

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«Южно-Уральский государственный университет  
(национальный исследовательский университет)»  
Институт лингвистики и международных коммуникаций  
Кафедра лингвистики и перевода

ДОПУСТИТЬ К ЗАЩИТЕ  
Заведующий кафедрой,  
д.филол.н., доцент  
\_\_\_\_\_ /Т.Н. Хомутова/

## **ПРОБЛЕМЫ АННОТИРОВАНИЯ (ТАГИРОВАНИЯ) ТЕКСТОВ В КОРПУСНОЙ ЛИНГВИСТИКЕ**

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА  
ЮУрГУ – 45.03.03.2018.286.ВКР

Руководитель,  
д. филол.н., профессор  
\_\_\_\_\_ /С.О. Шереметьева/  
« \_\_\_\_ » \_\_\_\_\_ 2018 г.

Автор  
студент группы ЛМ-437  
\_\_\_\_\_ /М.К.Махмутова/  
« \_\_\_\_ » \_\_\_\_\_ 2018 г.

Нормоконтролер,  
к. филол.н., доцент  
\_\_\_\_\_ /О.И. Бабина/  
« \_\_\_\_ » \_\_\_\_\_ 2018 г.

Работа защищена с оценкой  
\_\_\_\_\_  
« \_\_\_\_ » \_\_\_\_\_ 2018 г.

Челябинск  
2018

## ОГЛАВЛЕНИЕ

Введение.....	3
Глава 1 Основы аннотирования в корпусной лингвистике.....	5
1.1 Понятие и значение корпусных технологий в лингвистике.....	5
1.2 Понятие и типы аннотирования.....	10
1.3 Методы аннотирования.....	25
1.4 Области применения аннотирования.....	27
Выводы по главе 1.....	29
Глава 2 Аннотирование текстов предметной области дизайна модной одежды.....	30
2.1 Мотивация исследования.....	30
2.2 Материал исследования.....	34
2.3 Разработка тагера для текстов по дизайну модной одежды.....	35
2.3.1 Общая характеристика.....	35
2.3.2 База знаний.....	39
2.3.3 Алгоритм.....	39
2.3.4 Программа.....	40
Выводы по главе 2.....	45
Заключение.....	46
Библиографический список.....	47
Приложение 1.....	51

## ВВЕДЕНИЕ

В современной лингвистике корпусные исследования занимают значительное место. Корпусы текстов, как правило, покрывают разные функциональные стили: включая разговорную речь, художественную литературу, публицистический стиль. В зависимости от типа корпусы могут включать в себя лингвистическую разметку на морфологическом, синтаксическом, семантическом уровнях. Несмотря на обширный охват стилей, тексты специализированных журналов в моде в предметной области дизайна модной одежды, как правило, не включены в корпусы и не подвергаются тагированию. Как показывает практика, тагеры, настроенные на работу в ограниченной предметной области, отличаются хорошей производительностью.

**Актуальность** данной дипломной работы определяется необходимостью создания схемы семантического тагирования текстов в предметной области дизайна модной одежды, которую можно будет использовать для информационного поиска и семантического анализа.

**Объектом исследования** является весь процесс тагирования.

**Предметом исследования** является семантическое тагирование в предметной области дизайна модной одежды.

**Целью данного исследования** является создание схемы семантического тагирования, создание программного продукта для реализации схемы, написанного на языке программирования Python.

Проблема, цель, объект и предмет исследования определяют следующие **задачи**:

1. Анализировать теоретические источники, посвященные тагированию
2. Разработать базу знаний
3. Написать код программы
4. Создать собственный тагер

## 5. Апробировать тагер

**Материалом данного исследования** послужил корпус текста, состоящий из 4200 словоформ на английском языке, собранный с сайта издательства Vogue.com за последний год, в предметной области дизайна модной одежды, в частности дизайнера Алессандро Микеле из модного дома Gucci.

**Теоретико–методологической базой** исследования послужили работы Дж. Лича, В. П. Захарова, И. М. Богуславского, Т. Манкери, Э. Вилсона, В. Б. Виноградова, Т. Ф. Ефремова, А. Мустайоки, М. В. Копотева, Т.И. Резниковой, Р. П. Рогожникова, В.В. Рыкова.

**Структура работы** состоит из введения, теоретической части («Основы аннотирования в корпусной лингвистике»), где рассматриваются определение понятия «корпус», «корпусные технологии», «аннотирование», «тагирование», «разметка», а также рассматриваются типы, методы и области применения аннотирования, практической части («Анотирование текстов предметной области дизайна модной одежды»), которая посвящена разработке собственного тагера, заключения, библиографического списка.

**Объем работы** – работа состоит из 50 страниц, в том числе и библиографический список, состоящий из 27 наименований.

# ГЛАВА 1 ОСНОВЫ АННОТИРОВАНИЯ В КОРПУСНОЙ ЛИНГВИСТИКЕ

## 1.1 Понятие и значение корпусных технологий в лингвистике

Письменный и устный тексты являются основными источниками проведения лингвистических исследований. Термины «корпус» и «корпусная лингвистика» появились в 60-е гг. XX века. Термин «корпус» является неоднозначным, обозначая в лингвистике самые разнообразные объекты. Общим является то, что во всех случаях речь идет о некой подборке языкового материала и, как правило, имеется в виду подборка полных текстов или их достаточно крупных фрагментов.

К. Бауш дает следующее определение понятию «корпус»: корпусом является ограниченное количество языкового материала, состоящего исключительно из (письменных или устных) текстов или их частей.

Джон Синклеир из организации Tuscan Word Centre утверждает, что корпус представляет собой набор фрагментов текста в электронном виде, отобранных по видимым критериям для представления языка или языкового разнообразия в качестве источника данных для лингвистических исследований [Sinclair 2004].

В английских источниках также различаются понятия размеченного и неразмеченного корпуса. Неразмеченный корпус представляет собой простой, необработанный, “сырой” текст, в котором нет четкой лингвистической информации. Например, нет конкретного уточнения о том, что *present* выступает в роли существительного в данном контексте. Эта информация подразумевается и никак не проявляется. Размеченный корпус, в свою очередь, хранит в себе лингвистическую информацию. Например, *present – noun/verb*. Такой корпус уже не является простым текстом. Об этом говорит Ричард Сяо из Ланкастерского университета [Xiao 2003].

Корпусная лингвистика – раздел компьютерной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических

корпусов (корпусов текстов) с применением компьютерных технологий [Захаров 2011].

В свою очередь, ученые Ланкастерского университета Тони Макенри и Эндрю Вилсон говорят о корпусной лингвистике следующее: «Это изучение языка, включающего все процессы, связанные с обработкой, использованием и анализом письменных или устных машиночитаемых корпусов. Корпусная лингвистика - относительно современный термин, используемый для обозначения методологии, основанной на примерах использования «реальной жизни». В настоящее время эффективность и полезность корпусной лингвистики тесно связана с развитием информатики.

Под лингвистическим, или языковым, корпусом текстов понимается большой, представленный в машиночитаемом виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач. В.В. Рыков определяет корпус текстов как некоторое собрание текстов, в основе которых лежит логический замысел, логическая идея, объединяющая эти тексты [Рыков 2000].

Существует большое количество разных типов корпусов, что определяется многообразием исследовательских и прикладных задач, для решения которых они создаются, и различными основаниями для классификации. В зависимости от поставленных целей и классифицирующих признаков, можно выделить различные типы корпусов. Они описаны в таблице 1 [Захаров 2013].

Признак	Типы корпусов
Тип данных	Письменные Речевые Смешанные
Язык текстов	Русский

	Английский и т.д.
«Параллельность»	Одноязычные Двуязычные Многоязычные
«Литературность», специфичность	Литературные Диалектные Разговорные Терминологические Смешанные
Жанр	Литературные Фольклорные Драматургические Публицистические
Доступность	Свободно доступные Коммерческие Закрытые
Назначение	Исследовательские Иллюстративные
Динамичность	Динамические (мониторные) Статические
Разметка	Размеченные Неразмеченные
Характер разметки	Морфологические Синтаксические Семантические Просодические и т.д.
Объем текстов	Полнотекстовые

	«Фрагментнотекстовые»
Хронологический аспект	Синхронические Диахронические
«Общность»	Общие Одного писателя
Структура	Центральные и архивные Ядерные и периферийные

Таблица 1 – Классификация корпусов

Принципы построения любого корпуса текстов напрямую зависят от его практического назначения. Ниже представлены наиболее распространённые типы корпусов по классификации Hunston [Hunston 2000].

- Специализированный корпус. Представляет собой подборку текстов определенного типа: газетные тексты, научные статьи.
- Общий корпус. Включает тексты различных типов.
- Сопоставительный корпус. Состоит из двух или более подкорпусов различных языков, например, русский и немецкий, или вариантов одного языка, например, австрийский и швейцарский варианты немецкого языка.
- Параллельный корпус состоит из подкорпусов на различных языках, которые включают в себя идентичные тексты. Один текст является оригиналом, другой – переводом.
- Учебный корпус представляет собой набор текстов, написанных лицами, изучающими иностранный язык и его носителями.
- Педагогический корпус состоит из языковой информации, которая используется в процессе обучения иностранному языку.

Баранов А.Н. дает следующее определение корпусным технологиям: «Это одни из важнейших инструментов компьютерной лингвистики. Они позволяют решать большинство практических задач по анализу языка, представляя текст в формате лингвистических корпусов.» [Баранов 2001]. Развитие данного



направления дало возможность исследователю обращаться к корпусам текстов для того, чтобы наблюдать за реальным употреблением языковых единиц и их поведением в естественном контексте.

Суть первоначальных языковых исследований с использованием корпусов состояла в том, что подсчитывалась частота встречаемости различных языковых элементов. Такими элементами могли служить слова, словоформы, морфемы и словосочетания. Корпусы могут использоваться для получения разнообразных справок и статистических данных о языковых и речевых единицах [Чернякова 2012]. В частности, справляться и с задачами в области лексикографии, системах автоматической обработки текста. При решении более сложных лингвистических задач, как распознавание и синтез речи, осуществление автоматизированного и машинного переводов, проверка орфографии и грамматики также используются статистические методы. Например, на материале корпуса можно определить, какие слова относятся к группе устойчивых словосочетаний: для этого в полученных данных необходимо проверить какие слова регулярно встречаются вместе.

Со временем лингвисты стали признавать корпусы как один из самых мощных информационных ресурсов, тем самым используя их в рамках многих лингвистических направлений. Важно заметить, что использование корпусов ценится в лексикографии [Баранов 2003]. На их основе можно получить данные изменения частот использования языковых единиц и частот контекстов в различные периоды времени, данные о сочетании лексических единиц. Массив языковых данных за определенный срок, полученный в результате поиска, позволяет увидеть процесс изменения лексического состава языка, проводить анализ лексико-грамматических характеристик в разных жанрах и у разных авторов и т.д. [Рязанова 2012]. Также корпусы призваны служить источником и инструментом лексикографических работ по подготовке разнообразных исторических и современных словарей [Захаров 2011]. Данные корпусов могут

быть использованы для построения и уточнения грамматик и в целях обучения языку [Громыко 2008]. Благодаря им изменилась технология составления словарей, учебных материалов, ресурсов. На основе корпусов и с применением компьютеров словари составляются и пересматриваются за более короткий срок, тем самым предоставляется возможность фиксировать текущее состояние языка. Таким образом, информация, содержащаяся в словарях, не успевает устаревать за период их разработки и публикации. Следует отметить особый вид корпусов – *monitor corpora*, который позволяют лексикографам следить за появлением (изменением) неологизмов и новых значений у существующих слов.

Некоторые исследования в области семантики оказали влияние на исследования по лексикографии. Выявляя окружения той или иной лингвистической единицы в корпусе, становится возможным установить семантические признаки, которые характерны данной единице. Здесь следует говорить лишь о степени принадлежности слова к той или иной категории, так как чаще всего слово входит в несколько семантических категорий сразу. Исходя из вышесказанного, степень может быть определена путем подсчета частот распределения слова по различным категориям.

В университете Неймеген, в Нидерландах, разрабатываемые формальные грамматики проверяются на текстовых корпусах. На основе грамматики создается программа синтаксического анализа, которая обрабатывает корпус. Получаемые результаты обработки выявляют, насколько точно грамматика описывает данные. Исходя из этого, мы видим, что корпусы позволяют проверять новые лингвистические теории и системы автоматической обработки текста.

Сторонники осторожного отношения к использованию корпусных технологий в лингвистике полагают, что частотный подсчет и количественные данные не могут выступать критериями истинности утверждений о языковых

явлениях. Однако, благодаря репрезентативности, корпусные исследования дают возможность строить те или иные гипотезы, решать вопросы их валидности относительно языка и подтверждать их, используя корпус как источник примеров для иллюстрации различных положений. Таким образом, корпус является источником новых данных, которые необходимы для современных лингвистических исследований.

В нашей дипломной работе мы остановились на обзоре англоязычных и русскоязычных корпусов.

## **1.2 Понятие и типы аннотирования**

Изначально понятие «аннотирование» появилось в английском языке. Термин используется для обозначения процесса добавления дополнительной лингвистической информации к электронному корпусу устных или письменных данных [Leech 1997].

В русском языке понятие «аннотирование» - многозначный термин. Он может означать разметку предложения, тагирование (tagging) или сжатие текста (summarization). Мы используем данный термин в смысле разметки предложения.

В прикладной лингвистике термин «аннотирование» текстов трактуется как сообщение определенной дополнительной лингвистической информации о тексте, которое реализуется посредством его разметки в соответствии с определенной концепцией или теорией. Концепция, в рамках которой производится аннотирование, может определить задачи автоматической обработки текста и устанавливать новые стандарты. Аннотирование корпуса текстов является предварительным условием для многих методов машинного обучения при решении задач автоматической обработки текста [Бабина 2010].

Один из постулатов аннотирования, сформулированных в 1993 году одним из создателей корпуса Lancaster-Oslo/Bergen (LOB) и Британского

национального корпуса (BNC) Джеффри Личем, предполагает существование эксплицированных и доступных описаний лингвистической разметки.

Также, по его словам, разметка общедоступного корпуса должна подчиняться трем принципам.

1) Разметка должна основываться на доступной для пользователя в виде руководства или инструкции схеме анализа, в которой введение каждого параметра должно быть мотивировано.

2) Разметка общедоступного корпуса должна быть «теоретически нейтральна», то есть схема разметки по возможности должна опираться на знакомую всем систему понятий. Если корпус предназначен не для конкретного проекта, то при его разметке стоит избегать пусть и строгих, но авторских, необщепринятых классификаций, которые требуют предварительного знакомства с той или иной теорией.

3) Должно быть ясно, кто и как разрабатывает схему аннотации и каковы ограничения, например юридические или технические, при пользовании корпусом [Leech 1993].

Первый постулат Дж. Лича кажется очевидным, однако он реализуется не во всех корпусах. Конечно, все корпуса снабжены более или менее подробной справкой, но получение точных сведений о том, что означает каждый параметр описания, какие теоретические или чисто практические мотивировки стоят за выделением той или иной единицы – все это часто остается неэксплицированным или комментируется в узкоспециализированных сборниках статей.

Разметка — главная характеристика корпуса; она отличает корпус от простых коллекций (или «библиотек») текстов, в изобилии представленных в современном интернете. Подразумевается наличие в корпусе неких данных, не являющихся частью текста, но несущих какую-то информацию о нём (так

называемые метаданные). Простейший пример таких данных — отметки частей речи.

В 80-х годах был принят стандарт разметки электронных текстов под названием SGML (Standard Generalized Markup Language). Смысл SGML был в том, чтобы документы, набранные в разных текстовых процессорах, можно было редактировать, анализировать и изменять в любом из них. SGML ввёл концепцию тэгов. Тэги (англ. tags) — это служебные пометки в тексте, содержащие информацию о самом тексте. Для каждого случая можно определять собственные тэги.

Размечать большие корпуса вручную — занятие очень долгое и дорогое, поэтому уже в 70-х годах появились первые проекты по поручению этой задачи компьютеру. Тогда программа TAGGIT смогла корректно назначить тэги частей речи 77% слов в Брауновском корпусе. Остальные пришлось размечать вручную в течение 10 лет. В 80-е годы система CLAWS (Constituent Likelihood Automatic Word-tagging System) правильно разметила уже около 95% Брауновского корпуса. В ней использовался вероятностный подход.

