

ПРОМЕЖУТОЧНЫЙ ОТЧЕТ О ВЫПОЛНЕНИИ НИР (1 ЭТАП)
«Проблемы лингвистического обеспечения информационно-
поисковых систем для возможности сопоставления
наименований товаров и услуг из перечней
заявок на ТЗ с терминами МКТУ»

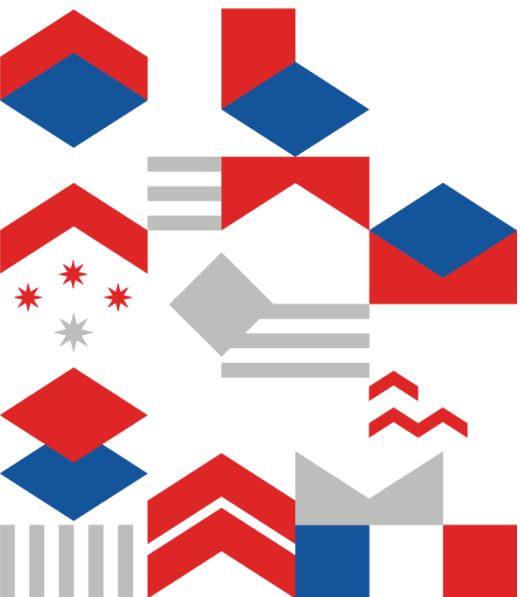
Шифр темы: **7-ИТ-2020**

*Сроки выполнения работ по 1 этапу НИР:
I квартал 2020 – IV квартал 2020*

Научный руководитель НИР
старший научный сотрудник
Маслова Е.В.

Ответственный исполнитель НИР
старший научный сотрудник
Ганин Д.Д.

Исполнители
Отдел 113





Цель исследования

Определение возможных способов автоматического сопоставления наименований товаров и услуг, включенных в перечни товарных знаков, с терминами из актуальной версии Международной классификации товаров и услуг (МКТУ) с учетом их семантики. Разработка инструментов для автоматической идентификации терминологических единиц, неоднозначных и производных терминов, встречающихся в перечнях заявок на ТЗ, для их накопления и учета при обработке перечней и экспертизе ТЗ.

Задачи исследования

1. Выявление проблем идентификации товаров и услуг.
2. Анализ способов ручного и автоматического сопоставления наименований товаров и услуг, а также способов поиска информации о товарах и услугах в существующих информационных системах.
3. Анализ основных методов исследования семантики и определение терминологической структуры для формирования смыслового содержания лингвистического справочно-информационного массива.
4. Разработка структуры лингвистического справочно-информационного массива наименований товаров и услуг.

Актуальность исследования

Актуальность работы обусловлена отсутствием механизмов решения проблем, возникающих при идентификации и классификации наименований, не входящих в действующую редакцию МКТУ, и отсутствием справочно-информационной системы по товарам и услугам.



Наименования товаров и услуг: (* красным выделены наименования, отсутствующие в МКТУ (2))

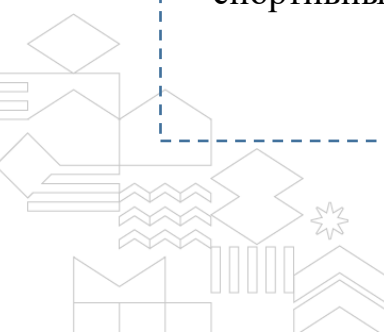
аппараты вендинговые; автоматы игровые с предварительной оплатой; **раковины (3);** акваланги (4); **антикоагулянты;** **антиперспиранты;** приборы и инструменты медицинские; **аппараты ортооптические;** **сетка армирующая (5);** аренда **радиостанций;** служба новостей...

Поисковый запрос: **Раковины**

Результат поиска по слову		Не найдены семантически близкие наименования: мидии (29,31 кл.); моллюски (29,31 кл.); морепродукты; раковины (11 кл.); ракушки (29,31 кл.); створки раковин устриц (20 кл.); устрицы (29,31 кл.) ... и т.д.
раковины (11 кл.); раковины (20 кл.); раковины для защиты паха спортивные (28 кл.)		

ПРОБЛЕМА: сложности при идентификации и классификации наименований товаров и услуг, отсутствующих в МКТУ, как на этапе составления перечня так и на этапах его обработки и экспертизы.

- 1 отсутствие единой терминологии, которой могли бы в равной степени пользоваться производители и потребители товаров и услуг, специалисты ФИПС, заявители, патентные поверенные и другие заинтересованные лица.
- 2 обеспечение потребностей заявителей в указании в перечнях заявок товаров и услуг, отсутствующих в МКТУ, и их правильной классификации.
- 3 проблема многозначности наименований товаров и услуг.
- 4 отсутствие механизма учета новых наименований товаров и услуг и их синонимов, а также недопустимых формулировок для включения в перечень ТЗ.
- 5 отнесение наименования к нескольким классам МКТУ.
- 6 отсутствие возможности семантического поиска в существующих поисковых системах.
- 7 отсутствие накопленной базы знаний .





