

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ИНСТИТУТ СОЦИАЛЬНО-ГУМАНИТАРНЫХ НАУК
Кафедра английского языка

РЕКОМЕНДОВАНО К ЗАЩИТЕ
В ГЭК И ПРОВЕРЕНО НА ОБЪЕМ
ЗАИМСТВОВАНИЯ

И.о. заведующего кафедрой

канд. филол. наук

 А.С. Остапенко

11 июля 2019 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(магистерская диссертация)

**КЛАССИФИКАЦИЯ ИНТЕРЕСОВ ПОЛЬЗОВАТЕЛЕЙ В СОЦИАЛЬНЫХ
СЕТЯХ НА ОСНОВЕ ЯЗЫКОВЫХ ПРИЗНАКОВ**

45.04.02 Лингвистика

Магистерская программа


«Теория и практика преподавания иностранных языков и культур»

Выполнила
Студентка 2 курса
очной формы обучения



Асатрян
Валентина
Артуровна

Научный руководитель
канд. филол. наук



Михалькова
Елена
Владимировна

Рецензент
доцент кафедры английской
филологии и перевода
Тюменского государственного
университета,
канд. филол. наук



Ковязина
Марина
Анатольевна

г. Тюмень, 2019

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
ГЛАВА 1. ВИРТУАЛЬНЫЕ СОЦИАЛЬНЫЕ СООБЩЕСТВА КАК ПРЕДМЕТ АНАЛИЗА.....	7
1.1. Классификация реальных и виртуальных сообществ	7
1.2. Лингвистические признаки сообществ в социальной сети.....	20
1.3. Тезаурус на основе интересов пользователей.....	28
ВЫВОДЫ ПО ГЛАВЕ 1	37
ГЛАВА 2. ТЕЗАУРУСНЫЙ МЕТОД ОРГАНИЗАЦИИ ИНТЕРЕСОВ	39
2.1. Анализ интересов	39
2.2. Составление тезауруса интересов пользователей.....	53
ВЫВОДЫ ПО ГЛАВЕ 2	61
ЗАКЛЮЧЕНИЕ	63
СПИСОК ИСТОЧНИКОВ И ЛИТЕРАТУРЫ.....	66
ПРИЛОЖЕНИЕ 1	72
ПРИЛОЖЕНИЕ 2	89

ВВЕДЕНИЕ

Исследование посвящено интересам пользователей в социальных сетях с лингвистической перспективы. На основе случайной выборки виртуальных русскоязычных сообществ ВКонтакте, включающих лексические единицы русского, английского и испанского языков, проанализированных по лингвистическим признакам, проведен эксперимент, идея которого – найти главный интерес, объединяющий все выделенные в ходе анализа интересы, и создать иерархию интересов, подкрепленную тезаурусом из лексических дескрипторов.

Цель работы: обобщить интересы сообществ социальных сетей в форме тезауруса и проследить их иерархию.

Задачи: 1) обобщить ключевые понятия исследования: тезаурус, социальная сеть, лингвистические дескрипторы, интерес; 2) изучить феномен «социальная сеть», лингвистические особенности коммуникации в социальных сетях (виртуальных и реальных) и особенности объединения в виртуальные сообщества через коммуникацию в Интернет; 4) создать выборку виртуальных сообществ, выделить лингвистические дескрипторы интересов и определить главный интерес сообществ; 5) выстроить и организовать в форме тезауруса иерархию полученных в ходе эксперимента интересов.

Объект исследования – интерес сообщества в социальной сети.

Предмет исследования – моноязыковые и мультязыковые сообщества в социальных сетях. Выбор сообществ случайный; текст сообществ собран посредством специализированного компьютерного приложения.

Актуальность работы обоснована тем, что анализ интересов, включающий метаданные и подкрепленный тезаурусом, будет пользоваться спросом не только у лингвистов, но и в сфере программирования,

социологии, SMM и др., а также облегчит пользователям процесс поиска нужной информации по интересам.

Существуют онлайн сервисы, скачиваемые программы и библиотеки для языков программирования, которые решают различные задачи автоматической обработки текста: GATE, MALLET, Apache OpenNLP, FreeLing, NLTK и др. Произвести обработку и анализ лингвистических данных со страниц пользователей можно также с помощью статистического программного обеспечения: SAS, STATISTICA, Oracle Data Mining, SPSS, STADIA, STATGRAPHICS, Mathcad, R. Сбор данных приходится осуществлять вручную, но после ввода информации в компьютер появляется возможность структурировать ее, создать удобные схемы, таблицы, диаграммы, произвести разметку.