Эта же система разметки использовалась при создании BNC (British National Corpus). Система пережила ряд изменений, прежде чем был получен финальный вариант CLAWS4. Система CLAWS1 была основана на скрытой марковской модели и была способна корректно разметить от 96 % до 97 % любого текста. При переходе от CLAWS1 к CLAWS2 пропала необходимость ручной подготовки текста перед запуском процесса разметки. В CLAWS4 вошли такие улучшения, как более мощные возможности по разрешению лексической многозначности и учёт вариаций в орфографии. Дальнейшая работа над системой разметки была ориентирована на увеличение показателей успеха автоматической разметки и на уменьшение ручной работы по подготовке текстов перед началом разметки путем введения в использование

дополнительного программного обеспечения для замещения ручной работы. Позже была добавлена разметка, указывающая на неоднозначность некоторых слов и выражений. При этом, несмотря на способность CLAWS4 автоматически определять типы речи и значения слов, необходимость в ручной разметке сохранилась, так как в CLAWS4 не поддерживаются языки, отличные от английского.

В настоящее время для основных европейских языков уже реализованы как автоматическая разметка частей речи (морфологический анализ, word-class tagging), так и автоматическая разметка членов предложения (синтаксический анализ, parsing). Эти достижения используются, в том числе, и в системах автоматического перевода и интернет поиска.

В этой связи нужно отметить немалый вклад рабочей группы учёных под названием «Автоматическая обработка текста» <http://www.aot.ru> . В основном они занимаются русским языком. Эта группа ученых факультета лингвистики РГГУ, которая занимается приложением теоретической лингвистики к современным компьютерным технологиям. Они разработали модули графематического (определение границ слов), морфологического (определение частей речи), синтаксического (определение членов предложения) и семантического (выявление семантических связей между словами) анализа текстов на русском, немецком и английском языках.

Разметка корпуса текстов с целью аннотирования может осуществляться на различных уровнях.

Морфологическая разметка содержит информацию о морфологических формах и значениях (часть речи, одушевленность, степень сравнения, краткость, вид, время, репрезентация, лицо, залог, род, число, падеж). Она необходима для того, чтобы осуществлять поиск нужных пользователю слов, форм слова, конструкций и т.д.

Элементы морфологической структуры слова распознаются с помощью различных алгоритмов морфологического анализа. К алгоритмам, широко применяемым на морфологическом уровне, относятся стемминг и лемматизация. Цель стемминга в отождествлении основы семантически схожих словоформ, что является необходимым аспектом для адекватного взвешивания терминов в процессе информационного поиска. На входе стеммера подается текст, на выходе мы получаем список основ слов входного текста.

Стеммеры классифицируются на алгоритмические и словарные [Hull 1996]. Первые функционируют на основе файлов данных, содержащих списки деривационных суффиксов и флексий, вторые на основе словарей основ слов. Словарные стеммеры обеспечивают большую точность поиска, алгоритмические – большую полноту, допуская больше ошибок, но несмотря на это они намного более распространены.

Для английского языка самыми известными стеммерами являются: стеммер Портера и ланкастерский стеммер.

Лемматизация тоже отождествляет основы слов, но проводится с учетом частей речи, к которым относятся словоформы. Например, стеммер отождествит read, reads, reader, readers с одной основой read, а лемматизер, напротив, отождествит глагольные формы с основой read, а именные формы с лексемой reader. Под леммой понимается лексема, а задачей лемматизации является отождествление словоформы, соотносящееся с одной лексемой.

Существуют лемматизированные и нелемматизированные списки, которые часто применяются для автоматической обработки английских текстов. Они были составлены А. Килгаррифом на основе Британского национального корпуса [Kilgarriff 1997]. По сути, эти списки являются словарями, в которых есть информация о частотностях и распределении лемм и словоформ по текстам и частям речи. Словари лемм широко используются в корпусной лингвистике для поддержки лингвистических исследований. Распределение по частям речи

также является существенным параметром при проведении автоматической классификации и категоризации текстов [Santini 2006].

Если брать во внимание русские корпуса, то можно отметить, что морфологической разметкой снабжен Уппсальский корпус, который лег в основу Тюбингенского корпуса (тем самым именно в рамках тюбингенского проекта изначально неаннотированный Уппсальский корпус получил морфологическую разметку). Общий объем морфологически аннотированных текстов – 2,3 млн. словоупотреблений. Разметка осуществлялась при помощи статистического морфологического анализатора ТпТ, разработанного Т. Брантсом (университет г. Саарбрюкен) [Brants 2006]. Программа не предназначена для анализа текстов на каком-либо конкретном языке, она может применяться к различным языкам и настраиваться с различными наборами параметров. Обучение программы для ее применения к русскому корпусу производилось на материале списков текстоформ и тэгов, а также с помощью размеченного вручную текста объемом 165 тыс. словоупотреблений. По оценке разработчиков корпуса, эффективность анализатора составляет 93-94% [Betsch 2001].

Другим примером выступает корпус Хельсинский аннотированный корпус русского языка, в котором процедура автоматической морфологической разметки осуществлялась при помощи системы автоматического распознавания русских слов RUSTWOL [Rustwol 1997]. Затем средствами отдельного модуля проводилось автоматическое снятие омонимии, при этом использовались лишь те правила, которые обеспечивают стопроцентную точность при устранении некоторого варианта морфологического разбора (например, случаи управления предлога определенным падежом). На следующем этапе осуществлялась ручная проверка результатов анализа.

На основном массиве текстов Национального корпуса русского языка морфологическая разметка осуществляется автоматически с помощью



программы MYSTEM, однако в части корпуса (на настоящий момент объемом свыше 4 млн. словоупотреблений) произведено ручное снятие омонимии и дополнительная коррекция результатов работы программы автоматического морфологического анализа.

Разные стратегии, выбранные разработчиками корпусов, порождают принципиальные различия в соотношении запрашиваемой пользователем и выдаваемой системой информации.

Синтаксическая разметка. Как известно, процесс внесения синтаксической информации в корпус в меньшей степени поддается автоматизации и требует больше ручной работы, чем морфологическая разметка. Поэтому на текстовых коллекциях большого объема задача детального синтаксического анализа обычно не ставится.

Главным алгоритмом, применяемым на синтаксическом уровне, является синтаксическая декомпозиция. На входе текст, на выходе выдается список предложений текста. Разработка таких алгоритмов началась в 1960-х гг. Алгоритмы синтаксической декомпозиции распознают предложения на основе символов: пробелов, знаков пунктуации, но здесь мы сталкиваемся с проблемой. Стандартное форматирование текста может отсутствовать, так как знаки, применяемые в качестве разделителей, могут использоваться не только в конце предложения, но и в середине. Оглавления, заглавия разделов, названия рисунков и таблиц, колонтитулы – все эти единицы текста могут не являться предложениями, но, тем не менее, будут форматироваться именно так. Ошибки в распознавании предложений очень сильно снижают эффективность таких систем.

Синтаксическая декомпозиция является основой для выполнения целого ряда алгоритмов распознавания фразовой структуры предложения. Довольно сильно распространены алгоритмы выделения n-gram. Они состоят из

словосочетаний, в которые входят два, три и более токена (биграмы, триграммы, тетрграммы и тд) [Bickel 2005].

Разбивка на словосочетания проводится с учетом позиции токена в предложении. Распознавание n-gram происходит на основе различных правил.

Анализ распределения n-gram помогает выявить статистически значимые словосочетания и часто применяется в стохастических алгоритмах аннотирования тегами частей речи. Начало и конец предложения обозначаются условными тегами, что позволяет рассматривать в качестве триграмм предложения состоящие из одного токена. Также, это помогает установить вероятностные параметры, необходимые для выбора тега.

Например, в корпусе современного американского варианта английского языка СОСА словосочетание I like to встречается 4810 раз, а словосочетание I live women – 29 раз. Это означает, что у первого словосочетания гораздо большая вероятность.

Для анализа морфологически значимых словосочетаний применяются чанкеры. На выходе они выдают списки фраз определенного типа (именные, глагольные, адъективные, адverbальные). Чаще всего распространены именные чанкеры (noun phrase). Они распознают словосочетание с управляющим существительным. Распознавание словосочетаний этого типа выполняется на основе предварительного аннотирования тегами частей речи и объединения отдельных частей речи во фразы на основе правил грамматики.

В русском языке в Тюбингенском корпусе к элементам синтаксической разметки можно отнести аннотирование устойчивых синтаксических оборотов, выступающих в качестве эквивалента слова (например, потому что, все равно, как раз, прежде всего). Также размечены предложные группы, т. е. можно осуществлять поиск по сочетанию предлога с именной группой в заданном падеже, определяя при этом входящее в ее состав существительное по признаку одушевленность/неодушевленность (запросы по «именным синтаксемам»

формулируются при помощи предлога и соответствующего вопросительного местоимения: в ком, в чем, в кого, во что и т. п.; всего в корпусе 109 типов таких групп).

Закономерно, что разработчики хельсинкского проекта, ограничившись небольшим объемом текстов с целью максимального охвата грамматической информации, предполагают включить в корпус детальную синтаксическую разметку. На настоящий момент результаты этой работы еще не доступны в интернет-версии хельсинкского корпуса, однако их размещение планируется осуществить в ближайшее время. По замыслу разработчиков, синтаксическая разметка будет нетривиальным образом совмещать (точнее, предлагать параллельно) две схемы аннотирования: в терминах членов предложения (что позволит описать синтаксические узлы на доступном широкому кругу пользователей метаязыке) и в терминах деревьев зависимостей (что в свою очередь позволит аккуратно представить прежде всего типы и иерархию синтаксических связей) [Мустайоки 2003]. Кроме того, в доступной версии ресурса, размечаются неоднословные устойчивые обороты: на основе списков из работ Ефремова [Ефремова 2004], Рогожниковой [Рогожникова 2003] выделяется приблизительно 2000 таких единиц.

В Национальном корпусе русского языка задача сплошной синтаксической разметки в объеме всего корпуса не ставится. К элементам синтаксического анализа, который предполагается реализовать на материале всего корпуса, относится аннотирование устойчивых оборотов, установление списка которых производится с учетом частотности в корпусе. Кроме того, на сайте Национального корпуса планируется поместить фрагмент синтаксически размеченного корпуса (общим объемом свыше 30 тыс. предложений) с использованием аппарата грамматики зависимостей.

Семантическая разметка предметно ориентирована, поскольку определяется онтологией ПО и производится экспертами. Процессу семантической разметки

специализированного корпуса текстов предшествует достаточно длительный (2–3 месяца) предварительный этап совместной работы экспертов, лингвистов и разработчиков системы, в рамках которого происходит обмен компетенциями, выработка и согласование признаков и принципов разметки. Результатом этого этапа является инструкция по семантическому аннотированию. Речь идет о двух видах семантического аннотирования:

- терминологическая разметка, которая в первую очередь предназначена для фиксации в тексте имен понятий ПО
- разметка отношений (или ситуаций, представляющих собой многоместные отношения), в которых размеченные сущности выступают в определенных семантических ролях.

Первые работы, связанные с полноценной семантической аннотацией текстов, предполагали привязку текстовых словоупотреблений к одному из значений толкового словаря. Самый известный эксперимент – это эксперимент с определением значения слова *bank* ('берег', 'учреждение' и др.) по словарю *Longmans Dictionary of Contemporary English (LDOCE)* [Wilks 1990]. Опираясь на кластеризацию слов в LDOCE (объединение частных значений в более общие группы), группа Уилкса определила значение слова в 200 предложениях. Оказалось, что точность автоматического распознавания на уровне кластеров достигала 90%, а на уровне частных она составляла всего 53%. В 1994 г. Брюс и Вибе продемонстрировали проект, в котором по словарю LDOCE вручную было размечено уже 2 476 употреблений слова *interest* 'интерес', 'прибыль' и др. [Bruce 1994]. Вполне естественно, что привязка семантической аннотации к индивидуальным толкованиям в словаре требовала работы с каждым словом, а следовательно, WSD могло быть проведено только выборочно, для одного или нескольких слов (*sample method*).

Современные системы семантической разметки используют привязку не к словарям, а к семантическим сетям или лексическим классификациям, среди

которых наиболее популярен WordNet <http://wordnet.princeton.edu/> , использующий разбиение на значения из словаря Oxford Advanced Learners Dictionary [Fellbaum 1998]. Первым на его основе был размечен подкорпус Брауновского корпуса, содержащий 234 136 размеченных словоупотреблений, из которых 186 575 многозначны [Miller 1993]. Затем появилась система LEXAS, в которой вручную были размечены 192 800 словоупотреблений, относящихся к двум сотням наиболее частотных существительных и глаголов [Lee 1996]. Корпус SemCor, созданный в Принстонском университете, содержал 700 000 слов, 200 000 из которых были вручную размечены по значениям WordNet 1.6, а впоследствии автоматически перекодированы в WordNet 1.7.-2.0.

На сегодняшний день в мире насчитывается небольшое количество корпусов с семантической разметкой. Например, в русском корпусе Газетных Текстов включены элементы семантической разметки. Во-первых, некоторым словам приписаны семантические признаки. Таким образом размечены прежде всего имена, обозначающие лиц и животных, которые разбиваются на 60 классов, организованных по семантическим и словообразовательным принципам. В остальные 10 классов попадают имена со значением действия (единый класс), имена речи, глаголы речи, прилагательные цвета и несколько других словообразовательно-лексических классов. Во-вторых, в корпусе размечены синонимические отношения между отдельными лексическими единицами. Тем самым пользователь может по одному синониму получить контексты употребления для всего синонимического ряда.

В Хельсинском аннотированном корпусе русского языка семантическая разметка находится в стадии первичной разработки и будет осуществлена после завершения синтаксической части. Планируется, что корпус будет содержать информацию о семантических категориях, список которых разрабатывается научным коллективом под руководством А. Мустайоки [Мустайоки 2005]. Как предполагают авторы, на базе морфологической и синтаксической разметки

создание функциональной части корпуса можно будет частично автоматизировать, однако планируется осуществлять и ручную работу.

В Национальном корпусе русского языка семантическая разметка осуществляется автоматически: большинству лемм в тексте приписывается один или несколько семантических и словообразовательных признаков. При этом подробная классификация охватывает не только предметные имена, но и непредметную лексику, прилагательные, глаголы и наречия. Важно отметить, что одна лемма может попадать одновременно в разные классы. Вместе с тем, поскольку все семантические пометы, присвоенные вокабуле в словаре, автоматически переносятся на любое ее вхождение в корпус, лексические омонимы не разводятся, совмещаясь в одной лемме. В силу этого поиск только одного члена полной омонимичной пары невозможен. Так, например, запрос «ЛУК: существительное: 'оружие'» выдает и контексты такого рода: «Золотистые связки лука над крыльцом». [Довлатов 1983]. Однако в настоящее время ведется работа по созданию и внедрению в корпус семантических фильтров, которые позволят по заданному контексту или конструкции автоматически снимать лексическую многозначность.

Много проблем связано с семантической разметкой корпуса, поскольку такого рода аннотирование связано не только с выбором определенного формализма, но и с решением вопроса, информация какого типа должна маркироваться семантическими тегами. В мировой практике эксперименты по реализации семантической разметки можно разделить на три класса. Под семантической разметкой может пониматься, во-первых, маркирование частных значений многозначных лексем (в этом случае разработчикам необходимо выбрать словарь, на основе которого будет производиться дифференциация значений), во-вторых – приписывание лексемам обобщенных семантических признаков на основе некоторой классификации лексики (в этом случае возникает проблема выбора стандартной, общепринятой классификации) и в-

третьих – отражение семантических отношений между словами в тексте (при этом нужно определить набор семантических ролей для каждой лексической единицы).

Метаразметка текстов является существенной частью поискового аппарата корпуса. Под метаразметкой понимается приписывание тексту атрибутов, характеризующих обстоятельства его создания, автора, тематику, жанровые особенности и так далее. Метаразметка нужна исследователю, пользующемуся корпусом для составления произвольных выборок текстов с заданными внешними параметрами. С помощью метаразметки становится несложным нахождение текстов, написанных мужчинами или женщинами, родившимися в определенный промежуток времени, нахождение текстов определенных жанров, автобиографий и т.д. Различные корпуса, в разной степени нуждаются в метаописании входящих в их состав текстов. Например, хельсинский корпус довольно однородный по внешним текстовым параметрам. Тюбенгенский корпус содержит разнородные текстовые коллекции. Система поиска допускает запросы не только по всему корпусу, но и по его части, однако формирование пользовательского подкорпуса возможно только путем простого выбора одной или нескольких из выделяемых в составе корпуса коллекций. Эти коллекции организованы по разным параметрам: это может быть автор текста, жанр текста, тип текста, источник текста. Кроме того, особую коллекцию образует Уппсальский корпус, поиск в котором можно вести отдельно по художественным и публицистическим текстам.

