globalisation
24. Trade Policy
25. Export
26. International Trade

Since our work is focused on the discourse in economics, we limited the articles to three subject areas — economics (ECON), social sciences (SOCI), and business (BUSI). We used the Python implementation of the Scopus API in the pybliometrics library [Rose, Kitchin, 2019] to collect the metadata of the selected articles. The final database contains information on 30,533 articles[2]. The metadata includes the names of the authors, their affiliation, article title, journal title, year of issue, issue number, volume of the journal, and the article abstract.

The distribution of articles in the final database through the years is presented below (see Figure 1). We observe a strong increasing trend from 1990 until present. The persistence of the trend can be attributed to either an increasing number of journals, indexed in Scopus, or an increasing interest in international trade.

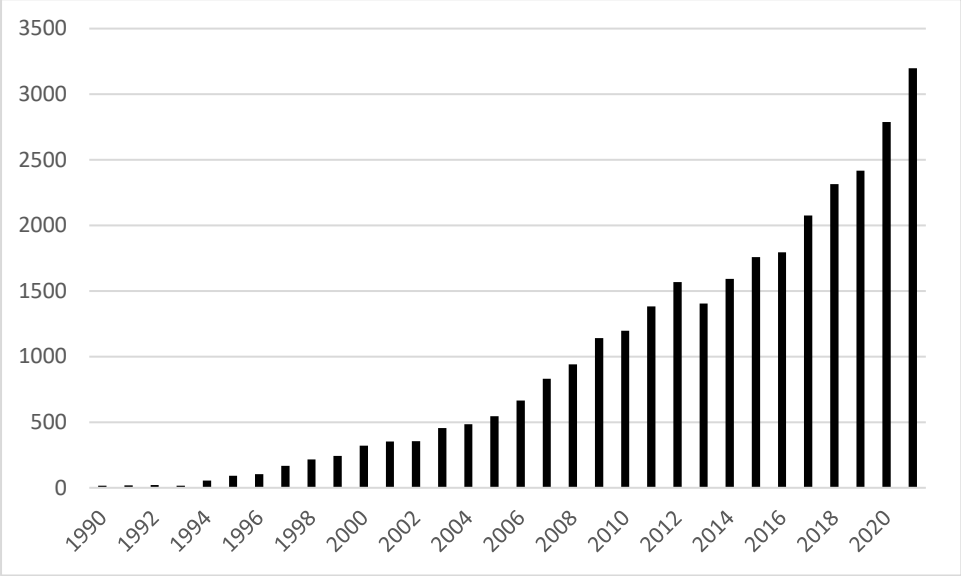

Figure 1. Number of published articles

3. Index construction

After the construction of the database, we proceeded to the analysis of semantic orientation. For this purpose, we focused specifically on articles' abstracts. The use of full texts is complicated by lack of open access papers and the technical complexity of their processing. On the one hand, the majority of journals in economics work on a subscription basis. The purchase of all the necessary subscriptions is impossible due to the financial constraints of our research. On the other hand, full texts are distributed as PDF-files in different templates. The variety of templates makes the task of text retrieval for a large number of articles impractically laborious. In addition, full texts contain a literature review, which does not represent the author's opinion and, thus, cloak the main idea of the paper. Unlike full texts, abstracts present key findings and the author's own views. Thus, the analysis of abstracts alone can provide the precise information of the semantic orientation of the paper.

The next step in our research was the preprocessing of abstracts using NLP tools. This procedure is standard for most works exploring the semantic orientation of speeches and texts in

[2] The data was downloaded from Scopus API between February 18 and 28, 2022 via http://api.elsevier.com and http://www.scopus.com.

economics. Firstly, the tokenisation of abstracts is performed using the Natural Language Toolkit (NLTK) module from Python [Bird, Loper, Klein, 2009]. Stop words and punctuation were excluded as interfering with further analysis. To determine stop words, we use a built-in list from NLTK. Secondly, we performed stemming — the reduction of words to their root form. We chose Lancaster Stemmer from NLTK for this task. Despite being aggressive, the algorithm is effective in reducing words to their initial forms [Paice, 1990]. Thus, we prepare a preprocessed corpus of abstracts, which allows us to search for the hits of keywords.

