

ЛИНГВИСТИЧЕСКИЙ АНАЛИЗ СТРАТЕГИИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ РЕСПУБЛИКИ КРЫМ

К. К. Боярский¹, к.ф.-м.н., Е. А. Каневский², к.т.н., Н. А. Рослякова², к.э.н.

¹Университет ИТМО, Санкт-Петербург, Россия

²Институт проблем региональной экономики РАН, Санкт-Петербург, Россия

LINGUISTIC ANALYSIS OF THE SOCIO-ECONOMIC DEVELOPMENT STRATEGY OF THE REPUBLIC OF CRIMEA

K. K. Boyarsky¹, Ph.D., E. A. Kanevsky², Ph.D., N. A. Roslyakova², Ph.D.

¹ITMO University, St. Petersburg, Russia

²The Institute for Regional Economic Studies RAS, St. Petersburg, Russia

Многообразие документов стратегического планирования формируют потребности разработки инструментария аналитической обработки таких текстов и выделения ключевых смысловых конструкций в них. В данной работе на примере Стратегии социально-экономического развития Республики Крым предлагается инструментарий предварительной обработки текста (пермутационный вывод) и обобщения смысловых конструкций текста (облако слов) стратегий с целью наглядного и концентрированного представления ключевых смыслов, заложенных в стратегический документ.

The variety of strategic planning documents form the needs for the development of analytical processing tools for such texts and the allocation of key semantic structures in them. In this paper, using the example of the Socio-Economic Development Strategy of the Republic of Crimea, we propose a toolkit for preprocessing of text (permutation output) and generalization of semantic constructions of text (word cloud) strategies for the purpose of visual and concentrated representation of the key meanings embedded in the strategic document.

Анализ документов стратегического планирования имеет большое значение как для адекватного и качественного исследовательского процесса, направленного на полное понимание и учёт аспектов, находящихся в поле внимания региональных и федеральных властей, так и для выработки действенных и реальных (а не декларативных) рекомендаций по совершенствованию системы государственного стратегического планирования и системы регулирования и стимулирования социально-экономического развития регионов России.

В этом отношении большое значение имеет глубокий содержательный анализ документов стратегического планирования. Существуют немногочисленные статьи, посвященные этим вопросам. Примером такой работы является [1], где автор акцентирует внимание на сопоставлении идеологических аспектов двух стратегий и делает вывод о трансформациях, произошедших в них. Однако следует подчеркнуть, что такого рода анализ является, в сущности, кропотливым «ручным трудом» аналитика, когда автор читает, осмысливает и самостоятельно проводит сопоставления. Многообразие документов стратегического планирования и усиление общественного и государственного запроса на их сущностный, содержательный анализ не позволяет рассматривать метод «ручного труда» как оптимальный. Примером работы, где ставятся вопросы анализа определённых содержательных конструкций, является [2]. Однако авторы оставляют в стороне вопрос разработки автоматизированных методов, также ориентируясь на «ручной» труд.

Кроме того, наличие в процедурах принятия и реализации стратегических документов процесса редактирования, доработки и пересмотра, свидетельствует в пользу того, что подобную процедуру для одного документа необходимо осуществить несколько раз, что ещё больше требует автоматизации этого процесса. Проведённый авторами анализ имеющейся литературы позволяет заключить, что не существует статей, где бы ставился вопрос углублённого анализа текстов стратегических документов машинными методами, с целью автоматизировать процесс

анализа и постановки задач на поиск определённых смысловых конструкций в текстах стратегических документов.

Таким образом, можно заключить, что содержательный анализ документов стратегического планирования представляет интерес и фактором, который также стимулирует поиск подходящего инструментария, является и развитие инструментария текстового и содержательного анализа, что позволяет ставить новые задачи для исследования текстов социально-экономического направления.