Метаописания газетных текстов, образующих корпус Газетных Текстов, включают в себя детальную жанровую классификацию статей. На основе анализа материала был выявлен круг основных жанрообразующих факторов, характеризующих предмет сообщения, его коммуникативную цель и композиционно-стилевую форму. По этим параметрам было выделено 9 жанровых типов (собственно информационные, информационно-

публицистические, собственно публицистические, художественно-публицистические, рекламные жанры и др.), которые распределяются между 96 конкретными жанрами. Таким образом, пользователь может для своих исследовательских целей формировать подкорпус на основе жанровых признаков текстов (задавая жанр и/или жанровый тип текстов). Кроме того, параметром при определении собственного подмассива текстов может служить наименование издания в сочетании с датой его выпуска. Использование такой подробной жанровой классификации представляется небесспорным. Во-первых, для исследования лингвистических особенностей того или иного жанра необходимо, чтобы каждому из них соответствовало значительное количество статей в корпусе. Во-вторых, в этом случае, как кажется, трудно избежать произвольных решений при отнесении той или иной статьи к конкретному жанру. Кроме того, метаразметка корпуса Газетных Текстов не учитывает ряд параметров, традиционно используемых для классификации текстов; и если, например, характеристика по полу и возрасту автора действительно не столь существенна в применении к газетным текстам, то тематика статьи (политика, спорт и т. п.) в некоторой степени определяет ее лингвистические особенности.

Значительный объем и разнородность текстов в Национальном корпусе русского языка потребовали разработки подробной системы метаразметки. За ее основу была взята классификация, предложенная в рекомендациях EAGLES, которая затем была адаптирована к русскому материалу с учетом отечественных традиций в области стилистики и типологии текстов. В итоге все тексты, входящие в НКРЯ, характеризуются с точки зрения множества различных параметров, в том числе по полу и возрасту автора, году создания текста, его объему, сфере его функционирования, его жанру, типу, тематике и др. Представлена сравнительно простая система метаразметки, подходящая для рядового пользователя, а не для специалиста по корпусной лингвистике.



Другие типы разметки. Наличием некоторых других типов лингвистической и экстралингвистической разметки характеризуется корпус Газетных Текстов. Большинству лемм в корпусе приписана их морфемная модель, т. е. схема с заполненными аффиксальными позициями и переменной для корня. Таких схем на материале интернет-версии корпуса объемом свыше 200 тыс. словоупотреблений обнаруживается более 5,5 тыс. Словообразовательная разметка позволяет осуществлять поиск слов (или включающих их контекстов), которые отвечают заданной морфемной модели или содержат заданные аффиксы.

Кроме того, большинству слов приписаны ранги в соответствии с их частотностью в корпусе. По уровню частотности все встречающиеся в корпусе леммы разбиты на 20 групп: первую образует слово с частотным рангом 1, последнюю – с частотными рангами 32769-65536. Тем самым одним из параметров при поиске могут служить частотно-ранговые характеристики лемм: можно определять уровень частотности той или иной единицы в корпусе (по принадлежности к одной из 20 групп), частично просматривать списки лемм, относящихся к определенной группе, а также ограничивать поиск какого-либо лингвистического явления словами определенной частотности.

Завершая обзор типов аннотирования, хотелось бы обратить внимание на справедливое замечание Дж. Лича о том, что разметка корпуса должна быть по возможности теоретически нейтральной.

Однако на практике выполнение этого постулата сталкивается с серьезными трудностями. Дело в том, что степень полноты и общепризнанность классификаций языковых уровней существенно различается. Например, в научной литературе по морфологии могут дискутироваться вопросы о количестве русских падежей, но не вызывает сомнения сам факт существования категории падежа. В области синтаксиса, как известно, такого единства нет. Широко распространенная в практике преподавания классификация,

опирающаяся на представление о главных и второстепенных членах предложения, не может считаться общепризнанной; современные синтаксические теории, описывающие синтаксические отношения в виде структуры составляющих, не имеют столь же широкого распространения, особенно в учебной практике; подходы функционального синтаксиса плохо согласуются с положениями «Русской грамматики» 1980-го года, и т. д.

Еще больше проблем связано с семантической разметкой корпуса, поскольку такого рода аннотирование связано не только с выбором определенного формализма, но и с решением вопроса, информация какого типа должна маркироваться семантическими тегами.

Таким образом, неравномерность аннотирования разных языковых уровней выявляет, среди прочего, две существенные проблемы современной русистики: отсутствие полных теоретически обоснованных и общепринятых классификаций, с одной стороны, и сложность (граничащая с невозможностью) автоматического аннотирования на основе этих классификаций, – с другой. В этом смысле всякий языковой корпус в силу необходимости тотального описания материала кристаллизует проблемные области в описании того или иного языка. Он оказывается не только инструментом для быстрого поиска примеров, но и источником совершенствования теоретических и чисто дескриптивных подходов к языку.

Ряд исследователей производят попытки создать универсальную схему аннотирования, которая бы покрывала все эти уровни.

### **1.3 Методы аннотирования**

Методы аннотирования бывают лингвистические и статистические. Методы аннотирования построены на концепциях. Лингвистические - на материале качественных правил. Статистические построены на количестве. Статистические методы стали популярными в 1950-х годах. К сожалению, развитие этих методов очень быстро закончилось. Это объясняется прежде

всего двумя факторами. Во-первых, проблема доступности данных. Одной из проблем применения статистических методов к языковым данным в то время было то, что наборы данных были в целом настолько малы, что невозможно было сделать интересные статистические обобщения по большому количеству языковых явлений. Во-вторых, произошел общий сдвиг в социальных науках. Позднее лингвист Ноам Хомский сосредоточился как на формальной методологии, так и на теории лингвистики, которые полностью отрицали статистический метод [Хомский, 1957].

Пользователей корпусов, как правило, интересует не содержание конкретных текстов, а их метатекстовая информация и примеры употребления тех или иных языковых элементов и конструкций. Это, в первую очередь, лингвисты. Первоначальные лингвистические исследования, проводившиеся с помощью корпусов, сводились к подсчету частот встречаемости различных языковых элементов. Статистические методики используются в решении сложных лингвистических задач, таких как машинный перевод, распознавание и синтез речи, средства проверки орфографии и грамматики и т.д. Так, устойчивые словосочетания представляют собой с семантической точки зрения неделимую смысловую единицу, что очень важно учитывать в лексикографии, системах автоматической обработки текста. На материале корпуса статистическими методами можно определить, какие слова встречаются вместе регулярно и, таким образом, могут быть отнесены к устойчивым словосочетаниям. Корпусы являются богатым источником данных для исследований по лексикографии и грамматике. С исследованиями по лексикографии тесно связаны исследования в области семантики. Наблюдая окружения той или иной лингвистической единицы в корпусе, можно установить определенные семантические признаки, характеризующие данную единицу.

Лингвисты-теоретики используют корпуса в качестве экспериментальной базы для проверки гипотез и доказательства своих теорий. Прикладные лингвисты (преподаватели, переводчики и т.п.) используют компьютерные корпуса при обучении языкам и для решения своих профессиональных задач. Особый класс пользователей представляют компьютерные лингвисты: они пытаются выявить и использовать статистические и лингвистические закономерности, присутствующие в текстах, для создания компьютерных моделей языка. Другие специалисты по языку (литературоведы, редакторы) также в ряде случаев могут получить ответы на интересующие их вопросы, обратившись к корпусу. Специалисты по общественным наукам (историки, социологи) также могут изучать свои объекты через язык, используя такие параметры текстов, как период, автор или жанр. Литературоведы используют корпуса для стилиметрических исследований. Наконец, корпуса используются для разработки и настройки различных автоматизированных систем (машинный перевод, распознавание речи, информационный поиск).

#### **1.4 Области применения аннотирования**

Тагирование было создано в качестве инструмента для помощи в классификации, обозначении принадлежности, указании границ, определении идентичности. Область применения тагирования достаточно широка. Тег может обозначать слово, изображение или являться любой другой идентификационной меткой. Аналогом в реальном мире являются научно-справочные картотеки экспонатов музеев. Идентификация и классификация объектов и информации с помощью слов возникла задолго до появления компьютеров.

Тагирование завоевало популярность из-за масштабного пользования социальными закладками, обмена изображениями и социальных сетей. Сайты позволяют пользователям создавать и управлять тегами, которые классифицируют контент с использованием простых ключевых слов.

В 2003 году вебсайт социальных закладок Delicious предложил своим пользователям добавлять теги к своим закладкам, чтобы ускорить последующий их поиск. Также появилась возможность видеть в одном месте всю информацию, связанную с конкретным тегом. Flickr предложил пользователям добавлять собственные текстовые метки каждой своей фотографии, создавая гибкую и простую систему метаданных для ускорения поиска изображения. Начиная с 2010 года YouTube также ввел систему тагирования. Теги — это ключевые слова, описывающие видео. Например, видео для серфинга может быть помечено как «серфинг», «вода» и «волны». Видео, связанное с этими тегами, будет отображаться в результатах поиска.

Также, для информационного поиска используется о — это визуальное представление списка категорий (или тегов, также называемых метками, ярлыками, ключевыми словами и т. п.) Обычно используется для описания ключевых слов (тегов) на веб-сайтах, или для представления неформатированного текста. Теги чаще всего представляют собой отдельные слова, и важность каждого тега обозначается размером шрифта или цветом. Такое представление удобно для быстрого восприятия наиболее известных терминов и для распределения терминов по популярности относительно друг друга.

Тагирование часто применяется для парсинга. Например, Stanford Parser. <http://nlp.stanford.edu:8080/parser/index.jsp> На входе подается предложение, на выходе мы получаем следующий список: наш запрос, протагированное предложение, разбор, универсальные зависимости и статистика, которая включает в себя количество токенов и время обработки запроса.

Одним из примеров также может послужить тагирование частей речи на сайте <http://parts-of-speech.info/>. Эта программа работает на базе Стенфордского таггера и отмечает слова заданного предложения различными цветами в зависимости от части речи.

Genia Tagger Demo – программа для автоматического анализа по непосредственным составляющим. <http://nactem7.mib.man.ac.uk/geniatagger/> После ввода предложения выдаются типы фраз, протагированные предложения, именованное распознавание объектов, вид распознавания и все эти данные сведенные в общую таблицу.

Еще одна область применения – формирование аннотаций ресурсов. Выполняется с применением тегов, наборов ключевых слов, наборов понятий и наборов триплетов. При использовании тегов аннотирование выполняется путем их добавления в тексты документов. Это делается для того, чтобы обрабатывающая программа могла определить формальный смысл выделенных с их помощью частей.

Другим способом аннотирования ресурсов является использование набора ключевых слов. Такое средство чаще применяется в электронных библиотеках для аннотирования научных статей. Ключевые слова для конкретной статьи выбираются их авторами, а затем хранятся в базе данных для поддержки процесса поиска.

Третьим способом аннотирования ресурса является его описание с помощью наборов у , имеющих следующий формат: субъект–предикат–объект, которые также называются триплетами. Данный способ является самым новым, и его использование еще в достаточной степени не исследовано.

### **Выводы по главе 1**

С появлением и развитием корпусных технологий у учёных появилась возможность автоматизировать лингвистические исследования. Основным инструментом корпусных технологий считается корпус.

Ряд исследователей производят попытки создать универсальную схему аннотирования корпуса, которая бы покрывала все уровни, но на сегодняшний день не существуют такой схемы аннотирования, так как естественный язык представляет собой «гибкую» систему, с трудом поддающуюся

формализованному описанию в рамках какой-либо из теорий. Разметка корпуса текстов с целью аннотирования может осуществляться на различных уровнях: морфологическом, семантическом, синтаксическом. Аннотирование корпусов текстов осуществляется вручную, либо с помощью автоматизации. В последнем случае используются морфологические и синтаксические анализаторы, тезаурусы и онтологии. Однако традиционно окончательным арбитром в процессе аннотации текстов остается человек, и автоматизированная процедура аннотирования корпуса текстов с необходимостью включает в себя этап интер- или постредактирования.

## ГЛАВА 2 АННОТИРОВАНИЕ ТЕКСТОВ ПРЕДМЕТНОЙ ОБЛАСТИ ДИЗАЙНА МОДНОЙ ОДЕЖДЫ

### 2.1 Мотивация исследования

Женская журнальная периодика – явление в системе СМИ, связанное с историей социализации женщины. Первоначально женские журналы носили элитарный характер, не пользуясь особой популярностью у широких слоев населения. Однако впоследствии за счет расширения тематики, более внимательного отношения к дизайну издания смогли значительно увеличить свою аудиторию и достигнуть миллионных тиражей. Правда, это удалось сделать лишь отдельным изданиям, которые известны нам и сегодня. Одним из таких журналов по сей день остается Vogue. С момента создания журнал занял прочную лидирующую позицию на рынке и зарекомендовал себя как «Библия Моды».

В нынешние дни журнал Vogue обладает большой значимостью на мировом рынке и среди обычных читателей. Каждый день сайт журнала генерирует большой поток информации, новостей мира моды и текстов о различных дизайнерах и людях, которые формируют моду каждый день. Именно тот факт, что Vogue является самым влиятельным журналом в этом направлении, обусловил выбор материала нашего исследования.

Мода, в свою очередь, — это не только одежда и форма прикладного искусства, это и социальное явление. Это манера жить, интерьер, курорты, домашние животные. Мода – это все, что нас окружает. На нее может повлиять, например, природа или политика: глобальное потепление, открытие новых залежей нефти, новая первая леди страны и тд. К сожалению, эта достаточно важная предметная область незначительно изучена в частности тагирования, несмотря на достаточно большой имеющийся корпус текстов.



Предметно ориентированных тагеров для текстов по модному дизайну найти не удалось. С трудом была найдена программа для парсинга одежды с фотографий и термин fashion tag, введенный компанией Apparel Search.

Термин Fashion tag - в переводе “фэшн таг”. Технически, этот термин, как и сама концепция тагирования, не были изобретены компанией Apparel Search. Тег/хэштег был придуман Крисом Мессина в 2007 году. Термин tag также зачастую используется в работе с вебсайтами. Компания пошла дальше и решила изобрести свой собственный термин. С помощью тагов на сайте представляется возможным найти информацию, фотографии или видео, которые нужны пользователю. Таги помогают категоризировать контент, который был загружен на вебсайт.

“Фэшн таг” — это ключевое слово или категория, которая используется на сайтах о моде для того, чтобы определить какие темы относятся к моде, одежде, стилю, тканям и тд. Например, таги, используемые для подразделения “фэшн блог” на сайте fashion industry network <http://www.fashionindustry.com> .

- style
- news
- designers
- accessories
- trends
- designer
- fashion colors
- apparel
- clothing stylist
- sexy
- fashion industry
- shoes

- handbags
- swimwear
- models
- lingerie
- trend
- design

Сайт FIN также использует теги в разделе фэшн фотографий, видео и на форуме, чтобы помочь пользователям определить категорию, в которую будет добавлено их фото или комментариев.

Основное внимание в классификации информации по модному дизайну относится к тагированию изображений. Например, проект представляет собой парсер, используемый для парсинга одежды с фотографий, представленный на площадке Clothing Parsing <http://clothingparsing.com/>.

Студенты Университета штата Нью-Йорк в Стоуни-Брук продемонстрировали эффективный метод парсинга одежды на фотографиях с изображением людей [Yamaguchi 2016]. Данная задача является достаточно сложной из-за огромного количества существующих предметов одежды, элементов, вариантов конфигурации и наложения. В программу входит большой набор данных. Корпус фотографий, состоящий из 158,235 фотографий с сайта Chictoria.com. В основном на фотографиях изображен один человек, стоящий прямо, на относительно однотонном фоне, что является идеальным вариантом для изучения и проведения парсинга одежды. Для независимого визуального поиска одежды также была проведена работа правильной идентификации позы.

В связи с отсутствием исследований тагирования в предметной области модной одежды и значимостью этой области, мы решили создать программу для тагирования текстов.

## **2.2 Материал исследования**

Издательство Condé Nast одно из первых отреагировало на начало цифровой эры в журналистике и глянце, и уже в 1995 году запустило свои первые сайты в США. Материалом исследования послужил корпус текста, собранный с сайта издательства Vogue.com с 2017 по 2018 года, в предметной области дизайна модной одежды, в частности дизайнера Алессандро Микеле из модного дома Gucci. Сам сайт состоит из шести разных категорий контента: мода, культура, красота, радости жизни, показы, видео, магазин. Для подписчиков сайта доступен архив изданий Vogue с 1892 года.