The analysis implies the construction of a "liberalism-protectionism" index. It reflects the semantic orientation of the analysed article in terms of adherence to either economic liberalism or economic nationalism. A similar approach to content analysis was presented in works studying the emotional semantic orientation of the text (the "negative-positive" scale) [Hutto, Gilbert, 2014] and the perception of central bank statements (the "hawkish-dovish" scale) [Banerjee et. al., 2021]. For our research we contrapose the doctrines and present a "liberalism-protectionism" scale embodied in the index.

To construct the index we form a dictionary of the words corresponding to either doctrine. Due to the lack of works on international trade discourse with the application of NLP methodology, we compiled the dictionary ourselves (see Table 2).

For the dictionary construction, we followed several rules described in the studies with a similar methodology. We took into account words with a high threshold of their normal form occurrences, particularly, more than 100 hits. Some words were not included in the list, since they are usually associated with non-economic information and, therefore, clog up the results of the index (e.g., "control"). We selected the words semantically oriented to either economic liberalism or economic nationalism in international trade. We paid special attention to verbs characterising authors' opinions on the political measures. In academic literature, authors abstain from using emotionally charged vocabulary, thus the verbs constitute the main source of semantic information.

For each article, we determined the number of hits of "liberal" and "protectionist" words (L_i and P_i, respectively). To calculate the index, we use the formula (1), based on the work [Banerjee et. al., 2021]. The formula presents the scaled difference3 in shares of "liberal" and "protectionist" words:

$$IL_i = \ln\left(\frac{L_i}{L_i + P_i} + 1\right) - \ln\left(\frac{P_i}{L_i + P_i} + 1\right) = \ln\left(\frac{2L_i + P_i}{2P_i + L_i}\right) \quad (1)$$

Accordingly, positive index values imply liberal orientation of the paper, while negative are associated with economic nationalism. The higher the index, the more we can classify the paper as liberal and vice versa.

[3] In this case, adding one into the logarithms in the formula helps to avoid infinite index values in the case of absence of keywords of one of the directions.

Table 2. List of stemmed words

Liberal words			Protectionist words		
Stemmed form	Hits	Example	Stemmed form	Hits	Example
op	8271	Open, openness	protect	4688	Protect, protectionism
coop	4155	Cooperate, cooperation	tax	2865	Tax, taxes
partn	3984	Partner, partnership	pref	2263	Preference, preferential
negoty	3164	Negotiations, negotiate	barry	2247	Barrier
neolib	2742	Neoliberal	restrict	1662	Restrict, restriction
facilit	1703	Facilitate, facilitation	clos	1616	Close
bargain	1213	Bargain	subsidy	1550	Subsidy
mitig	743	Mitigate, mitigation	interv	1110	Intervention, intervene
dereg	317	Deregulate	impos	879	Impose
allevy	230	Alleviate	oblig	554	Oblige
eas	193	Ease	bloc	409	Bloc
Total	26715		ban	392	Ban
			regul	232	Regulate, regulation
			prohibit	186	Prohibit, prohibition
			subsid	176	Subsidize
			confin	112	Confine
			Total	20941	

We also take into account the presence of negatives and "polar" words that change the contribution to the article's semantic orientation. We apply templates as seen in [Banerjee et. al., 2021].

The first structure is:

$$(Polar\ word/verb\ +\ n't/verb\ +\ not)\ +\ keyword$$

Here the keyword changes its meaning to the opposite due to the presence of the word coming before. To check the polarity of the previous word, we used the Python VADER library.

Within the library's dictionary, each word was assigned with a rating of its shade [Hutto, Gilbert, 2014].

We take into account only negative polar words, since the change in the meaning occurs only if there is a negative polar word before it. Positive polar words reinforce the meaning of the keyword, but do not change the full context. Therefore, we do not consider them in any special way. A total of 877 such negations were found.