В этом отношении можно сказать, что актуальные на современном этапе системы машинного обучения и искусственного интеллекта, например, популярный GPTChat [3], не в состоянии решать подобные задачи, так как по сути они являются системой генерации (условно, синтезирования) нового контента, однако они не дают возможности аналитической работы с текстом, направленной на выявление смыслов, сетевых и содержательных конструкций и т.п. Соответственно, возможности, анализа с использованием ИИ систем сводится к компилированию из уже имеющегося материала на основе слов маркеров с настройкой подражанию тому или иному стилю изложения материала. В этом отношении наш подход к работе с текстом стратегий имеет прямо противоположный характер. В рамках данной работы было сформулировано несколько запросов:

- 1) оценка частоты упоминания слов и степени их связанности в тексте;
- 2) отбор слов и более глубокий анализ контекста и окраски (положительной или отрицательной) при их упоминании;
- 3) фиксация некоторых слов, словосочетаний или смысловых конструкций и отслеживание, как меняется контекст к ним с течением времени.

Для анализа была отобрана Стратегия социально-экономического развития Республики Крым до 2030 г. [4]. Её изучение актуально с точки зрения выделения актуальных направлений и аспектов развития, которые целесообразно внести в программу социально-экономического развития новых регионов России (Луганской Народной Республики, Донецкой Народной Республики, Запорожской и Херсонской областей).

Для оценки частотности используется система ДИСКАНТ (вариант ВЕГА) [5], которая при наличии морфоанализатора позволяет составлять алфавитные и частотные словари. В табл. 1 приведены данные о 20 наиболее часто упоминаемых словах.

Таблица 1

Частотный словарь, сформированный по тексту Стратегии Республики Крым [4]

Слово	Число упоминаний	Слово	Число упоминаний
Развитие	971	Население	227
Крым	940	Уровень	197
Республика	934	Центр	195
Год	489	Регион	182
Комплекс	340	Строительство	182
Микрорегионы	333	Санаторно-курортный	180
Создание	314	Организация	179
Производство	305	Район	179
Являться	266	Инфраструктура	174
Система	249	Транспортный	173

Из табл. 1 можно видеть, что имеют место как рамочные слова (год, является, Республика Крым, (РК как объект документа стратегического планирования), регион (как синоним РК)), так и слова, предполагающие смысловые конструкции, связанные с реализацией мероприятий (создание, производство, строительство) и получением определённых эффектов (транспортный, санаторно-курортный). Однако сам словарь и частота упоминания отдельных слов являются малопоказательными для концентрации смысла и сущности предлагаемых в стратегии решений. Гораздо более продуктивным является путь построения так называемого облака слов (с использованием инструментария открытого доступа [6]), предполагающего учёт не только частоты, но и тесноты связи слов в тексте. И в этом отношении инструментарий ДИСКАНТА предлагает дополнительные возможности. В частности, он обеспечивает пермутационный вывод

— выведение интересующего слова и некоторого количества окружающих его символов. На рис. 1 и 2 представлены облака слов без учёта и с учетом пермутационных выводов. Отметим, что рамочные слова (Республика Крым, РК, год, тысячи, является) и служебные слова были исключены из анализа.

Сопоставляя рисунки, первое на что стоит обратить внимание — это повышение уровня связанности терминов при анализе текста (отражается через интенсивность цвета букв). Так, наиболее употребляемое слово «микрорегион», которое на рис. 1 более светлое, на рис. 2 оказывается в центре и демонстрирует более высокую степень связанности. Также более ярко проявляются понятия «системы» и «строительство», которые в общем тексте размывается понятием «создание». Это происходит от того, что структура текста стратегических документов часто имеет специфическую структуру, в частности описание сфер деятельности носит не структурно связанный характер (например, сельское хозяйство связано с пищевой промышленностью, розничной и оптовой торговлей, а строительство с финансово-экономическим положением предприятий региона, социально-экономическим положением населения и потребностями государства в создании объектов инфраструктуры и безопасности, которые предъявляют или не предъявляют спрос на стройку). Однако чаще всего описание идёт согласно с классификацией отраслей ОКВЭД, что снижает уровень и качество связанности смыслов, мероприятий и эффектов в них. Реструктуризация текста с использованием пермутационного вывода отчасти позволяет скорректировать эти разрывы.