Модный дом Gucci - итальянский дом моды, производитель одежды, парфюмерии, аксессуаров и текстиля. Принадлежит французскому конгломерату Kering. Дом моды «Гуччи» был основан Гуччио Гуччи во Флоренции в 1904 году. С 2015 креативным директором дома был назначен Алессандро Микеле.

Алессандро Микеле родился в Риме в 1972 году. Закончил Академию костюма и моды. Алессандро начал свою карьеру в модном доме Fendi как младший дизайнер аксессуаров. В 2002 году его нанял Том Форд для работы в офисе Gucci. С тех пор он проработал на разных позициях модного дома 12 лет.

Нами был собран корпус, состоящий из текстов категории «мода», подкатегории «дизайнеры» сайта Vogue.com. В корпус входят авторские тексты о новых коллекциях модного дома Gucci, их показах и деятельности Алессандро Микеле. Объем корпуса составляет 4200 словоупотреблений.

## **2.3 Разработка тагера для текстов по дизайну модной одежды**

### **2.3.1 Общая характеристика**

о предметной

—  
предметной области. Такая задача эффективно решается методами корпусной лингвистики, то есть путем

создания и исследования специализированного корпуса текстов, представляющего собой достаточный объем снабженных экспертной интерпретацией лингвистических данных, который может служить основой формирования системы автоматического анализа текстов, т.е. выступать в роли обучающего корпуса. В данном случае в состав корпуса текстов отбираются фрагменты из журнальной периодики посвященные определенной тематике. При создании специализированных корпусов текстов обычно производится лингвистическое аннотирование (морфологическое, синтаксическое), не зависящее от ПО и осуществляемое автоматически и/или вручную. Семантическая разметка, напротив, предметно ориентирована, поскольку определяется онтологией ПО и производится экспертами.

С помощью семантически аннотированного корпуса можно решать самые разные лингвистические задачи. В частности, такие корпуса широко используются для информационного поиска.

Задача, решаемая нашей программой – автоматическое тагирование текста из предметной области модной одежды (на материале собранного корпуса текста с сайта Vogue.com).

Цель программы – морфологическое и семантическое тагирование текстов на английском языке, подъязыке дизайна.

### **2.3.2 База знаний**

Схема семантического тагирования состоит из трех компонентов: базы знаний, алгоритма, программы. База знаний включает в себя список слов с присвоенными тегами.

Чтобы автоматизировать создание базы знаний мы использовали программу TagAnt. TagAnt – бесплатная программа для некоммерческого пользования, созданная Энтони Лоуренсом. Собранный корпус мы загрузили в программу TagAnt и получили на выходе список из слов с присвоенными к ним тегами. Для работы используется тагсет Penn Treebank II. Теги части речи

присваиваются одному слову в соответствии с его ролью в предложении. Тагсет Penn Treebank II выглядит следующим образом:

- **CC** - coordinating conjunction сочинительный союз
- **CD** - cardinal number количественное числительное
- **DT** – determiner определяющее слово
- **EX** - existential there слово there
- **FW** - foreign word иностранное слово
- **IN** - preposition/subord. conj. предлог/подчинительный союз
- **IN/that** – complementizer комплементатор
- **JJ** – adjective прилагательное
- **JJR** - adjective, comparative прилагательное в сравнительной степени
- **JJS** - adjective, superlative прилагательное в превосходной степени
- **LS** - list marker список
- **MD** – modal модельные глаголы
- **NN** - noun, singular or mass существительное
- **NNS** - noun plural существительное во множественном числе
- **NP** - proper noun, singular имя собственное, единственное число
- **NPS** - proper noun, plural имя собственное, множественное число
- **PDT** - predeterminer слово, предшествующее определяющему
- **POS** - possessive ending притяжательное ‘s
- **PP** - personal pronoun личное местоимение
- **PP\$** - possessive pronoun притяжательное местоимение
- **RB** – adverb наречие
- **RBR** - adverb, comparative сравнительное наречие
- **RBS** - adverb, superlative наречие в превосходной степени
- **RP** – particle частица
- **SENT** - end punctuation знак пунктуации
- **SYM** - symbol символы

- **TO** – *to* частица *to*
- **UH** – interjection восклицание
- **VB** - verb *be*, base form базовая форма глагола *be*
- **VBD** - verb *be*, past глагол *be* в прошедшем времени
- **VBG** - verb *be*, gerund/participle герундий/причастие от глагола *be*
- **VBN** - verb *be*, past participle причастие от глагола *be* в прошедшем времени
- **VBZ** - verb *be*, pres, 3rd p. глагол *be* в настоящем времени 3 лицо
- **VBP** - verb *be*, pres non-3rd p. глагол *be* в настоящем времени 1,2 лицо
- **VD** - verb *do*, base form базовая форма глагола *do*
- **VDD** - verb *do*, past прошедшая форма глагола *do*
- **VDG** - verb *do* gerund/participle герундий/причастие от глагола *do*
- **VDN** - verb *do*, past participle причастие от глагола *do* в прошедшем времени
- **VDZ** - verb *do*, pres, 3rd per.sing глагол *do* в настоящем времени 3 лицо, единственное число
- **VDP** - verb *do*, pres, non-3rd per. глагол *do* в настоящем времени 1,2 лицо, единственное число
- **VH** - verb *have*, base form базовая форма глагола *have*
- **VHD** - verb *have*, past глагол *have* в прошедшем времени
- **VHG** - verb *have*, gerund/participle герундий/причастие от глагола *have*
- **VHN** - verb *have*, past participle причастие от глагола *have* в прошедшем времени
- **VHZ** - verb *have*, pres 3rd per.sing глагол *have* в настоящем времени 3 лицо, единственное число
- **VHP** - verb *have*, pres non-3rd per.sing глагол *have* в настоящем времени 1,2 лицо, единственное число
- **VV** - verb, base form базовая форма глагола

- **VVD** - verb, past tense глагол в прошедшем времени
- **VVG** - verb, gerund/participle герундий/причастие
- **VVN** - verb, past participle причастие в прошедшем времени
- **VVP** - verb, present, non-3rd p. глагол в настоящем времени
- **VVZ** - verb, present 3d p. sing. глагол в настоящем времени
- **WDT** - wh-determiner слова на wh
- **WP** - wh-pronoun местоимение с wh
- **WP\$** - possessive wh-pronoun притяжательное местоимение с wh
- **WRB** - wh-abverb наречие с wh
- **:** - general joiner символ
- **\$** - currency symbol символ валюты

Далее был проанализирован весь протагированный корпус, были выявлены и исправлены ошибки, показанные в таблице 2.

Ошибки	Исправленный вариант
There's _NP	There's _EX
that_WDT	that_IN/that
surgeon's_NNS	surgeon's_POS
LED_JJ	LED_NN
Haraway's_NP	Haraway's_POS
Cyborg_NP	Cyborg_NN
Manifesto_NP	Manifesto_NN
Science_NP	Science_NN
Technology_NP	Technology_NN
Twentieth_NP	Twentieth_JJ
Century_NP	Century_NN
Of_IN course_NN	of course_RB
manga_NP	manga_NN

Instagram_NN	Instagram_NP
no_DT one_CD	No one_PI
PVC_NP	PVC_NN
et_CC al_NP	et al_FW
Diamanté_JJ	diamanté_FW
close-to-the-head_NN	close-to-the-head_JJ
zeitgeists_NNS	zeitgeists_FW
Michele's_NP	Michele's_POS
double-G_NN	double-G_JJ
Bailey's_NP	Bailey's_POS
label's_JJ	label's_POS
'80s_NN	'80s_NNS
M_NN M_NP	M/M_NP
Cofounder_NP	cofounder_NN
Social_JJ media_NNS	social media_NN
Anyone_NN	anyone_PI
Younger_NP	Younger_JJR
À_NP la_FW	À la_FW
out-of-the-box_NN	out-of-the-box_JJ
Matters_NNS	matters_VVZ
Gucci's_NP	Gucci's_POS
Fendi's_NP	Fendi's_POS
customer's_JJ	customer's_POS
bamboo-handle_NN	bamboo-handle_JJ
Guccio's_NP	Guccio's_POS
Everyone_NN	everyone_PI
New_NP	New_JJ



Beauty_NP	Beauty_NN
Dare_NP	Dare_NN
Go_NP	Go_VV
Girlie_NP	Girlie_NN
Goth_NP	Goth_NN
Resort_NP	Resort_NN
Antoinette-esque_NP	Antoinette-esque_JJ
marcel-meets-art-school_NN	marcel-meets-art-school_JJ
Wow_VVP	wow_UH
time-collapsing_VVG	time-collapsing_JJ
Arlesian_NP	Arlesian_JJ
Memento_NN mori_NN	Memento mori_FW
super-'70s_NNS	super-'70s_JJ
Chanel-esque_NP	Chanel-esque_JJ
Generation_Np	Generation_NN
today's_NP	today's_POS
C'était_NP un_VVZ Rendez-vous_JJ	C'était un Rendez-vous_FW
Lucan_JJ Gillespie_NP	Lucan Gillespie_NP
Gillespie's_NP	Gillespie's_POS
Vogue's_NP	Vogue's_POS
values_VVZ	values_NNS
in-demand_NN	in-demand_JJ
Burberry_NN	Burberry_NP

Таблица 2 – исправленные ошибки TagAnt

Мы пришли к решению создать свои собственные таги для различных семантических классов. Для этого была написана программа, вычленяющая

существительные из всего корпуса и тагирующая их в зависимости от семантического класса. Мы создали следующие категории: бренды, haute couture, обувной бренд, бренды аксессуаров, части одежды, тип одежды, верхняя одежда, креативные директора модных домов, главные редакторы Vogue, бывшие главные редакторы Vogue, цвет, цвет Pantone, платья, нижнее белье, обувь, свитера, рубашки, юбки, головные уборы. Текст будет тагироваться в зависимости от категорий, к которым принадлежат существительные. Для создания категории «бренды» был проанализирован сайт tsum.ru. На нем были собраны самые известные бренды. Для создания категории «haute couture» и «цвет» были проанализированы соответствующие страницы в Википедии. Категория «цвет Pantone» создавалась благодаря информации с вебсайта <http://www.pantone-colours.com>. Анализ словаря синонимов Collins Dictionary <https://www.collinsdictionary.com> помог собрать все слова из категорий «обувной бренд», «бренды аксессуаров», «части одежды», «тип одежды», «верхняя одежда», «платья», «нижнее белье», «обувь», «свитера», «рубашки», «юбки», «головные уборы».

В таблице 3 продемонстрированы таги, созданные для наших семантических классов.

Семантический класс	Тег
Бренды	_BR
Haute couture	_HC
Обувной бренд	_SB
Бренды аксессуаров	_FAB
Части одежды	_POC
Тип одежды	_TOC
Верхняя одежда	_OW
Креативные директора модных домов	_CR _EIC

Главные редакторы Vogue	_XEIC
Бывшие главные редакторы Vogue	_CLR _CLRP
Цвет	_DR
Цвет Pantone	_UW
Платья	_SH
Нижнее белье	_SW
Обувь	_SHRT
Свитера	_SKRT
Рубашки	_HAT
Юбки	
Головные уборы	

Таблица 3 – Семантические классы

Далее в таблицах 4 и 5 приведены фрагменты из слов, включенных в категорию «haute couture» и категорию «части одежды».

Adeline André	Alexandre Vauthier	Alexis Mabille
Anne Valérie Hash	Armani	Atelier Gustavolins
Azzedine Alaïa	Béatrice Demulder Ferrant	Bouchra Jarrar
Carlota Alfaro	Chanel	Christian Dior
Christophe Josse	Dior	Dominique Sirop
Elie Saab	Franck Sorbier	Giambattista Valli
Giorgio Armani	Givenchy	Hervé L. LeRoux
Iris van Herpen	Jean Paul Gaultier	Julien Fournié
Maison Martin Margiela	Maison Rabih Kayrouz	Marco Zanini
Maurizio Galante	On Aura Tout Vu by Yanna Samouilov and	Rad Hourani

	Livia Stoianova	
Ralph & Russo	Schiaparelli	Serkan Cura
Stéphane Rolland	Ulyana Sergeenko	Yanina Couture
Valentino	Valentin Yudashkin	Versace
Yiqing Yin	Zuhair Murad	

Таблица 4 – haute couture

armhole	armlet	bodice
buttonhole	collar	cuff
dicky	epaulette	founce
gusset	hem	hemline
hood	jabot	lapel
leg	lining	neckline
patch pocket	pocket	seam
shawl collar	shoulder	sleeve
tail	train	waist
waistline	yoke	

Таблица 5 – части одежды

Полный перечень слов, включенный в каждый семантический класс можно увидеть в приложении №1.

### **2.3.3 Алгоритм**

Сначала текст поступает в тагер, потом происходит поиск каждой словоформы в словаре и найденной словоформе присваивается соответствующий тег из словаря.

### **2.3.4 Программа**

Сначала мы открываем файл `DB_creator.py`, который поможет нам создать базу данных. Подключаем `'sqlite3'` как `'SQLite'`, подключаемся к базе данных. Создаем таблицу `'Lexicon'` со столбцами `'word'` и `'tag'`. `Word` = слово, к которому будут присваиваться таги (`tag`). Заполняем таблицу, сохраняем результат. Извлекаем таблицу и отключаемся от базы данных.

После заполнения базы данных, подключаем ее к программе. Создаем документ с текстом о модном доме Gucci с расширением `.txt` и сохраняем его в той же директории, что и остальные файлы. Запускаем файл `program.py`, подключаем `'sqlite3'` как `'SQLite'`, подключаем регулярные выражения, они необходимы для поиска слов в тексте. После запуска программы на входе пишем название документа (в нашем случае `Gucci_1.txt`). Программа подключается к созданной нами ранее базе данных, извлекает из нее слова и ищет их в тексте. Регулярное выражение `re.search()`, которое мы подключили в начале программы, ищет только первое совпадение в тексте. В переменную заносятся координаты, если эти координаты есть, то выполняется цикл `while`. Первая координата должна быть равна началу строки, последняя координата не равна букве или первая координата должна быть не равна букве или открывающейся скобке, а последняя не равна букве. Если слово нашлось, тогда обращаемся в базу данных для получения его тега. Находим границы (первую и последнюю буквы) слова в тексте, чтобы программа могла безошибочно добавить тег к слову. Если последний символ равен любому символу (но не букве), то программа его удаляет, если первый символ тоже равен любому символу (но не букве), то он удаляется и к последней координате добавляется тег. Если программа не находит символ (но не букву) слева, то сразу присваивается тег. Этот цикл продолжается до тех пор, пока программа не протагирует все слова в тексте. После завершения цикла выводится полностью протагированный текст. Программа, создающая базу данных, показана на рисунке 1.

```
DB_creator.py - /Users/marusiamak/Desktop/ЛПО/Экранацционн
# Подключаем 'sqlite3' как 'SQLite'.
import sqlite3 as SQLite

# Подключаемся к базе данных.
con = SQLite.connect('data.db')
cur = con.cursor()

# Создаём таблицу 'Lexicon' со столбцами 'word' и 'tag'.
#cur.execute('CREATE TABLE Lexicon(word, tag)')

#file = open('Gucci_1 words.txt', 'r', encoding = 'utf-8')
#text = file.read().split()

#for each in text:
#    #word = each.split('_')[0]
#    #tag = each.split('_')[1]
#    #cur.execute('INSERT INTO Lexicon VALUES(?, ?)', (word, tag))
#    #con.commit()

word = input('Word: ')
tag = input('Tag/tags: ')

# Заполняем таблицу.
cur.execute('INSERT INTO Lexicon VALUES(?, ?)', (word, tag))
# Сохраняем результат.
con.commit()

# Извлекаем таблицу из базы данных.
cur.execute('SELECT * FROM Lexicon')
data = cur.fetchall()

print(data)

# Отключаемся от базы данных.
con.close()

#file.close()
```

Рисунок 1 – код программы, создающей базу данных.