The second structure implies the double negation, which nullifies the negative effect:

$$verb + not/n't + polar\ word + keyword$$

We did not find any examples of such structures in the collected database. We assume that this is due to the fact that the academics try to simplify their texts and avoid complex constructions, such as double negations. We refrain from analysing more complex negations, due to the absence of double negations in the data.

After calculating the index for each abstract, we aggregate them across time. The aggregated index reflects the orientation to either economic nationalism (negative index values) or economic liberalism (positive index values) of the academia in general. Namely, we average the index values for annual intervals (2), where N is the number of articles published during the selected period. The choice of the period is based on the following observations. Firstly, we account for the difference in periodicity of the journals. Scopus includes annual editions, semi-annual editions as well as quarterly and monthly journals. Thus, an annual period would reflect the total contribution of all possible editions. Secondly, the publication process is time-consuming, so annual intervals encompass the general direction of the discourse development. Finally, such periods allow tracing the relationship between the discourse and the implemented policies in international trade.

$$IL = \frac{\sum_1^N IL_i}{N} \tag{2}$$

We plot the calculated index for the period 1990-2021 (see Figure 2). The index significantly fluctuates from 1990 till the early 2000s and then shows stable growth. The fluctuations in the 1990s can be attributed to the small number of papers in the collected database — the number of papers does not exceed 100 before 1996 and is less than 300 before 2000.

In 1990, the "liberalism-protectionism" index is positive, thus indicating that liberalism was prevailing in the academia. However, in the subsequent years the index drops and remains negative. This implies that the academia switched its orientation and became significantly more favourable towards economic nationalism. Despite the fluctuations the prevalence of the protectionist doctrine continued until 2000. During the next 21 years the index remains positive, indicating economic liberalism as the dominant doctrine in the discourse. The persistent positive trend of the "liberalism-protectionism" index from 2000 till 2021 signifies the formation of consensus embodied in the liberal doctrine, its development and popularisation.

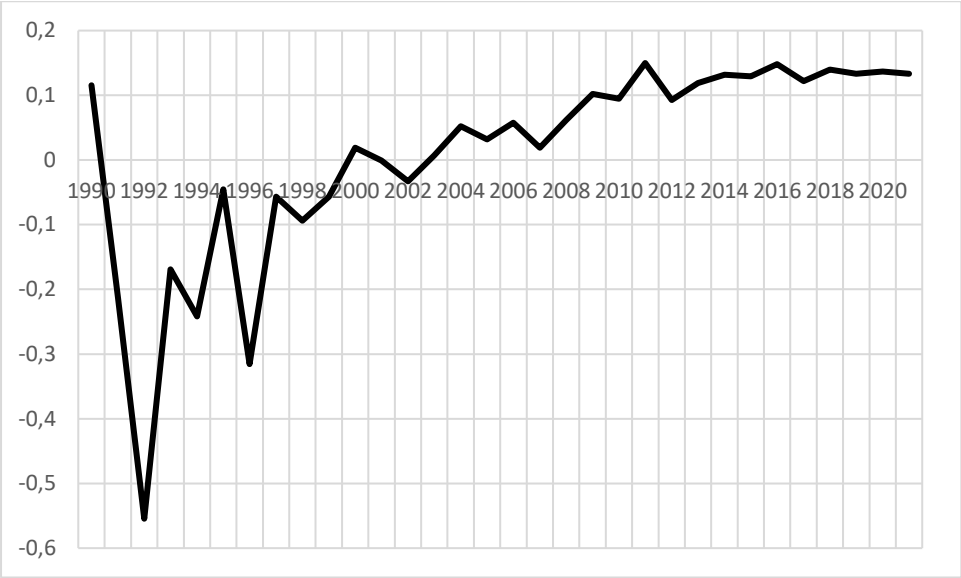

Figure 2. Dynamics of the index "liberalism-protectionism"

4. Comparison of Theory and Practice

To study the relationship between the discourse and the implemented policies, we use data on political measures from the Trade Monitoring Database [WTO, 2022]. We focus on measures regulating goods turnover between countries, since the international goods trade is of significant interest in academia while services in trade flows constitute only a small share [WTO, 2019].