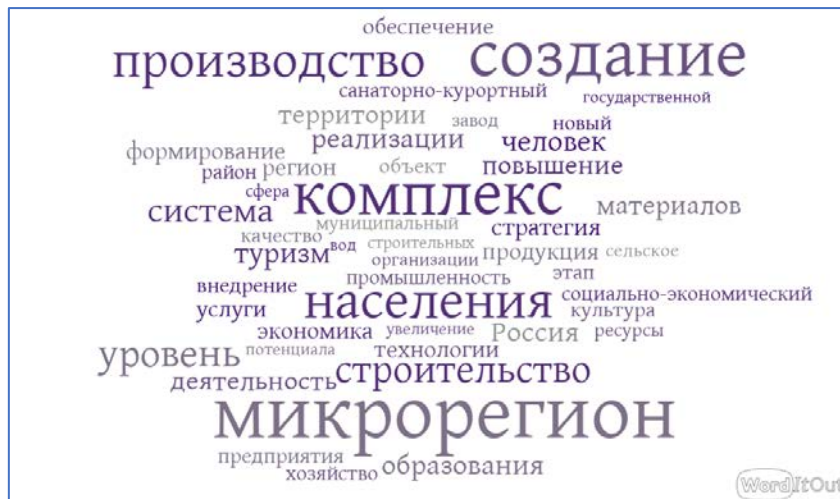


Рис. 1. Облако слов, построенное по тексту Стратегии Республики Крым [4].



Рис. 2. Облако слов, построенное по подготовленному с использованием системы ДИСКАНТ пермутационному выводу текста Стратегии Республики Крым.

Также на рис. 2 более чётко можно проследить ключевые вектора: культура–туризм–санаторно-курортные услуги и государственная деятельность, связанная с ними; потенциал инвестиций в отрасли и реализация промышленного потенциала России; технологии строительства и развитие территорий; усилия власти по формированию строительных и сельскохозяйственных организаций; комплексная и централизованная система управления; комплексный и системный подход для улучшения социально-экономического положения населения.

Литература

1. *Сургуладзе В. Ш.* Идеологическое измерение стратегии национальной безопасности Российской Федерации: Сравнительный анализ документов 2015 и 2021 годов / Гуманит. науки. Вест. Фин. Ун-та. 2022. № 12 (1). С. 60–69.
2. *Рубцов Г. Г., Литвиненко А. Н.* Использование ценностно-ориентированного подхода в стратегическом планировании на примере реализации региональных стратегий развития субъектов Северо-Западного федерального округа // *Вопр. Управ.* 2020. № 3 (64). С. 65–77.
3. ChatGPT на Русском: Чатбот от OpenAI / Портал чатбота. 2023. URL: <https://chat-gpt.org/ru>
4. Закон Республики Крым от 9 января 2017 года № 352-ЗРК/2017 «О стратегии социально-экономического развития Республики Крым до 2030 года» [по состоянию на 26.12.2022].
5. *Каневский Е. А., Боярский К. К.* ВЕГА — инструмент для лингвистических исследований // *Прикладна лінгвістика та лінгвістичні технології. MegaLing-2012.* К.: Довіра, 2013. С. 113–123.
6. WordItOut / Портал анализатора. 2023. URL: <https://worditout.com/word-cloud/create>

Статья подготовлена в рамках государственного задания ИПрЭ РАН «Механизмы формирования новых подходов к пространственному развитию экономики РФ, обеспечивающей устойчивое развитие и связанность ее территорий в условиях глобальных вызовов XXI века», АААА-А21-121011290083-2.