Код основной программы показан на рисунке 2.

```
import sqlite3 as SQLite
import re

file_name = input('Enter file name for POS-tagging: ')
file = open(file_name, 'r', encoding = 'utf-8')
text = file.read()

# Удаляем '\ufe00' из начала текста.
text = text[1:]

print('\n\t"Text before POS-tagging"')
print(text)

# Подключаемся к базе данных.
con = SQLite.connect('data.db')
cur = con.cursor()

# Извлекаем из базы данных слова.
cur.execute('SELECT word FROM Lexicon')
data = cur.fetchall()

stop = ('"', "'", ' ', '-', '\n')

# Ищем извлечённые слова в тексте.
for each in data:
    searcher = re.search(r'^'+each[0]+'[^a-z_]|[^a-z_]'+each[0]+'[^a-z_]', text.lower()) # регулярное выражение,
    while searcher:

        cur.execute('SELECT tag FROM Lexicon WHERE word=' + "'" + each[0] + "'") # запрос в базу данных
        data = cur.fetchall()

        first_letter = searcher.span()[0]
        last_letter = searcher.span()[-1]

        # Находим границы( первую и последнюю буквы) слова в тексте.
        if text[last_letter - 1] in stop: # справа
            last_letter -= 1 # если первая буква = пробелу то -1
            if text[first_letter] in stop: # слева
                first_letter += 1 # если последняя буква = пробелу то +1
            # К последней границе прбавляем тег.
            text = text[:last_letter] + '_' + data[0][0] + text[last_letter:]
            searcher = re.search(r'^'+each[0]+'[^a-z_]|[^a-z_]'+each[0]+'[^a-z_]', text.lower())
        else:
            # К последней границе прбавляем тег. если не нашел в начале пробел
            text = text[:last_letter] + '_' + data[0][0] + text[last_letter:]
            searcher = re.search(r'^'+each[0]+'[^a-z_]|[^a-z_]'+each[0]+'[^a-z_]', text.lower())
```

```

searcher = re.search(r'^(?=[^a-z_]+[a-z_])', text.lower())
else:
    if text[first_letter] in stop:
        first_letter += 1 # удаление пробела
        #К последней границе прбавляем тег.
        text = text[:last_letter - 1] + '_' + data[0][0] + text[last_letter - 1:]
        searcher = re.search(r'^'+each[0]+'[Aa-z_]|[Aa-z_]'+each[0]+'[Aa-z_]', text.lower())
print('\n\t"Text after POS-tagging"')
print(text)

```

Рисунок 2 – код программы

После запуска программы на входе пишем название документа (в нашем случае Gucci\_1.txt).

```

Python 3.6.3 (v3.6.3:2c5fed86e0, Oct  3 2017, 00:32:08)
[GCC 4.2.1 (Apple Inc. build 5666) (dot 3)] on darwin
Type "copyright", "credits" or "license()" for more information.
>>>
== RESTART: /Users/marusiamak/Desktop/ЛПО/Экраниционная работа/program.py
Enter file name for POS-tagging: Gucci_1.txt|

```

Рисунок 3 – вход программы

На выходе получаем копию текста из текстового файла, который мы задали на входе и протагированный тренировочный текст. Выход программы показан на рисунке 4.



#### "Text before POS-tagging"

Last season of fashion shows was splendid. We could see models, wearing Chanel, Gucci, Marni, Moschino, Tom Ford, Valentino at the catwalks.

There were Russian babushka headscarves and modest, covered-up folk-costume dresses next to jewellery brands such as Fiorelli Fiorucci, Flowfold, Safor. A lot of slouch hats and pea jackets; floral white nightdress and Aran sweater. Gucci logos were everywhere.

Among the brands which represent couture we could see Versace, Ralph & Russo, Schiaparelli, Dior, Elie Saab, Franck Sorbier, Giambattista Valli.

All parts of clothes (we are talking about armhole, armlet, bodice, collar, cuff, epaulette) were way too colourful.

Despite of the fact that Pantone Institute chose Almond as a colour of the year, we could also see such colours as Shamrock

Caribbean Green, Pine Green, Robin Egg Blue, Aquamarine, Turquoise Blue.

Among the star guests we saw (of course) Carine Roitfeld, Anna Wintour, Edward Enniful and Emmanuelle Alt.

#### "Text after POS-tagging"

Last\_JJ season\_NN of\_IN fashion\_NN shows\_NNS was\_VBD splendid\_JJ. We\_PP could\_MD see\_VV models\_NNS, wearing Chanel\_NP\_BR\_HC, Gucci\_NP\_BR\_FAB, Marni\_NP\_BR, Moschino\_NP\_BR, Tom Ford\_NP\_BR\_SB, Valentino\_NP\_BR\_HC at\_IN the\_DT catwalks\_NNS.

There\_EX were\_VBD Russian\_JJ babushka\_NN\_HAT headscarves\_NNS\_POC and\_CC modest, covered-up\_RP folk-costume dresses\_NNS\_DR next to\_TO jewellery\_NN brands\_NNS such as\_IN Fiorelli\_NP\_BR\_FAB, Fiorucci\_NP\_BR\_FAB, Flowfold\_NP\_BR\_FAB, Safor\_NP\_BR\_FAB. A\_DT lot\_NN of\_IN slouch hats\_NNS\_HAT and\_CC pea jackets\_NNS\_TOC; floral white\_JJ\_CLR nightdress\_NN\_DR and\_CC Aran sweater\_NN\_TOC. Gucci\_NP\_BR\_HC logos\_NNS were\_VBD everywhere\_PI.

Among\_IN the\_DT brands\_NNS which\_WDT represent\_VV couture\_NN we\_PP could\_MD see\_VV Versace\_NP\_BR\_HC, Ralph & Russo\_NP\_BR\_HC, Schiaparelli\_NP\_BR\_HC, Dior\_NP\_BR\_HC, Elie Saab\_NP\_BR\_HC, Franck Sorbier\_NP\_BR\_HC, Giambattista Valli\_NP\_BR\_HC.

All\_DT parts of\_IN clothes\_NNS (we are\_VBP talking\_VVG about\_IN armhole\_NN\_POC, armlet\_NN\_POC, bodice\_NN\_POC, collar\_NN\_POC, cuff\_NN\_POC, epaulette\_NN\_POC) were\_VBD way\_NN too\_RB colourful\_JJ.

Despite of\_IN the\_DT fact\_NN that\_IN Pantone\_NN Institute\_NN chose\_VVD Almond\_NN\_CLR\_CLRP as\_IN a\_DT colour\_NN of\_IN the\_DT year\_NN, we\_PP could\_MD also\_RB see\_VV such colours as\_IN Shamrock\_NN\_CLRP, Caribbean Green\_NN\_CLRP, Pine Green\_NN\_CLRP, Robin Egg Blue\_NN\_CLRP, Aquamarine\_NN\_CLR\_CLRP, Turquoise Blue\_NN\_CLR\_CLRP.

Among\_IN the\_DT star\_NN guests\_NNS we\_PP saw\_VVD (of course\_RB) Carine Roitfeld\_NP\_XEIC, Anna Wintour\_NP\_XEIC, Edward Enniful\_NP\_XEIC and\_CC Emmanuelle Alt\_NP\_XEIC.

Рисунок 4 – выход программы

## Выводы по главе 2

Во второй главе был проанализирован материал из различных источников, который способствовал созданию семантических классов в предметной области дизайна модной одежды. Материал включает в себя словарь синонимов Collins, сайты Tsum.ru, Vogue.com, Wikipedia и Pantone. Были определены следующие классы: бренды, haute couture, обувной бренд, бренды аксессуаров, части одежды, тип одежды, верхняя одежда, креативные директора модных домов, главные редакторы Vogue, бывшие главные редакторы Vogue, цвет, цвет Pantone, платья, нижнее белье, обувь, свитера, рубашки, юбки, головные уборы.

Был создан морфологическо-семантический тагер, который позволяет определить, к какой части речи относится каждое слово в тексте и к какому семантическому классу относятся существительные.

С помощью семантически аннотированного корпуса можно решать самые разные лингвистические задачи. В частности, такие корпуса широко используются для информационного поиска.

Семантическая разметка является одной из самых сложных разметок. Она предметно ориентирована, поскольку определяется онтологией ПО и производится человеком.

## ЗАКЛЮЧЕНИЕ

В процессе данной дипломной работы были теоретически осмыслены понятия «корпус», «корпусные технологии», «аннотирование», «тагирование», «разметка», а также проанализированы способы, методы и области применения аннотирования. В ходе работы был собран корпус с сайта издательства Conde Nast Vogue.com в предметной области дизайна модной одежды, в частности дизайнера Алессандро Микеле из модного дома Gucci. Во второй главе описано создание собственного тагера, который позволяет определить, к какой части речи относится каждое слово в тексте и к какому семантическому классу относятся существительные. Был проанализирован материал из различных источников, который способствовал созданию семантических классов в предметной области дизайна модной одежды. Материал включает в себя словарь синонимов Collins, сайты Tsum.ru, Vogue.com, Wikipedia и Pantone. Были определены следующие классы: бренды, haute couture, обувной бренд, бренды аксессуаров, части одежды, тип одежды, верхняя одежда, креативные директора модных домов, главные редакторы Vogue, бывшие главные редакторы Vogue, цвет, цвет Pantone, платья, нижнее белье, обувь, свитера, рубашки, юбки, головные уборы. Была написана программа для реализации схемы семантического тагирования на языке программирования Python.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Андрющенко, В. М. Концепция и архитектура машинного фонда русского языка [Текст] / В. М. Андрющенко. – М.: Изд-во Наука, 1989. – 200 с.
2. Аннотированный корпус русских текстов: концепция, инструменты разметки, типы информации [Текст]/И.М. Богуславский  
длин, В.З. Санников, Н.Е. Фрид // Труды международного семинара по компьютерной лингвистике и ее приложениям «Диалог-2000» – под ред. А.С. Нариньяни. – Протвино: Изд-во Истина, 2000. – Т. 2 – С. 41-47.
3. Бабина, О.И. Автоматический морфологический анализ флективных языков [Текст] / О.И. Бабина, Н.Ю. Дюмин // Наука  
- научной конференции. Секции естественно-научных и гуманитарных наук. – Т. 2 – Челябинск: Издательский центр ЮУрГУ, 2010. – С. 38–44.
4. Баранов, А. Н. Введение в прикладную лингвистику [Текст]: учебное пособие / А.Н. Баранов. – М.: Изд-во Эдиториал УРСС, 2001. – 347 с.
5. Борискина, О. О Корпусное исследование языка: мода или необходимость [Текст] / О. О. Борискина // Вестник ВГУ. Сер. Лингвистика. – Воронеж, 2015. – №3. – С. 24–27.
6. Венцов, А. В. Словарь для модели восприятия речи / А. В. Венцов. – СПб.: Вести С.-Петербургского ун-та, 1998. – С. 32-39.
7. Виноградова, В. Б Компьютерный корпус текстов русских газет конца XX века: создание, категоризация, автоматизированный анализ языковых особенностей [Текст] / В. Б. Виноградова // «Русский язык: исторические судьбы и современность»: международный конгресс русистов-исследователей. – М.: Изд-во МГУ, 2001. – С. 398-399.

8. Громько, А. А. Образы России и Великобритании: реальность и предрассудки [Текст] / А. А. Громько. – М.: Изд-во Русский сувенир, 2008. – 96 с.
9. Ефремова, Т. Ф. Толковый словарь служебных частей речи русского языка / Т. Ф. Ефремова. – М.: Изд-во Астрель-АСТ, 2004. – 814 с.
10. Захаров, В. П. Корпусная лингвистика [Текст]: учебное пособие / В. П. Захаров, С.Ю. Богданова. – Иркутск: ИГЛУ, 2011. – 161 с.
11. Захаров, В. П. Корпусная лингвистика [Текст]: учебное пособие / В. П. Алефиренко. – СПб.: СПбГУ, 2005. – 48 с.
12. Михайлов, М. Н. Параллельные корпуса художественных текстов [Текст] / М. Н. Михайлов. – Тампере: Изд-во Унив-а Тампере, 2003. – 348 с.
13. Мустайоки, А. Принципы создания Хельсинкского аннотированного корпуса русских текстов ХАНКО в сети Интернет / Научно-техническая информация. – 2003. – Сер. 2, № 6. – С. 33-37.
14. Мустайоки, А. Теория функционального синтаксиса [Текст] / А. Мустайоки. – М.: Изд-во Языки славянской культуры, 2010. – 512 с.
15. Принципы синтаксической разметки Хельсинкского аннотированного корпуса русских текстов ХАНКО [Текст] / Мустайоки, А., Копотев М. В., Гурин Г. Б., Саломатина М. С. // Труды международной конференции «MegaLing'2005. Прикладная лингвистика в поиске новых путей». – СПб.: Изд-во Институт русского языка им. В. В. Виноградова, 2005. – С. 90-95.
16. Резникова, Т. И. Лингвистически аннотированные корпуса русского языка [Текст] / Т. И. Резникова. – М.: Индрик, 2005. – С. 31-59.
17. Рогожникова, Р. П. Толковый словарь сочетаний, эквивалентных слову / Р. П. Рогожникова. – М.: Изд-во Астрель-АСТ, 2003. – 416 с.
18. Рязанова, Е. А. Методика формирования грамматических навыков речи студентов на основе лингвистического корпуса [Текст] / Е. А. Рязанова. – Тамбов, 2012. – 146 с.

- 19.Чернякова, Т. А. Методика формирования навыков студентов на основе лингвистического корпуса [Текст] / Т. А. Чернякова. – Тамбов, 2012. – 149 с.
- 20.Betsch, M. Automatic Annotation of Russian texts: Evaluation of Different Tagging Methods / R.Meyer, P.Kosta, J. Blaszczak, J. Frasek, L. Geist, M. Zygis // Investigations into Formal Slavic Linguistics. Contributions of the Fourth European Conference on Formal Description of Slavic Languages (FDSLIV) held at Potsdam University, 2003. – 234.
- 21.Fillmore Ch. 'Corpus linguistics' vs. 'Computer-aided armchair linguistics' / Ch. Fillmore // Directions in Corpus Linguistics. Proceedings from a 1992 Nobel Symposium on Corpus Linguistics. – Berlin, New York, 1992. – Pp. 35-60.
- 22.Leech, G. Corpus annotation schemes / G. Leech Literary and Linguistic Computing, 1993. – 8/4. – P. 275-281.

#### Электронные ресурсы

- 23.Корпусная лингвистика. Курс лекций. – URL: <http://www.rykov-cl.narod.ru/>, свободный. – Загл. с экрана. – (Дата обращения: 18.04.2017).
- 24.Корпусные методы и применение корпусов в лингвистических исследованиях. – URL: <http://tpl-it.wikispaces.com/Корпусные+методы+и+применение+корпусов+в+лингвистических+исследованиях/>, свободный. – Загл. с экрана. – (Дата обращения: 20.12.2017).
- 25.Корпусы и корпусная лингвистика. – URL: <https://www.myfilology.ru/7/korpusy-i-korpusnaya-lingvistika-osnovnye-onytiya/>, свободный. – Загл. с экрана. – (Дата обращения: 20.03.2017).
- 26.A System for Automatic Recognition of Russian Words. – URL: <http://www.lingsoft.fi/doc/rustwol/rustwol.txt/>, свободный. – Загл. с экрана. – (Дата обращения: 08.04.2017).

27. Statistical Part-of-Speech Tagging. – URL: [www.coli.uni-sb.de/~thorsten/publications/Brants-TR-TnT.pdf](http://www.coli.uni-sb.de/~thorsten/publications/Brants-TR-TnT.pdf), свободный. – Загл. с экрана. – (Дата обращения: 15.01.2017).