The database provides classification of various legislative acts, labelling them as restrictive measures, facilitating measures or trade remedies. The available period from 2008 to 2021 provides information on 5,963 measures. While interpretation of restrictive and facilitating measures in international trade is quite straightforward, trade remedies present a puzzle in their adherence to either economic nationalism or economic liberalism. Trade remedies imply "actions taken in response to subsidies (countervailing duties), sales at less than fair value (antidumping) and import surges (safeguards)" [United States Trade Representative, 2022]. In this paper we consider these measures to be neutral and do not attribute them to any doctrine. In general, we observe an approximately equal amount of restrictive and facilitating measures for the period in question (see Figure 3).

Considering trade remedies as neutral, we focus on facilitating and restrictive acts as the implementation of protectionist or neoliberal doctrines respectively. The information on the policies has to be aggregated to be comparable with the "liberalism-protectionism" index. We suggest two approaches to the aggregation of data on political measures. Firstly, we calculate the share of net facilitating measures:

$$\overline{MI_t} = \frac{F_t - R_t}{F_t + R_t} \tag{3}$$

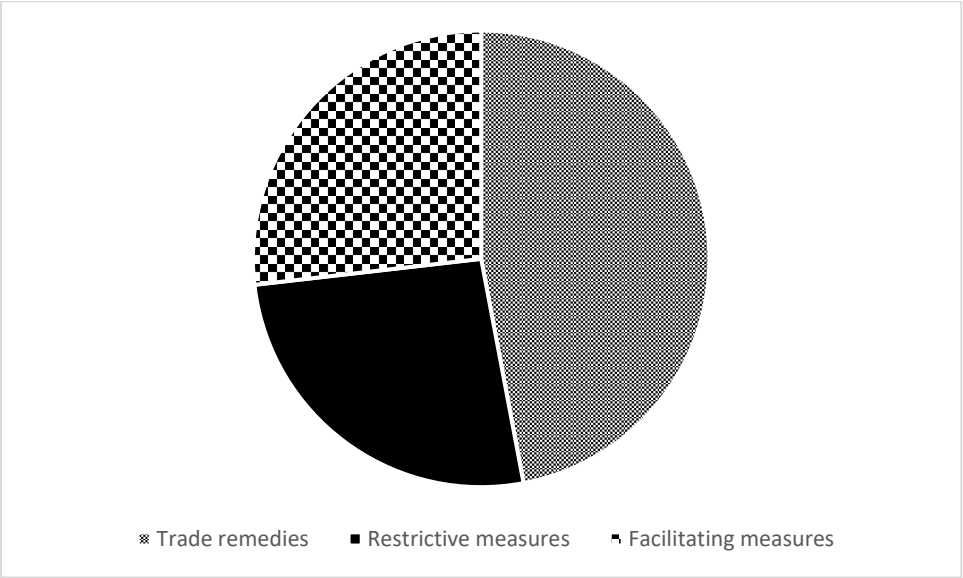

Figure 3. The share of the types of measures in the database, 2008–2021

Here F_t, R_t are the facilitating and restricting measures in year t respectively. The suggested approach to aggregation is intuitive; however, it lies within a scale different from the "liberalism-protectionism" index. Thus, secondly, we aggregate the data on political measures using the same formula as (1):

$$MI_t = \ln\left(\frac{F_t}{F_t + R_t} + 1\right) - \ln\left(\frac{R_t}{F_t + R_t} + 1\right) = \ln\left(\frac{2F_t + R_t}{2R_t + F_t}\right) \quad (4)$$

The constructed trade measures index is directly comparable with the "liberalism-protectionism" index.

We plot all the indices on one graph, with the "liberalism-protectionism" index is plotted along the auxiliary axis on the left (see Figure 4). The trade measures and share of net facilitating measures move alongside throughout the whole period. They demonstrate a significant decrease in 2009 as both indices hit all time lowest negative values. This implies the turn to the protectionist policies (probably due to the Great Recession). In the subsequent years the indices fluctuate and by 2014 they start following an increasing trend and take positive values indicating the liberalisation of international trade. However, the trend's direction changes to decreasing in 2017. We observe a dramatic fall in the policies indices in 2020, associated with the beginning of COVID-19 pandemic.