Однако тот факт, что предлагаемые системы анализа опираются на данные, указанные пользователями, является их слабой стороной. Исходя из описания ресурсов, анализ текстовых данных из комментариев пользователей, постов на стене и т.д. вообще не используется, скорее всего, в силу отсутствия модели для анализа этих данных. Анализ социальных сетей применяется как вспомогательный инструмент при поиске сообществ, близких по интересам, и до сих пор остается слабо автоматизированной областью знания. Данное исследование может помочь в построении «системы интересов», и полученные данные могут быть использованы как часть базы данных или материал для иных целей.

Материалом исследования послужили тексты, собранные в общедоступных активных тематических моноязыковых (русскоязычных) и мультязыковых (с лексическими единицами английского и испанского языков) сообществах социальной сети «ВКонтакте» за период 2018-2019 гг. Текст каждого анализируемого сообщества составлял до 10000 слов и включал 100 последних постов, описание сообщества, комментарии, правила, обсуждения. Всего вручную было проанализировано 95 сообществ.

Методологической основой послужили работы лингвистов, программистов, социологов, переводчиков, компьютерных лингвистов. Исследование опирается на работы Е. В. Михальковой, С. С. Фролова, Н. В. Лукашевич, М. С. Мельниковой, Вл. А. Лукова, Э. К. Погорского, А. В. Коршунова, В. А. Лукова, Ю. Н. Караулова, П. Сринивасан и др.

Основными **методами** послужили метод системно-функционального анализа языковых единиц, описательно-аналитический метод, метод ключевых слов и дескрипторов, тезаурусный метод, установление парадигматических отношений между дескрипторами, анализ лингвистических пересечений.

Структура работы. Исследование состоит из введения, двух глав, заключения, 6 рисунков, 1 таблицы, библиографического списка и 2 приложений.

В Главе 1 рассматриваются признаки социальных сообществ, отличия и сходства виртуальных и реальных сообществ и их классификации. Также описываются лингвистические особенности виртуального языка, его компоненты, нехарактерные для реальной речи, и невербальные компоненты. Сравнивается письменная речь реального и виртуального общения. Третий пункт описывает типы тезаурусов, их характеристики и структурные особенности, функционал, процесс построения, системы внутренних отношений терминов/дескрипторов, возможные модификации.

В Главе 2 описывается ход эксперимента. Анализ интересов подразумевает выделение главного интереса анализируемых сообществ на основе дескрипторов и целей сообщества. Описывается иерархия интересов, построенная на основе анализа лингвистических пересечений. Второй пункт отведен для описания полученного тезауруса с наглядными примерами, характеристики его структуры и системы условных обозначений.

Выделяются принципы выбора дескрипторов для наполнения тезауруса, описываются факторы принципа «значимость».

В заключении анализируются результаты эксперимента по разработке тезауруса.

ГЛАВА 1. ВИРТУАЛЬНЫЕ СОЦИАЛЬНЫЕ СООБЩЕСТВА КАК ПРЕДМЕТ АНАЛИЗА

1.1. Классификации реальных и виртуальных сообществ

Социальная коммуникация в современном обществе является связующим звеном всех коммуникативных сфер человека. Появление глобальной компьютерной сети Интернет стало одним из наиболее важных факторов развития общества. Виртуализация процесса коммуникации привела к изменению в сфере межличностного общения, позволив осуществлять коммуникацию на дальних расстояниях, для чего в сети Интернет созданы особые площадки – социальные сети, которые на данный период времени пользуются огромной популярностью среди аудитории различных возрастов.

У пользователя социальной сети формируется «сетевая картина мира» [Матусевич 2016]. Знакомые и незнакомые с интернет-общением люди – это два отдельных мировосприятия и типа коммуникантов, вне зависимости от того, принадлежат ли они одним реальным группам по возрасту и социальному статусу. В реальной коммуникативной ситуации они могут столкнуться с трудностями в понимании друг друга (в качестве типичной ситуацией можно привести пример непонимания старшим поколением «словечек», которые используются молодыми поколениями).

Социальные сети представляют собой огромную базу данных, материалы которой можно использовать для различных научных исследований, в частности касающихся изучения человеческой природы. С каждым днем к виртуальной жизни присоединяется все большее количество людей, что расширяет возможности интернет пространства.

Для упрощения работы с имеющимся материалом специалисты из разных областей создают программы, целью которых является упорядоченность базы данных. Один из вариантов упорядоченности –

тематическая кластеризация. В данной работе нас интересует тематическая кластеризация на основе интересов пользователей.