1

ЭТАП - Анализ способов автоматического систематизирования и обработки данных в информационных системах *(I квартал 2020 – IV квартал 2020)*

1.1

Исследование представления информации в виде данных и интерпретации данных в процессе функционирования информационных систем
(I квартал 2020 г.)

1.2

Формирование смыслового содержания информационного массива в виде, удобном для обработки и обеспечивающем автоматизированный поиск информации
(II квартал 2020 г.)

1.3

Обоснование структуры лингвистического обеспечения в системах многокритериального поиска
(III квартал 2020 г.)

1.4

Исследование метода компонентного анализа для моделирования фреймовой структуры базы данных
(III-IV квартал 2020 г.)





Основные результаты 1 этапа работы:

1. Вариант идентификационного указателя, который содержит алфавитный перечень порядка 1000 наименований товаров и услуг из перечней заявок и свидетельств на ТЗ, интеллектуально сопоставленных с терминами актуальной редакции МКТУ-11 (2020) на основе анализа их характеристик.
2. Описание структуры информационного лингвистического массива, построенного на базе частотного словаря (лексические единицы из перечней заявок и свидетельств на ТЗ) и предметного указателя МКТУ (на основе базовых номеров) в виде таблиц, которые будут включать в себя лексические единицы в нормализованном виде, их контекстные синонимы, а также данные из отраслевых классификаторов (гармонизация с МКТУ).

Результаты работы могут быть использованы:

1. В качестве справочного инструмента для идентификации и классификации товаров и услуг, отсутствующих в МКТУ:
 - на этапе составления перечня заявки на ТЗ (автоматический подбор альтернативных наименований, допустимых для включения в перечень заявки на ТЗ);
 - на этапе обработки перечней и экспертизы ТЗ (уменьшение количества трудно идентифицируемых наименований и выявление сходных наименований товаров и услуг вне зависимости от их лексического выражения);
 - на этапе создания лингвистического справочно-информационного массива наименований товаров и услуг и дальнейшего включения в автоматизированную поисково-экспертную систему средств индивидуализации.

Потребители результатов исследования

Специалисты ФИПС, задействованные в экспертизе заявок на ТЗ (отд. 24, 71, 50, 42), заявители, патентные поверенные, и другие заинтересованные лица.



Вариант идентификационного указателя, где наименования товаров и услуг из перечней ТЗ вручную сопоставлены с наименованиями товаров и услуг МКТУ

Представлен вариант идентификационного указателя, в котором наименования товаров и услуг из перечней заявок и свидетельств на ТЗ и других источников (столбец 1) сопоставлены с базовыми номерами наименований товаров и услуг, включенных в актуальную редакцию МКТУ (столбцы 3,4). Напротив наименований товаров, являющиеся названиями ТЗ, в столбце 2 стоит помета «ТМ».

Новизна результатов НИР состоит в разработке справочно-информационного инструмента для ручной и автоматической идентификации наименований товаров и услуг, отсутствующих в МКТУ.

Наименование товара/услуги	ТЗ	Базовый № МКТУ	Термин МКТУ	Класс МКТУ
1	2	3	4	05
Автоматы вендинговые		070537	Автоматы торговые	07
Акваланги	ТМ	090355	Аппараты дыхательные для подводного плавания	09
Антикоагулянты		050069	Препараты фармацевтические	05
Антиперспиранты		030180; 030162	Дезодоранты для человека или животных; Средства туалетные против потения [туалетные принадлежности]	03
Аппараты ортооптические		100221	Аппаратура для функционального восстановления тела для медицинских целей	10
Аренда радиостанций		380033	Прокат оборудования для телекоммуникационной связи	38
...	
Сетка армирующая		060275; ... 170089; 190191	Материалы армирующие для бетона металлические ... Материалы армирующие для труб неметаллические; Материалы армирующие строительные неметаллические	06; ... 17; 19
...	



Структура информационного лингвистического массива

- ★ Определена структура для формирования лингвистического справочно-информационного массива наименований товаров и услуг, которая включает в себя две таблицы:
Таблица 1 содержит перечень наименований товаров и услуг из МКТУ, соотнесенных с родовыми группами и сопоставленных с кодами отраслевых классификаторов.

Таблица 1

Баз.№ МКТУ	Термин МКТУ	Лемма	I*	Контекстный синоним	Родовая группа	ОКПД2	ТНВЭД ЕАС
300280	булочки с шоколадом		0	шоколатины	изделия хлебобулочные сдобные	10.71.11.130	1905 90 600 0
		булочка	1	булка			
			1	плюшка			
			1	сайка			
			1	слойка			
			1	бриошь			
		шоколад	2	начинка			
			2	крем			

*I -индекс, отражающий последовательность лексических значений в наименовании.





Таблица 2 представляет собой нормализованный частотный словарь, составленный из слов-синонимов первой таблицы массива, связанных с базовыми номерами МКТУ, в наименованиях которых встречается ключевое слово.

В таблице указана ~ частота встречаемости слова в заявках и свидетельствах на ТЗ.