The "liberalism-protectionism" index stays positive for all the period in question, which indicated the prevalence of liberal doctrine in the academic literature. Despite a significant decrease in 2012 the index demonstrates a steady increasing trend unlike trade measures and share of net facilitating measures.

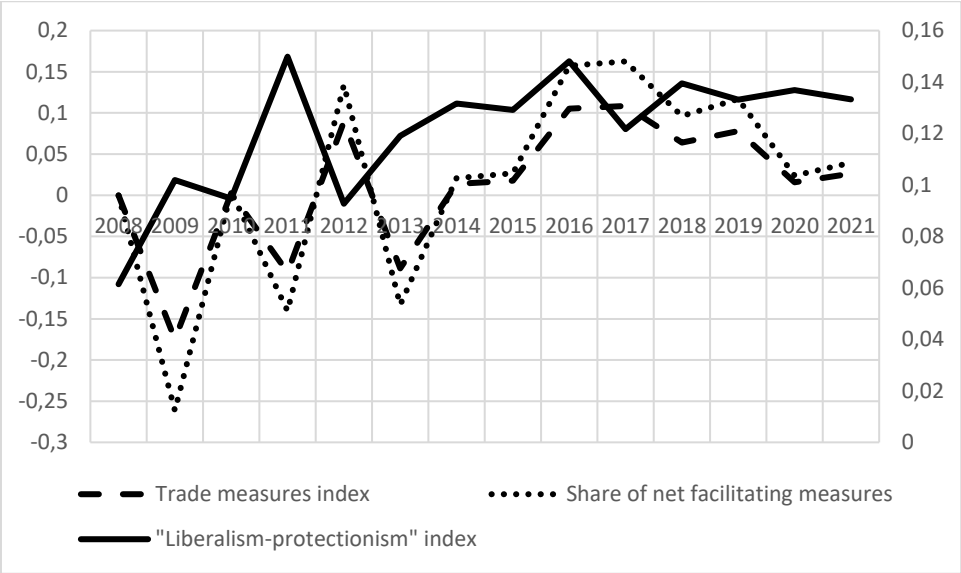

Figure 4. Discourse and implemented policies dynamics

The primary analysis of the graph suggests a correlation between the variables related to the policy and the discourse index. To conclude whether the variables are directly related or there is an intertemporal relationship, we calculate the correlation matrix. The matrix demonstrates the relationship between the "liberalism-protectionism" index (IL_t or its first lag IL_{t-1}), the share of net facilitating measures ($\overline{MI_t}$ or its first lag $\overline{MI_{t-1}}$) and trade measures index (MI_t or its first lag MI_{t-1}) (see Table 3).

Table 3. Correlation matrix

	$\overline{MI_t}$	$\overline{MI_{t-1}}$	MI_t	MI_{t-1}
IL_t	0.159	0.545	0.155	0.545
IL_{t-1}	0.919	—	0.919	—

The first lag of the "liberalism-protectionism" index shows strong correlation with the indices reflecting the dynamics of trade policies (91,9% for both $\overline{MI_t}$ and MI_t). However, we cannot imply the causal relationship. The study of causality between the discourse and the implemented policies requires time-series analysis, which is beyond the scope of this paper.

5. Conclusion

The paper presents early steps in the research of the relationship between the discourse in academic literature and the implemented measures in international trade. We construct an index that reflects the article's semantic orientation in terms of adherence to either economic liberalism or economic nationalism using NLP algorithms. We compare the index dynamics with the dynamics of the aggregated number of legislative acts in the Trade Monitoring Database. We find a direct relation between political measures and the first lag of the "liberalism-protectionism" index from 2008 to 2021.

The paper demonstrates two important conclusions. Firstly, though we cannot state the causal relationship between the discourse and implemented trade policies, the existence of such a relationship is important itself. Economics, for the last decade, remains a practice-oriented discipline, despite the accusations of being "divorced from real life". Secondly, the results show the applicability of NLP methods for discourse analysis. The interdisciplinary approach to the history of economic analysis allows us to quantify the development of science. This case opens up a promising path for further quantitative content analysis, as opposed to the generally accepted qualitative approach.