## ПРИЛОЖЕНИЕ 1

### КАТЕГОРИЯ «БРЕНДЫ», ТЕГ \_BR

A La Russe	adidas by	Aigle	Alexander
A. Testoni	Stella	Aj Arabia	McQueen
A.F.Vandervo	McCartney	Ajmone	Alexander
rst	adidas	Akid	McQueen
A.W.A.K.E.	Originals	Akira Naka	Perfumes
Aalto	adidas	Akris Punto	Alexander
Acne	Originals by	Alaia	Terekhov
Studios	Alexander	Alan Paine	Alexander
Acqua di	Wang	Alanui	Wang
Parma	Aecawhite	Albert	Alexandra
Act n1	Aedes de	Thurston	Golovanoff
Adam	Venustas	Aldo Brue	Alexandre
Selman	Aesop	Aldo	Birman
Add	Affenzahn	Coppola	Alexandre
Addicted	Affex	Alejandro	De Paris
Aden+Anais	Ag	Ingelmo	Alexandre.J
adidas	Age of	Aleksandersi	Alexis
Kolor	Innocence	radekian	Bittar
adidas by	AGL	Alessandra	Alice +
Raf Simons	Agnelle	Rich	Olivia
adidas by	Agolde	Aletta	Alice
Rick Owens	Agonist	Alexachung	McCall

Allude	Ava Adore	Barbara I	Bilancioni
Altaia	Avery	Gongini	Billionaire
Altea	Aviator	Barbed	Billionaire
Alternà	Nation	Barrett	Boys Club
Altuzarra	Aviu	Barrett	Biotherm
Amen	B3.0	Black	Bitto
Ami	Baby T	Barrie	Blancha
Amina	Baccarat	Basix Black	Blauer
Rubinacci	Back Label	Label	Blaze
Amiri	Bacon	Be Positive	Milano
AMO	Badelaine	Beats	Blood
Amouage	Paris	Beberlis	Concept
Ancient	Bagutta	Bellefontaine	Blugirl
Greek	Baldan	Belstaff	Blumarine
Sandals	Balenciaga	Ben	Boadicea the
Andrea	Bally	Taverniti	Victorious
Campagna	Balmain	Unravel	Bobbi
Atkinsons	Balmain Hair	Project	Brown
Atlanta	Couture	Benjamins	Boglioli
Mocassin	Balmain	Benoit	Bogner
Atlantic	Kids	Missolin	Bogner
Stars	Balmuir	Berluti	Kids
Atlantique	Bao Bao	Bernardaud	Bogs
Ascoli	ISSEY	Bertoni	Bonfanti
Attico	MIYAKE	Berwich	Bordelle
Au Jour Le	Barba	Beth	Borgo De
Jour	Barbara	Richards	Nor
Aurora	Casasola	Beyond	Borsalino



BOSS	Calvin Klein	Castaner by	Citizens Of
Both	Jeans	Manolo	Humanity
Bottega	Calvin Klein	Blahnik	Cividini
Veneta	Swimwear	Catya	Clarins
Bowers &	Calvin Klein	Cedric	Clinique
Wilkins	Underwear	Charlier	Clive
BOYY	Cami NYC	Celestine	Christian
Brian	Canada	Céline	Cloud Nine
Atwood	Goose	Céline	Clu
Bric`s	Canali	Eyewear	Clube
Brioni	Candyshop	Cellcosmet&	Bossa
Burberry	Russia	Cellmen	Coach
Burlington	Cantarelli	CHADO	Coccinelle
Burri	Capobianco	Chanel	Colette
Milano	Carine	Charlotte	Malouf
Buscemi	Gilson	Olympia	Coliac
Buttero	Carolee	Charvet	College
BVLGARI	Carolina	Chepe	Colmar
By Malene	Herrera	Chloé	Colombo
Birger	Caron	Chopard	Colors of
Byredo	Carrera	Christopher	California
C.P.	Cartier	Kane	Come
Company	Caruso	Church's	Comme des
Caf	Carven	Cinemood	Garcons
CALVIN	Casadei	Cinzia	Comme des
KLEIN	Casasola	Araia	Garcons
205W39NY	Castaner	Cirri_Cirri	GIRL

C

Comme Des	Daniele	DESA	Dom Rebel
Garcons	Fiesoli	Design	Dorateymur
Homme	Daum	Letters	Dorin
Plus	David	Diadora	Dr. Brandt
Comme des	Charles	Heritage	Dr. Levy
Garcons	David	Diana	Dr.Jart+
Play	Koma	Broussard	Dr.Sebagh
Comme des	David	Diane Von	Dr.Vranjes
Garcons	Yurman	Furstenberg	Dries Van
SHIRT	De Carlis	Diego M	Noten
BOYS	De Fonseca	Diego	DRKSHDW
Common	Deakin&Fra	Venturino	DROME
Projects	ncis	Diemme	Drykorn
Comptoir	Declare	Diesel	Dsquared2
Sud	Dedicated	Dior	Ea 7
Pacifique	Deha	Diptyque	Eddie
Converse	Del Carlo	Dirk	Borgo
D.Exterior	Delpozo	Bikkembergs	Edit
Dal Lago	Denim X	Disney	Eduard
Dalood	Alexander	Dita	Dressler
Dami	Wang	Djaba	Edun
Damir	Denim&Sup	Diassamidze	Egg By
Doma	ply by Ralph	DKNY	Susan Lazar
Dana	Lauren	Dodo Bar	Eight &
Pisarra	Dennis	Or	Bob
Daniele	Basso	Dolce &	Elevenparis
Basta	Derek Lam	Gabbana	Eleventy
	Derek Rose	Dolfie	

Eleventy	Erickson	Etudes	Fendi
Platinum	Beamon	Eugenia	Fenty Puma
Eleventy	Erika	Kim	by Rihanna
UOMO	Cavallini	EVARAE	Filles A
Elie Saab	Ermanno	EviDenS de	Papa
Elie Tahari	Ermanno	Beaute	Filling
Elina	Scervino	Ex Nihilo	Pieces
Linardaki	Ermanno	Eytys	Fine Edge
ElioKap	Scervino	F.R.S.	Fioroni
Elizabeth	Ermenegildo	F66	Flagpole
and James	Zegna	Fabrizio Del	Fleur du Mal
Ella K	Escada	Carlo	X Kilian
Ellery	Escada	Fabrizio	Fleur of
Emilia	Sport	Viti	England
Wickstead	Escentric Mo	Facto	Florian
Emilio	lecules Esm	Faith	Denicourt
Conte	eralda	Connexion	Flower
Emilio	Resort	Falcotto	Mountain
Pucci	Esteban	Faliero	Fogal
Emporio	Cortazar	Sarti	For Art's
Armani	Estée	Falke	Sake
Epiqual	Lauder	Fausto	Forte Dei
Equipment	Estella	Puglisi	Marmi
Erborian	Bartlett	Fay	Couture
Erdem	Ethan K	Fay Junior	Forte_forte
Eres	Etincelles	Fedeli	Fracap
Eric Javits	Eton	FEMME by	Franco
	Etre Cecile	Michele R	Ferrari

Frankie	Gentle	Histoires de	Icons of
Morello	Monster	Parfums	California
Frapin	Germano	History	IENKI
Fratelli	GFase	Repeats	IENKI
Rossetti	GHC	Hitch-hiker	iKins
Frederic	Giamba	Hogan	Il Gufo
Malle	Giambattista	Homers At	Il Profvmo
Frenchy	Valli	Home	Il Trenino
Yummy	Gianantonio	Hood by	Ilaria Nistri
Frency&Mer	Paladini	Air	Illesteva
cury	Giannico	Houbigant	Imec
Frescobol	Gianvito	House Of	Incotex
Carioca	Rossi	Holland	Initio
FTC	Giorgio	House Of	INUIKII
Furla	Armani	Lafayette	Inverni
G-star RAW	Giorgio	House of	Ioanna
+ Aitor	Brato	Sillage	Deschamps
Throup	Helmut	Howlin	Paris
Gabriela	Lang	Hublot	Iphoria
Hearst	Heritage	Hudson	IQ+Berlin
Gallo	Hermès	HUGO	Iro
Gallucci	Herno	Hunting	Iro Anja
Ganni	Heroine	Season	Rubik
Garcon	Sport	I Pinco	Isa Arfen
Francais	Hide&Jack	Pallino	Isaac
Gem	Hillier	ICE	Sellam
Kingdom	Bartley	Iceberg	Isabel
Gemma. H	Hiltl		Benenato

Isabel	Jean Paul	Juicy	Keiko
Marant	Gaultier	Couture	Mecheri
Isabel	Jennifer	Julius	Kenzo
Marant	Behr	Junefee	Kissy Kissy
Etoile	Jennifer	Junya	Kiton
Isaia	Ouellette	Watanabe	Koché
Issey	Jenny	Jupe by	Kolor
Miyake	Packham	Jackie	Koral
Ivory & Co.	Jessie	Just Cavalli	Korres
J Brand	Western	Just Mobile	Ktz
J.F.	Jil Sander	K Jacques	Kub0raum
Lazartigue	Jimmy	Kappa	Kure
J.W.	Choo	Kontrol	Bazaar
Anderson	Jitrois	Karakoram	Kuxo
Jack Spade	JM Studio	accessories	Kyo
New York	Jog Dog	Karl	L A A I N
Jacob	John	Lagerfeld	L.B.M.
Cohen	Smedley	Kids	1911
Jacquemus	John	Kashja`	L.G.B.
Jacques	Varvatos	Cashmere	L'Artisan
Britt	Johnstons Of	Kate Moss x	Parfumeur
Jakioo	Elgin	Equipment	La
James Perse	Joseph	Kate Spade	Biosthetique
jane iredale	Joshua	New York	
Jane	Sanders	Катя	La
Woolrich	JOSHUA`S	Dobryakova	Collection
Jason Wu	KIDZ	Kazuyuki	Essence
	Jour/ne	Kumagai	La Colline

La Martina	Lemon	Maison	Marco de
La Mer	Jelly	Margiela	Vincenzo
La Perla	Lempelius	Maison	Marco
La Prairie	Lengling	Michel	Pescarolo
La reveche	Leoca	Maiyet	Marcoliani
La Ric	Leonard	malo	Margherita
Lafayette	Paris	Malone	Marina
148	Leonor	Souliers	Fossati
Lalique	Greyl	Man &	Mark Cross
Lamy	Les Coyotes	Wood	Markus
Lancaster	de Paris	Manebi	Lupfer
Lancome	Les	Manolo	Marlies
Lanvin	Hommes	Blahnik	Moller
Lanvin	Letasca	Mansur	Marlino
Contemporar y	Lila Eugenie	Gavriel Manuel	Marni Marsell
Larisa Barrera	Made Gold Madpax	Barcelo Manuel	Martiniano Mary
Laurence	Magda	Marte	Katrantzou
Dacade	Butrym	Mara	Marysia
Lazul	Magnanni	Hoffman	Marysia
Le Kasha	Maison	Marc	Bumby
Le Silla	Boinet	Jacobs	Masnada
Lele	Maison	Marcelo	Mastercraft
Sadoughi	Francis	Burlon	Union
Lemargo	Kurkdjian	Marcelo	Mastermind
LemLem	Maison Mangostan	Burlon Kids of Milan	X RTA

Mate	Michael Lo	Molli	Mykita
Official	Sordo	Moma	Mykita
Mattel	MICHAEL	Moncler	Mylon
Matthew	Michael	Moncler	N21#Kartell
Williamson	Kors	Genius	Naeem
Maurizzio	Michele	Monnalisa	Khan
Massimino	Binda	Monreal	Nancy
Max&Moi	MiH Jeans	London	Gonzalez
Maxverre	Mikimoto	Monrow	Napoleone
MCM	Mikimoto	Monster	Erba
MCQ	Cosmetics	Montale	Narciso
MD 75	Mimi	Montblanc	Rodriguez
Medical	Minna	Montcarotte	NARS
Beauty	Parikka	Montecore	Nasaseasons
Research	Mini	Montelpare	Natasha
Melissa	Mira Mikati	Tradition	Zinko
Melissa	Miriamqueve	Moorer	NATAYAKI
Odabash	do	Moose	M
Memo	Mischka	Knuckles	Native
Menard	Aoki	Mophie	Union
Meng	Missoni	Moreschi	Naturino
Mercury	Missouri	MRZ	Neil Barrett
Meteo Yves	Miu Miu	MSGM	Neil Barrett
Salomon	Mm6	Mugler	Kids
Mey	Mo&Co	Mukluks	New
Michael	MOA	Mulberry	Balance
Kors	Moka Da	Musher	Nicholas
Collection	Due	Muso	Kirkwood

Nike	Officine	Paco	Perris Monte
NikeLab	Generale	Rabanne	Carlo
Nilos	Oh My	Paige	Pertini
Nima	Gosh	Pal Zileri	Peter
Nina Donis	Olaplex	Palm	Pilotto
Nina Ricci	OLGANA	Angels	Petit Nord
Nine in the morning	Oliver	Paloma	Peuterey
NO KA'OI	Peoples	Barcelo	Philip
No. 21	Olivia Von	Panchic	Kingsley
Norma	Halle	Pantherella	Philip
Kamali	Olympia Le-	Paolo Pecora	Treacy
Norveg	Tan	Milano	Philipp
Norwegian	Omy	Parajumpers	Plein
Rain	One-T-Shirt		Philo
Not Shy	Oneteaspoon	Parfums de	Phytomer
Nove	Opening	Marly	Piacenza
NSF	Ceremony	Pasotti	Cashmere
Nude	Oribe	Ombrelli	1733
Numero 10	Ortigni	Pasquale	Pierre
NUNUNU	Oscar de la	Bruni	Hardy
O.X.S.	Renta	Patek	Pietro
Oamc	OSEREE	Philippe	Brunelli
Oeuf	Osman	Paul&Joe	Pitaka
Off-White	OXS rubber	Paul&Shark	Plein Sport
Officine	soul	Penhaligon's	PleinAir
Creative	Oxydo	Perofil	Poeme
	OYE	Perricone	Bohemien
		MD	



Polo Ralph	Pryma	Rebecca	ROBERT
Lauren	Headphones	Taylor	CLERGERI
Ports 1961	PT01	REDVALEN	E
Poustovit	Public	TINO	Robert
Prabal	School	Reebok	Piguet
Gurung	Puma	Reebok X	Roberto
Prada	Quis Quis	Cottweiler	Cavalli
Prada Linea	R13	Rejina Pyo	Roberto
Rossa	Rachel	Religion	Collina
PREEN by	Comey	Renco	Roberto
Thornton	Rachel Zoe	Rene	Ricci
Bregazzi	Racil	Caovilla	Designs
Premiata	Radical	Rene	Roberto
Premiata will	Chic	Furterer	Ricetti
be	Raf Simons	RH45	Robinson
Pretty	Rag&Bone	Rhea Costa	Les Bains
Ballerinas	Rails	Riccardo	Rocco P.
Prim by	Ralph &	Comi	Rochas
Michelle	Russo	Rick Owens	Rockins
Elie	Ralph	Rick Owens	Rodarte
Primaforma	Lauren	Lilies	Roeckl
Primordial is	Raparo	Ritratti	Roja Dove
Primitive	Raquel	Milano	Roland
Proenza	Allegra	Rivieras	Mouret
Schouler	Rasario	Leisure	Rolax
Profumum	Ray-Ban	Shoes	Ron Dorff
Roma	Re/Done	Roberi &	Rondinella
Prudence		Fraud	Roque

Rosantica	Sara	Simone	Sonia Rykiel
Rosetta	Battaglia	Rocha J	Enfant
Getty	sara roka	Brand	Sonix
Rosie	Sarah Jane	Simonetta	Sonrisa
Assoulin	Sartoria	Simonetta	Sophia
Saint	Latorre	Mini	Webster
Laurent	Sartoria	Simonetta	Sophia
Saks Potts	Tramarossa	Ravizza	Webster
Salvatore	Saryna Key	Simonnot-	kids
Ferragamo	Satechi	Godard	Sophie
Salvatore	Saucony	Sisley	Anderson
Santoro	SB-26	Skandia	Sorel
Sam	See by	Skin	Sorley
Sanayi313	Chloé	Smith and	Socks
Sand	self-portrait	Cult	SPEKTRE
Sanetta	Sensai	Smythson	Spider-man
Sanetta	Senscience	Solace	Spin Master
Fiftyseven	Sensi	Sold out	Spotlight
Sans-	Studio	Solingen	Sprung
Arcidet	Serapian	Soloviere	SPRWMN
Santoni	Serge	Somedaysoo	St. John
Santoni	Lutens	n	Steiger
edited by	SIMIMI	Sonia by	Stella Jean
Marco	Simon	Sonia	Stella Jean
Zanini	Miller	Rykiel	Kids
Santoni	Simon Scott	Sonia	Stella
Precious	Simone	Rykiel	McCartney
	Rocha		

Stephen	Tadashi	The Oozoo	Turnbull &
Webster	Shoji	The Row	Asser
Steve J &	Tadashi	The Sleep	Twelve
Yoni P	Shoji Kids	Shirt	South
Sticky-	Tak.Ori	The Spiders	Twins
Fudge	Talitha	from Arts	Florence
Stone	Tara	The Volon	Twinset
Island	Jarmon	Theodent	Two Women
Stuart	Tartine Et	Theory	In The
Weitzman	Chocolat	Three	World
Stussy	Tata Naka	Graces	uBear
Subterranei	Tateossian	Tiffany &	UGG
Sunday	Tegin	Co.	Ultracor
Somewhere	Ten C	Timberland	Ultru
Suno	Terekhov	Tingley	Uma Wang
Sunuva	Girl	Tod's	Umit Benan
Svevo	TFK The	Tom Ford	Un_furtive
Swarovski	Fragrance	Tomas	Undercover
Swims	Kitchen	Maier	Unfortunate
Swiss Line	Thakoon	Tome	portrait
Swiss Smile	The	Tommy	Unlabel
Swissgetal	Beautiful	Hilfiger	Urban
T by	Mind	Tse	Decay
Alexander	The Editor	TSUM	Valentino
Wang	The Elder	Collection	Valextra
Tabitha	Statesman	Tu es mon	Valmont
Simmons	The House Of Oud	TRESOR	Van Laack