Новизна результатов НИР состоит в разработке подхода к решению вопросов идентификации наименований товаров и услуг через построение лингвистического массива на основе анализа терминологической лексики. Использование результатов работы позволит расширить возможности поисковых инструментов, уменьшить трудозатраты на обработку перечня ТЗ и повысить качество услуг, предоставляемых федеральной службой.

Таблица 2

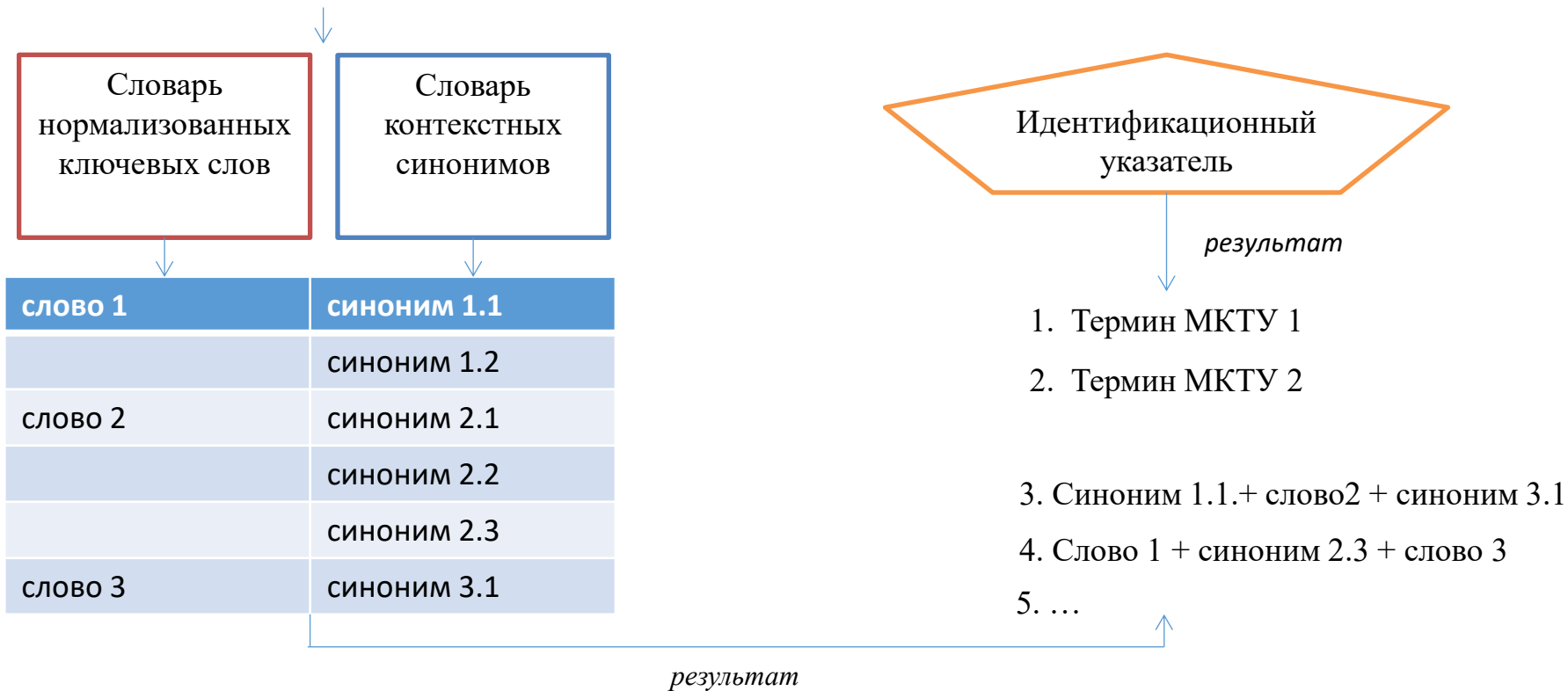
Ключевые слова	Частота встречаемости в заявках	Перевод	Индекс МКТУ
1	2	3	4
бриошь	17427	brioche	300023
булка	25464	loaf	-
булочка	934	bun	-
лепешка	25153	scone	300178
плюшка	6	roll	-
сайка	3	bread roll	-
слойка	77	puff	-
начинка	33105	stuffing; filling	300029
шоколад	10943	chocolate	300038
шоколатин	101	chocolatines	300280





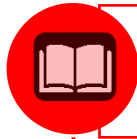
Пример графического представления процесса построения вариаций наименований товаров с использованием словарей лингвистического массива

Наименование А = слово 1 + слово 2 + слово 3



Использование лингвистического массива в поисковых системах позволяет вручную и автоматически находить и обрабатывать близкие по значению термины, включая их контекстные синонимы, в перечнях товаров и услуг, как из фонда заявок РФ, так и среди вновь поступающих наименований новых товаров и услуг, не учтенных в МКТУ.





По результатам исследования опубликованы:

- 1 Климова В.А. Возможно ли в настоящее время построение единой терминосистемы?// Интеллектуальная собственность. Промышленная собственность. 2020. № 1. С. 51-56.





Спасибо за внимание!