References

Banerjee S., Pooter M., Grishchenko O.V., Strum B., Walsh C. (2021). Gauging the Sentiment of Federal Open Market Committee Communications through the Eyes of the Financial Press. *Board of Governors of the Federal Reserve System Working paper*.

Bird S., Loper E., Klein E. (2009). *Natural Language Processing with Python*. Sebastopol, CA: O'Reilly Media Inc.

Elseveir (2020). Scopus. Content Coverage Guide. URL: https://www.elsevier.com/__data/assets/pdf_file/0007/69451/Scopus_ContentCoverage_Guide_WEB.pdf

Gentzkow M., Kelly B., Taddy M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535–574.

Handoko L. H. (2020). History of Islamic Economic Thought: A Content Analysis. *Library Philosophy and Practice* (e-journal), 4409.

Hutto C., Gilbert E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, 8(1), 216–225.

McDermott J. (1999). Mercantilism and modern growth. *Journal of Economic Growth*, 4(1), 55–80.

Paice C.D. (1990, November). Another stemmer. In *ACM Sigir Forum*, 24(3). New York: ACM, 56–61.

Rose M.E., Kitchin R.J. (2019). Pybliometrics: Scriptable bibliometrics using a Python interface to Scopus, SoftwareX 10. 100263.

Stiglitz J., Schoenfelder L. (2003). Challenging the Washington consensus. *The Brown Journal of World Affairs*, 9(2), 33–40.

World Trade Organisation (2019). World Trade Report 2019: The future of services trade. https://www.wto.org/english/res_e/booksp_e/00_wtr19_e.pdf

World Trade Organisation (2022). TMDB @ Home – Open Trade Data Portal https://tmdb.wto.org/en

United States Trade Representative (2022). Trade Remedies https://ustr.gov/trade-agreements/wto-multilateral-affairs/wto-issues/trade-remedies#:~:text=Trade%20remedies%20are%20actions%20taken,and%20import%20surges%20(safeguards).

Жохов, М. М., Галеев, А. В.

Трансформация международной торговой политики: контент-анализ истории дискуссий с применением NLP [Электронный ресурс] : препринт WP11/2022/01 / М. М. Жохов, А. В. Галеев; Нац. исслед. ун-т «Высшая школа экономики». – Электрон. текст. дан. (500 Кб). – М. : Изд. дом Высшей школы экономики, 2022. – (Серия WP11 «Экономические реформы конца XX в.: опыт и уроки новейшей истории»). – 14 с.

В работе приводится исследование развития научного дискурса в сфере реформ международной торговли и его влияния на проводимую политику. Подготовлена база статей по международной торговле, индексируемых в базе данных научного цитирования Scopus. Рассчитан индекс, отражающий смысловую направленность статьи с точки зрения приверженности либо экономическому либерализму, либо экономическому национализму с использованием алгоритмов NLP. Приводится сравнение динамики индекса с динамикой агрегированного количества законодательных актов в базе данных мониторинга торговли. Анализ доказывает существование прямой связи между политическими мерами и первым лагом индекса «либерализм-протекционизм» с 2008 по 2021 г.

Ключевые слова: NLP, контент-анализ, дискурс-анализ, международная торговля, экономический либерализм, экономический национализм.

Жохов Михаил Михайлович – стажер-исследователь Центра истории и методологии экономической науки Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ); РЭШ.

Галеев Антон Владиславович – научный сотрудник Центра истории и методологии экономической науки Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ).

.

Препринты Национального исследовательского университета «Высшая школа экономики» размещаются по адресу: http://www.hse.ru/org/hse/wp

Препринт WP11/2022/01
Серия WP11
Экономические реформы конца XX в.:
опыт и уроки новейшей истории

Жохов Михаил Михайлович, Галеев Антон Владиславович

Трансформация международной торговой политики: контент-анализ истории дискуссий с применением NLP

(*на английском языке*)

Изд. № 2598