Vanda	Vintage	YGGY	(+)People
Catucci	Shades	Yohji	0711
Vanessa	Vionnet	Yamamoto	10 Crosby
Seward	Vipe	YOLKE	Derek Lam
Vannina	Vita Fede	Young	10sei0otto
Vesperini	Vivetta	Soles	11 by Boris
Varley	Vivienne	YSL	Bidjan
Versace	Westwood	Yves	Saberi
Versus	Voile	Salomon	111SKIN
Versace	Blanche	Yves	12 Francais
Vertu	Voodoo	Salomon	Parfumeurs
Vetements	Jewels	Enfant	120% Lino
Vic Matie	W.Gibbs	Z Zegna	2 Men
Victoria	Walk of	Zac Posen	Jeans
Beckham	Shame	Zaccone	2Mestyle
Victoria,	Weill	Zadig&Volta	3,0 Berlin
Victoria	Wildfox	ire	3.1 Phillip
Beckham	William	Zegna	Lim
Victoria/To	Sharp	Couture	3LAB
mas	Windsor	Ziggy Chen	3x1
Vika	Wolford	Zilli	5PREVIEW
Gazinskaya	Woolrich	Zimmerli	
Viktor&Rolf	Xr	Zimmerman	
Vilebrequin	Y-3	n	
Vilshenko	Yeah Right	Zuhair	
Vince	Yeezy	Murad	

## КАТЕГОРИЯ «ОБУВНОЙ БРЕНД», ТЕГ \_SB

Adidas	Colchester Rubber Co.	Givova
Airness	Cole Haan	Globe International
Alden Shoe Company	Columbia Sportswear	Golden Goose Deluxe
Allen Edmonds	Common Projects	Brand
Anta Sports	Crocs	Goodwill Shoe
Ari Football	Damani Dada	Company
Ariat	DC Shoes	Grand Sport Group
ASICS	Diadora	Gravati
Atletica	Dr. Martens	Havaianas
Avia	Duarig	Heelys
Bally Shoe	Dunlop Sport	Herbert Levine
Banana Republic	DVS Shoes	Hi-Tec
Bata Shoes	ECCO	Humanic
Belle International	Edward Green Shoes	The Hundreds
Bontoni	EMU Australia	Hush Puppies
British Knights	Erreà	Insolia
Brooks Sports	Etnies	Invicta
Bucketfeet	Falc	Irregular Choice
Caleres	Fallen Footwear	Jello Shoecompany
Calvin Klein	FBT	Johnston & Murphy
Camper	FEIT	KangaRoos
Caterpillar Inc.	Feiyue	Kappa
Chaco	Fila	Kashi Kicks
Charles David	John Fluevog	Kat Maconie
ChenOne	Foot Locker	Keds
Church's	Foreva	Kickers
Circa	Geox	Kinney Shoes

Kubba Sportswear	Pearl Izumi	Starbury
LA Gear	Pediped Footwear	Start-rite
Leder und Schuh	PF Flyers	Stefano Bemer
Legea	Pony International	Superga
Lescon	Puma	Thom McAn
Li-Ning	Real United	Tod's
Loake	Red or Dead	Toms Shoes
John Lobb Bootmaker	Reebok	Tredair
Lotto Sport Italia	Reebok Ventilator	Troentorp Clogs
Lugz	Reef	UGG
M.Dia	Rockport	Unification Shoes
Steve Madden	Rossi Boots	United Nude
Melville Shoe Corporation	Rupani Footwear	Vans
Merrell	Russell & Bromley	Veja Sneakers
Montrail	SAS	Veritas Bespoke
Moon Boot	Saucony	Volley
Munich	Sebago	Vratim
Naot	Servis Shoes	Stuart Weitzman
New Balance	Sessions	J. M. Weston
Nine West	ShoeDazzle	Wildsmith Shoes
Nocona Boots	Shy	Xtep
Vanessa Noel	Skechers	Yohji Yamamoto
The Original Car Shoe	SLAM	Yull shoes
Original Penguin	Paul Smith	Giuseppe Zanotti
Osiris Shoes	Solovair	
Pacific Brands	SOURCE	
	Source Sandals	

**КАТЕГОРИЯ «БРЕНДЫ АКСЕССУАРОВ», ТЕГ \_FAB**

Ally Capellino	Countess Mara	Harveys Seatbelt Bags
Armani	DKNY	Daniel Hechter
Aspinal of London	Dooney & Bourke	Hermès
Asprey	Drake's	Hervé Chapelier
Dominique Aurientis	Benoît-Pierre Emery	Hield
Balenciaga	Enfants Riches	John Varvatos
Banana Republic	Déprimés	JOOP!
BeachMint	English Retreads	Alfred Karram
Belber	Fabletics	Kate Spade New York
Belen Echandia	Fastrack	Kenneth Cole
Bellroy	Fauré Le Page	Productions
Benetton Group	FEED Projects	Kenzo
Blanc & Eclare	Finnigans	Kooba
Vera Borea	Fiorelli	Kygo Life AS
Bottega Veneta	Fiorucci	La Perla
BP Studio	Flowfold	Lacoste
Bulgari	Fossil Group	Lancel
Burberry	Frey Wille	Lanvin
Cambridge Satchel	Fuchsia	Guy Laroche
Company	Gap Inc.	Léo Marciano
Pierre Cardin	General Eyewear	LeSportsac
Carrera y Carrera	Emily Gimmel	Lewis Leigh
Dean and Dan Caten	Giusto Manetti	Accessories
Cerruti	Battiloro	Longchamp
Chanel	Givenchy	Luxe Link
Charvet Place Vendôme	Gottex	Luxottica
Chrome Hearts	Goyard	M0851
Coach New York	Gucci	Maison Maquet

Maison Margiela	Roberto Cavalli	The Row
Malo	Roberto Cavalli	Tiffany & Co.
Marchesa	Rodarte	Rafé Totengco
Michael Kors	Rosetti	Tous
Miche Bag Company	Christian Roth	ToyWatch
Miu Miu	Sonia Rykiel	The Trump
Mouawad	Safilo Group	Organization
Moynat	Safor	Valentino SpA
Mulberry	Samantha Thavasa	Verdura
MZ Wallace	Scappino	Viktor & Rolf
Ted Noten	Sermoneta Gloves	Louis Vuitton
Old Navy	Shiatzy Chen	WonderLuk
Orient Watch	Sljeme	Yoshida & Co., Ltd.
Oroton	Paul Smith	David Yurman
Parfois	Stefano Ricci	Yves Saint Laurent
Petunia Pickle Bottom	Strathberry	Giuseppe Zanotti
Pomellato	Tally Weijl	Ermenegildo Zegna
Ports International	The Daily Edited	Zlatarna Celje

### **КАТЕГОРИЯ «ТИП ОДЕЖДЫ», ТЕГ \_ТОС**

academic dress	beachwear	boubou
apron	bib and brace	braces
armour	bikini	bustier
babywear	black tie	canonicals
baldric	blouse	cardigan
basque	body	chapeau
bathing suit	body stocking	chausses
bathrobe	bodysuit	cilice



civvies	Highland dress	mantle
civies	hose	millinery
clericals	hose	mitten
coat	hosiery	morning dress
coatee	housecoat	muff
codpiece	jacket	mufti
coordinates	jerkin	neckwear
coveralls	jubbah	negligee or negligée
cummerbund or kummerbund	jump suit	nightclothes
dolman	jumper	nightshirt
dress	jupon	nightwear
dressing gown	kaftan	overall
dungarees	caftan	overcoat
evening dress	kameez	overgarments
fancy dress	kanzu	overskirt
fatigues	kaross	oversleeve
frock	kilt	paletot
froufrou	kimono	pallium
galluses	kittel	partlet
gambeson	knitwear	pashmina
garter	leotard	peignoir
gilet	lingerie	plaid
glove	livery	pullover
gown	loincloth	pyjamas
haik	breechcloth	robe
halter	long-coats	rompers
hauberk	maillot	sackcloth
	manteau	samfoo

sanbenito	sock	necktie
sari or saree	sporrان	tights
sarong	sportswear	hose
sash	surcoat	toga
separates	swaddling clothes	tunic
serape	sweater	undergarment
shalwar	swimming trunks	underthings
shawl	trunks	underwear
shift	swimsuit	uniform
shirt	swimwear	waistcoat
shoe	T-shirt	weepers
shorts	tee shirt	white tie
skirt	tallit	widow's weeds
skivvies	tanga	wrap
skivvy	tank top	wrapper
slops	thong	yashmac
smock	tie	yashmak

### **КАТЕГОРИЯ «ВЕРХНЯЯ ОДЕЖДА», ТЕГ \_OW**

afghan	coatee	duster coat
balmacaan	cope	fearnought
bathrobe	covert coat	fearnaught
box coat	cutaway	frock coat
cape	dolman	fur coat
capote	domino	gabardine
capuchin	dreadnought	greatcoat
cardinal	dreadnaught	hacking jacket
chesterfield	dress coat	himation
coat dress	duffel coat	housecoat

Inverness	newmarket	roquelaure
jellaba	opera cloak	sheepskin jacket
jellabah	opera hood	sherwani
Jodhpuri coat	overcoat	slicker
joseph	paletot	snorkel
mac	pea jacket	spencer
mack	peacoat	surcoat
Mackinaw coat	parka	surtout
Mackintosh	pelisse	tailcoat
macintosh	peplum	tails
manta	poncho	tippet
mantelet	Prince Albert	topcoat
mantilla	raglan	trench coat
mantle	raincoat	ulster
morning coat	redingote	undercoat

### **КАТЕГОРИЯ «КРЕАТИВНЫЕ ДИРЕКТОРА» ТЕГ \_CR**

Alexander McQueen – Sarah Burton\_CRALE

Balenciaga – Demna Gvasalia\_CRBAL

Balmain – Olivier Rousteing\_CRBALM

BottegaVeneta – Daniel Lee\_CRBOT

Brioni – Nina-Maria Nitsche\_CRBRI

Burberry – Christopher Bailey\_CRBUR

Calvin Klein – Raf Simons\_CRCAL

Celine – Hedi Slimane\_CRCEL

Chanel – Karl Lagerfeld\_CRCHA

Chloe – Natacha Ramsay-Levi\_CRCHL

Christopher Kane – Christopher Kane\_CRCHR

Diane Von Furstenburg – Nathan Jenden\_CRDIA  
Dior – Maria Grazia Chiuri\_CRDIO, Kris Van Assche\_CRDIO  
Dolce&Gabbana – Domenico Dolce, Stefano Gabbana\_CRDOL  
Dries Van Noten – Dries Van Noten\_CRDRI  
Dsquared2 – Dean Caten, Dan Caten\_CRDSQ  
Escada – Niall Sloan\_CRESC  
Fendi - Karl Lagerfeld\_CRFEN  
Givenchy – Clare Waight Keller\_CRGIV  
Gorgio Armani – Gorgio Armani\_CRGOR  
Gucci – Alessandro Michele\_CRGUC  
Isabel Marant – Isabel Marant\_CRISA  
Jeremy Scott – Jeremy Scott\_CRJER  
Jil Sander – Luke Meier, Lucie Meier\_CRJIL  
Jimmy Choo – Sandra Choi\_CRJIM  
John Galliano – Bill Gaytten\_CRJOH  
Kate Spade New York – Nicola Glass\_CRKAT  
Kenzo – Carol Lim\_CRKEN, Humberto Leon\_CRKEN  
Lanvin – Olivier Lapidus\_CRLAN  
Loewe – Jonathan Anderson\_CRLOE  
Louis Vuitton – Nicolas Ghesquiere\_CRLOU, Virgil Abloh\_CRLOU  
Maison Margiela – John Galliano\_CRMAI  
Marc Jacobs – John Targon\_CRMARC  
Marni – Francesco Risso\_CRMAR  
Michael Kors – Michael Kors\_CRMIC  
Miu Miu – Miuccia Prada\_CRMIU  
Moncler – Remo Ruffin\_CRMON  
Moschino – Jeremy Scott\_CRMOS  
Mugler – Casey Cadwallader\_CRMUG

№21 – Alessandro Dell’Acqua\_CR№21  
Off-white – Virgil Abloh\_CROFF  
Oscar de la Renta – Laura Kim\_CROSC, Fernando Garcia\_CROSC  
Philipp Plein – Philipp Plein\_CRPHI  
Prada – Miuccia Prada\_CRPRA  
Raf Simons – Raf Simons\_CRRAF  
Ralph Lauren – Kevin Carrigan\_CRRAL  
Roberto Cavalli – Paul Surridge\_CRROB  
Saint Laurent Paris – Anthony Vaccarello\_CRSAI  
Salvatore Ferragamo – Paul Andrew\_CRSAL  
Stella McCartney – Stella McCartney\_CRSTE  
Tod’s – Andrea Incontri\_CRTOD  
Tom Ford – Tom Ford\_CRTOM  
Valentino – Pierpaolo Piccioli\_CRVAL  
Versace – Donatella Versace\_CRVER  
Vetements – Demna Gvasalia\_CRVET

**КАТЕГОРИЯ «ГЛАВНЫЕ РЕДАКТОРЫ VOGUE», ТЕГ \_EIC**

Australia – Edwina McCann\_EIC  
USA – Anna Wintour\_EIC  
UK – Edward Enniful\_EIC  
France – Emmanuelle Alt\_EIC  
Italy – Emanuele Farneti\_EIC  
Brazil – Silvia Rogar\_EIC  
Germany – Christiane Arp\_EIC  
Spain – Eugenia de la Torriente\_EIC  
Singapore – Nancy Pilcher\_EIC  
South Korea – Myung Hee Lee\_EIC  
Taiwan – Sky Wu\_EIC

Russia – Masha Fyodorova\_EIC  
Japan – Mitsuko Watanabe\_EIC  
Mexico & Latin America – Karla Martinez\_EIC  
Greece – Elena Makri\_EIC  
Portugal – Sofia Lucas\_EIC  
China – Angelica Cheung\_EIC  
India – Priya Tanna\_EIC  
Turkey – Seda Domanic\_EIC  
Netherlands – Karin Swerink\_EIC  
Thailand – Kullawit Laosuksri\_EIC  
Ukraine – Olga Sushko\_EIC  
Arabia – Manuel Arnaut\_EIC  
Poland – Filip Niedenthal\_EIC  
Czech Republic & Slovakia – Andrea Behounkova\_EIC

**КАТЕГОРИЯ «БЫВШИЕ ГЛАВНЫЕ РЕДАКТОРЫ VOGUE»,  
ТЕГ \_XEIC**

EX USA	Elizabeth Penrose_XEIC
Josephine Redding_XEIC	Audrey Withers_XEIC
Marie Harrison_XEIC	Ailsa Garland_XEIC
Edna Woolman Chase_XEIC	Beatrix Miller_XEIC
Jessica Daves_XEIC	Anna Wintour_XEIC
Diana Vreeland_XEIC	Liz Tilberis_XEIC
Grace Mirabella_XEIC	Alexandra Shulman_XEIC
EX UK	EX FRANCE
Elsbeth Champcommunal_XEIC	Cosette Vogel_XEIC
Dorothy Todd_XEIC	Main Bocher_XEIC
Alison Settle_XEIC	Michel de Brunhoff_XEIC

Edmonde Charles-Roux\_XEIC  
Françoise de Langlade\_XEIC  
Francine Crescent\_XEIC  
Colombe Pringle\_XEIC  
Joan Juliet Buck\_XEIC  
Carine Roitfeld\_XEIC  
EX AUSTRALIA  
Rosemary Cooper\_XEIC  
Sheila Scotter\_XEIC  
Eve Harman\_XEIC  
June McCallum\_XEIC  
Nancy Pilcher\_XEIC  
Marion Hume\_XEIC  
Juliet Ashworth\_XEIC  
Kirstie Clements\_XEIC  
EX ITALY  
Consuelo Crespi\_XEIC  
Franco Sartori\_XEIC  
Franca Sozzani\_XEIC  
EX BRAZIL  
Luis Carta\_XEIC

Andrea Carta\_XEIC  
Patricia Carta\_XEIC  
Daniela Falcão\_XEIC  
EX SPAIN  
Luis Carta\_XEIC  
Yolanda Sacristán\_XEIC  
EX RUSSIA  
Aliona Doletskaya\_XEIC  
Victoria Davydova\_XEIC  
EX JAPAN  
Hiromi Sogo\_XEIC  
EX MEXICO & LA  
Eva Hughes\_XEIC  
Kelly Talamas\_XEIC  
EX PORTUGAL  
Paula Mateus\_XEIC  
EX UKRAINE  
Masha Tsukanova\_XEIC  
EX ARABIA  
Deena Alijuhani Abdulaziz\_XEIC

### **КАТЕГОРИЯ «ЦВЕТ», ТЕГ \_CLR**

Absolute Zero	Air Force blue	Alloy orange
Acajou	Air superiority blue	Almond
Acid green	Alabama crimson	Aluminum
Aero	Alabaster	Aluminum foil
Aero blue	Alice blue	Amaranth
African violet	Alizarin crimson	Amaranth deep purple

Amaranth pink	Apricot	Banana Mania
Amaranth purple	Aqua	Banana yellow
Amaranth red	Aquamarine	Bangladesh green
Amazon	Arctic lime	Barbie Pink
Amber	Argent	Barn red
American blue	Army green	Battery charged blue
American bronze	Artichoke	Battleship grey
American brown	Arylide yellow	Bayside
American gold	Asda Green	Bazaar
American green	Asda Orange	Beau blue
American orange	Ash gray	Beaver
American pink	Asparagus	Beer
American purple	Ateneo blue	Begonia
American red	Atomic tangerine	Beige
American rose	Auburn	B'dazzled blue
American silver	Aureolin	Big dip o'ruby
American violet	Avocado	Big Foot Feet
American yellow	Awesome	Bisque
Amethyst	Axolotl	Bistre
Anti-flash white	Azure	Bistre brown
Antique brass	Azure mist	Bitter lemon
Antique bronze	Azureish white	Bitter lime
Antique fuchsia	Baby blue	Bittersweet
Antique ruby	Baby blue eyes	Bittersweet shimmer
Antique white	Baby pink	Black
Ao	Baby powder	Black bean
Apple	Baker-Miller pink	Black chocolate
Apple green	Ball blue	Black coffee



Black coral	Bole	Bronze
Black leather jacket	Bondi blue	Bronze
Black olive	Bone	Bronze
Black Shadows	Bored accent green	Bronze yellow
Blanched almond	Bottle green	Brown
Blast-off bronze	Boysenberry	Brown-nose
Bleu de France	Boy red	Brown sugar
Blizzard blue	Brandeis blue	Brown Chocolate
Blond	Brandy	Brown Coffee
Blood orange	Brass	Brown Yellow
Blood red	Brave orange	Brunswick green
Blood	Brick red	Bubble gum
Blue	Bright cerulean	Bubbles
Blue bell	Bright gray	Bud green
Blue bolt	Bright green	Buff
Blue cola	Bright lavender	Bulgarian rose
Blue-gray	Bright lilac	Burgundy
Blue-green	Bright maroon	Burlywood
Blue jeans	Bright navy blue	Burnished brown
Blue lagoon	Bright pink	Burnt orange
Blue-magenta violet	Bright turquoise	Burnt sienna
Blue raspberry	Bright ube	Burnt umber
Blue sapphire	Bright yellow	Button blue
Blue-violet	Brilliant azure	Byzantine
Blue yonder	Brilliant lavender	Byzantium
Blueberry	Brilliant rose	Cadet
Bluebonnet	Brink pink	Cadet blue
Blush	British racing green	Cadet blue

Cadet grey	Carmine pink	Charcoal
Cadmium blue	Carmine red	Charleston green
Cadmium green	Carnation pink	Charm
Cadmium orange	Carnelian	Charm pink
Cadmium purple	Carolina blue	Chartreuse
Cadmium red	Carrot orange	Cheese
Cadmium yellow	Castleton green	Cherry
Cadmium violet	Catalina blue	Cherry blossom pink
Café au lait	Catawba	Chestnut
Café noir	Cedar Chest	Chickadee
Cal Poly Pomona green	Ceil	China pink
Calamansi	Celadon	China rose
Cambridge blue	Celadon blue	Chinese black
Camel	Celadon green	Chinese blue
Cameo pink	Celeste	Chinese bronze
Camouflage green	Celestial blue	Chinese brown
Canary	Celtic blue	Chinese green
Canary yellow	Cerise	Chinese gold
Candy apple red	Cerise pink	Chinese orange
Candy pink	Cerulean	Chinese pink
Capri	Cerulean blue	Chinese purple
Caput mortuum	Cetacean blue	Chinese red
Caramel	Cerulean frost	Chinese red
Cardinal	CG blue	Chinese silver
Caribbean green	CG red	Chinese violet
Caribbean blue	Chamoisee	Chinese white
Carmine	Champagne	Chinese yellow
Carmine	Champagne pink	Chlorophyll green

Chocolate brown	Copper	Cyan
Chocolate cosmos	Copper	Cyan azure
Chocolate Kisses	Copper penny	Cyan-blue azure
Chocolate	Copper red	Cyan cobalt blue
Chocolate	Copper rose	Cyan cornflower blue
Chrome aluminum	Coquelicot	Cyan
Chrome yellow	Coral	Cyber grape
Cinereous	Coral pink	Cyber yellow
Cinnabar	Coral red	Cyclamen
Cinnamon	Coral reef	Daffodil
Cinnamon Satin	Coral Reef	Dandelion
Citrine	Cordovan	Dandelion
Citrine Brown	Corn	Dark blue
Citron	Cornell red	Dark blue-gray
Claret	Cornflower blue	Dark bronze
Classic rose	Cornflower	Dark bronze
Cobalt blue	Cornsilk	Dark brown
Cocoa brown	Cosmic cobalt	Dark brown-tangelo
Coconut	Cosmic latte	Dark byzantium
Coconut White	Coyote brown	Dark candy apple red
Coffee	Cotton candy	Dark cerulean
Cola	Cream	Dark charcoal
Columbia Blue	Crimson	Dark chestnut
Conditioner	Crimson glory	Dark chocolate
Congo pink	Crimson red	Dark chocolate
Cool black	Crystal	Dark cornflower blue
Cool grey	Crystal blue	Dark coral
Cookies and cream	Cultured	Dark cyan

Dark electric blue	Dark puce	Deep carrot orange
Dark gold	Dark purple	Deep cerise
Dark goldenrod	Dark raspberry	Deep champagne
Dark gray	Dark red	Deep chestnut
Dark green	Dark salmon	Deep coffee
Dark green	Dark scarlet	Deep dumpling
Dark gunmetal	Dark sea green	Deep fuchsia
Dark imperial blue	Dark sienna	Deep green
Dark jungle green	Dark silver	Deep green-cyan
Dark khaki	Dark sky blue	turquoise
Dark lava	Dark slate blue	Deep jungle green
Dark lavender	Dark slate gray	Deep koamaru
Dark lemon lime	Dark spring green	Deep lemon
Dark liver	Dark tan	Deep lilac
Dark magenta	Dark tangerine	Deep magenta
Dark medium gray	Dark taupe	Deep maroon
Dark midnight blue	Dark terra cotta	Deep mauve
Dark moss green	Dark turquoise	Deep moss green
Dark Navy	Dark vanilla	Deep peach
Dark olive green	Dark violet	Deep pink
Dark orange	Dark yellow	Deep puce
Dark orchid	Dartmouth green	Deep red
Dark pastel blue	Davy's grey	Deep ruby
Dark pastel green	Debian red	Deep saffron
Dark pastel purple	Deep amethyst	Deep sky blue
Dark pastel red	Deep aquamarine	Deep Space Sparkle
Dark pink	Deep carmine	Deep spring bud
Dark powder blue	Deep carmine pink	Deep taupe

Deep Tuscan red	Durian Yellow	Emerald green
Deep violet	Dust storm	Eminence
Deer	Dutch white	English green
Delicate girl blue	Earhtone	English lavender
Denim	Earth yellow	English red
Denim blue	Ebony	English vermilion
Desaturated cyan	Ecru	English violet
Desert	Eerie black	Eton blue
Desert sand	Eggplant	Eucalyptus
Desire	Eggshell	Facebook Blue
Diamond	Egyptian blue	Fallow
Dim gray	Elderberry	Falu red
Dingy Dungeon	Electric blue	Fandango
Dirt	Electric brown	Fandango pink
Dirty brown	Electric crimson	Fashion fuchsia
Dirty white	Electric cyan	Fawn
Dodger blue	Electric green	Feldgrau
Dodie yellow	Electric indigo	Feldspar
Dogwood rose	Electric lavender	Fern green
Dollar bill	Electric lime	Ferrari red
Dolphin gray	Electric orange	Field drab
Douban Blue	Electric pink	Fiery rose
Douban Green	Electric purple	Firebrick
Douban Light Yellow	Electric red	Fire engine red
Drab	Electric ultramarine	Fire opal
Duke blue	Electric violet	Flame
Durian	Electric yellow	Flamingo pink
Durian White	Emerald	Flattery

Flavescent	Forest green	French sky blue
Flax	Forest green	French violet
Flesh	French beige	French wine
Flickr Blue	French bistre	Fresh Air
Flickr Pink	French blue	Frostbite
Flirt	French fuchsia	Fuchsia
Floral white	French lilac	Fuchsia
Flower girl	French lime	Fuchsia pink
Fluorescent blue	French mauve	Fuchsia purple
Fluorescent orange	French pink	Fuchsia rose
Fluorescent pink	French plum	Fulvous
Fluorescent yellow	French puce	Fuzzy Wuzz
Folly	French raspberry	
Forest green	French rose	

### **КАТЕГОРИЯ «ЦВЕТ PANTONE», ТЕГ \_CLRP**

Mahogany	Sepia	Tan
Fuzzy Wuzzy Brown	Orange	Peach
Chestnut	Burnt Orange	Macaroni and Cheese
Red Orange	Copper	Apricot
Sunset Orange	Mango Tango	Neon Carrot
Bittersweet	Atomic Tangerine	Almond
Melon	Beaver	Yellow Orange
Outrageous Orange	Antique Brass	Gold
Vivid Tangerine	Desert Sand	Shadow
Burnt Sienna	Raw Sienna	Banana Mania
Brown	Tumbleweed	Sunglow

Goldenrod	Sky Blue	Hot Magenta
Dandelion	Outer Space	Purple Pizzazz
Yellow	Blue Green	Razzle Dazzle Rose
Green Yellow	Pacific Blue	Orchid
Spring Green	Cerulean	Red Violet
Olive Green	Cornflower	Eggplant
Laser Lemon	Midnight Blue	Cerise
Unmellow Yellow	Navy Blue	Wild Strawberry
Canary	Denim	Magenta
Yellow Green	Blue	Lavender
Inch Worm	Periwinkle	Cotton Candy
Asparagus	Cadet Blue	Violet Red
Granny Smith Apple	Indigo	Carnation Pink
Electric Lime	Wild Blue Yonder	Razzmatazz
Screamin Green	Manatee	Piggy Pink
Fern	Blue Bell	Jazzberry Jam
Forest Green	Blue Violet	Blush
Sea Green	Purple Heart	Tickle Me Pink
Green	Royal Purple	Pink Sherbet
Mountain Meadow	Purple Mountains'	Maroon
Shamrock	Majesty	Red
Jungle Green	Violet	Radical Red
Caribbean Green	Wisteria	Mauvelous
Tropical Rain Forest	Vivid Violet	Wild Watermelon
Pine Green	Fuchsia	Scarlet
Robin Egg Blue	Shocking Pink	Salmon
Aquamarine	Pink Flamingo	Brick Red
Turquoise Blue	Plum	White

Timberwolf

Silver

Gray

**КАТЕГОРИЯ «ПЛАТЬЯ», ТЕГ \_DR**

ballgown

nightdress

burka

nightrobe

busuuti

nightgown

chemise

nightie

cheongsam

nightshirt

chiton

overdress

coat dress

peignoir

cocktail dress

riding habit

dirndl

sack

gymslip

sari

kaftan

saree

caftan

sheath

kimono

shift

mantua

shirtdress

maxidress

shirtwaister

microdress

shirtwaist

midi

sundress

minidress

sweater dress

Mother Hubbard

tea gown

muu-muu

tunic

negligee

wedding dress

negligée

**КАТЕГОРИЯ «НИЖНЕЕ БЕЛЪЕ», ТЕГ \_UW**

Balmoral

bloomers

body stocking

basque

body

boxer shorts



boxers	foundation garment	shorts
brassiere	French knickers	undershorts
bra	girdle	singlet
briefs	G-string	step-ins
broekies	half-slip	suspender
underpants	waist-slip	garter
bustle	jockstrap	suspender belt
camiknickers	athletic support	garter belt
camisole	knickers	teddy
chemisette	lingerie	thermals
combinations	long johns	trunks
union suit	pannier	underpants
corselet	panties	undershirt
corset	pants	underskirt
crinoline	panty girdle	Y-fronts
drawers	petticoat	
underdrawers	shift	

### **КАТЕГОРИЯ «ОБУВЬ», ТЕГ \_SH**

ankle boot	buskin	crowboot
arctic	chopine	deck shoe
Balmoral	chopin	Doc Marten
biker boot	chukka boot	espadrille
blucher	clog	field boot
bootee	cothurnus	flat
bottine	cothurn	flatty
brogan	court shoe	flip-flop
brogue	cowboy boot	football boot
brothel creeper	creeper	gaiter

spat	mule	slip-on
galosh	overshoe	slipper
ghillie	Oxford	sneaker
golf shoe	pantofle	snowshoe
gumboot	pantoffle	spike
gumshoe	pantoufle	stiletto
gym shoe	platform	surgical boot
half boot	plimsoll	tennis shoe
Hessian boot	plimsole	top boot
high heel	pump	track shoe
hobnail boot	racket	training shoe
jackboot	racquet	trainer
Jandal	running shoe	veldskoen
kitten heel	rock boot	wader
larrigan	sabot	wedge
loafer	sandal	wedge heel
moccasin	sandshoe	welly
moonboot	scuff	winkle-picker
mukluk	slingback	

### **КАТЕГОРИЯ «СВИТЕРА», ТЕГ \_SW**

Aran sweater	crew-necked sweater
cardigan	Fairisle
Cowichan sweater	Guernsey
Indian sweater	hoodie
siwash	Icelandic
siwash sweater	jersey
cowl-necked sweater	jumper
crew-neck	polo

polo neck  
pullover  
rollneck  
skivvy

slipover  
sloppy joe  
sweatshirt  
turtleneck

### **КАТЕГОРИЯ «РУБАШКИ», ТЕГ \_SHRT**

banyan  
blouse  
boiled shirt  
bush shirt  
camise  
chemise  
cover-shoulder  
dashiki  
dress shirt

garibaldi  
grandad shirt  
guimpe  
Jacky Howe  
hair shirt  
kerbaya  
kurta  
khurta  
lava-lava

middy blouse  
overblouse  
polo shirt  
sark  
skivvy  
sports shirt  
Swanndri  
T-shirt  
tee shirt

### **КАТЕГОРИЯ «ЮБКИ», ТЕГ \_SKRT**

A-line  
button-through  
crinoline  
dirndl  
divided skirt  
drop-waisted  
filibeg  
fillibeg  
philibeg  
full skirt  
fustanella  
fustanelle

gaberdine  
gabardine  
grass skirt  
half-slip  
waist-slip  
hobble skirt  
hoop skirt  
kilt  
lava-lava  
maxiskirt  
microskirt  
midiskirt

miniskirt  
overskirt  
pencil skirt  
petticoat  
puffball skirt  
ra-ra skirt  
riding skirt  
sarong  
tutu  
underskirt  
wrapover  
wrapround

## КАТЕГОРИЯ «ГОЛОВНЫЕ УБОРЫ», ТЕГ \_НАТ

Akubra	boater	crash helmet
anadem	bonnet	crown
babushka	bowler	curch
Balaclava helmet	derby	curchef
Balaclava	broadbrim	deerstalker
Balmoral	busby	diadem
bluebonnet	calash	Dolly Varden
bandanna	caleche	dunce cap
bandana	calotte	earmuff
bandeau	calpac	fascinator
barret	calpack	fedora
baseball cap	kalpak	fez
basinet	capuche	flat cap
bascinet	capouch	fool's cap
beanie	castor	forage cap
beany	chaplet	frontlet
bearskin	circlet	frontal
beaver	cloche	Gandhi cap
beret	cloth cap	glengarry
billycock	cocked hat	hard hat
bowler	coif	havelock
biretta	commode	headband
berretta	coonskin	headdress
blackcap	cornet	heaume
bluebonnet	coronet	helmet
bluecap	cossack hat	homburg

hood	topee
Juliet cap	topi
keffiyeh	poke
kaffiyeh	poke bonnet
kufiyah	porkpie hat
kepi	sailor hat
laurels	sallet
leghorn	salet
liberty cap	salade
lum-hat	shako
mitre	shacko
mobcap	shovel hat
mob	shower cap
montero	silk hat
morion	
mutch	
nightcap	
opera hat	
gibus	
Panama hat	
paper hat	
peaked cap	
petasus	
Phrygian cap	
picture hat	
pillbox	
pinner	
pith helmet	