Социальная сеть представляется как виртуальная платформа, предоставляющая пользователю площадку для обмена информацией. «ВКонтакте» – проект Павла Дурова, созданный по подобию популярной за рубежом виртуальной социальной платформы Facebook для легкого общения с широким кругом интересных людей и поддержания связей с уже знакомыми пользователями. У пользователей есть возможность на постоянной основе делиться интересами и быть частью сообществ по интересам. К главным признакам социальной сети относятся: 1) создание личного профиля (странички); 2) взаимодействие между пользователями (личные сообщения, просмотр профилей, комментарии, участие в обсуждениях и др.); 3) накопление и обмен информацией (участие в сообществах/создание сообществ; «лайки» и репосты).

Интерес пользователя, несмотря на свой индивидуальный характер, в виртуальном пространстве социальной сети имеет социальный характер. Он не может существовать сам по себе и требует коллективной организации. Отсюда появляются различного рода тематические сообщества – паблики – объединяющие людей в социальные кластеры и привлекающие новых участников. Из этого следует, что интерес пользователя будет рассматриваться как интерес целого сообщества, как признак группы пользователей в сети. Получается, что если мы классифицируем интересы, то мы классифицируем сообщества (кластеры), которые проявляют этот признак, и, следовательно, пользователей (индивидов), которые составляют эти сообщества. То есть, классификация интересов – это классификация сообществ (кластеров).

Сообщество – это группа людей, связанных определенными признаками, целями, идеями. Оно может принадлежать более крупному объединению — социальной группе. Е. В. Михалькова выделяет для социальной группы ряд

наследуемых (передаются кластерам и индивидам) и ненаследуемых признаков:

1. место пребывания группы: ненаследуемый (устанавливается на любой территории – волонтерами «Greenpeace» являются люди с разных стран и ведут активность по всему миру);

2. время пребывания группы: ненаследуемый (участники не зависят от временных рамок, могут вступить и выйти из сообщества в любое время);

3. численность участников: ненаследуемый (деятельность сообщества динамична – одни участники выходят, другие – вступают);

4. возраст участников: ненаследуемый (участник может быть любого возраста, даже если это сообщество рассчитано на определенный возраст – ввиду анонимности этот фактор сложно отследить);

5. пол участников: ненаследуемый (исключение могут составлять сообщества, посвященные радикальному феминизму или, наоборот, мизогинии; опять же срабатывает фактор анонимности);

6. лидеры группы: наследуемый – признанные лидеры всей группы (возможно, авторы литературных, философских произведений, на которые ориентируется группа, политические лидеры и т.п.); ненаследуемый – лидеры в кластере (закономерно появляются в любой группе);

7. деятельность: наследуемый – виды активности, которые поддерживаются всеми участниками группы (например, cosplay); ненаследуемый – деятельность отдельных кластеров, индивидов может отличаться от эталона, но квалифицироваться как деятельность именно участников группы;

8. ресурсы: наследуемый – финансовая поддержка группы, идеологическое влияние какой-либо партии и политических активистов,

сотрудничество группы со СМИ может разделяться всеми участниками группы, тем более, если все они получают доступ к этим ресурсам; ненаследуемый – в кластерах может возникать сотрудничество и контакты, которые недоступны и неизвестны другим участникам группы;

9. наименование участников: наследуемый (гринписовцы);

10. атрибутика: наследуемый (радужный флаг у ЛГБТ+ сообществ);

11. стиль речи: наследуемый (сообщества любителей кошачьих используют такие лексические единицы, как «кусь», «буп», «у меня лапки»; часто используются уменьшительно-ласкательные суффиксы: «котенька», «котечка», «лапочки»; детская манера речи: «нинаю», «патьяки», «апути»);

12. идеи: наследуемый (для сообщества, посвященного транс-людям, главная идея – поддержка транс-людей и борьба с трансфобией);

13. ценности: наследуемый (для гринписовцев это права и свобода животных);

14. поведенческие нормы: наследуемый (уважение идей, ценностей и участников сообщества);

15. санкции: наследуемый (бан за оскорбление идей, ценностей и участников сообщества);

16. отношения с другими группами: наследуемый (феминистки борются с сексистами и последователями патриархата).

Наличие данных признаков варьируется в зависимости от типа сообщества (как виртуального, так и реального) – не обязательно наличие всех вышеперечисленных признаков. Тем не менее, ни один из этих признаков по отдельности не определяет тип сообщества. При описании различных групп данные признаки легко формализуются и служат средством дифференциации между группами. Таким образом, социальное сообщество –

это набор тех или иных признаков, по которым сообщества отличаются друг от друга по своему типу. Далее мы рассмотрим некоторые классификации виртуальных и реальных сообществ и разберем отличия, если таковые имеются, виртуальных сообществ от реальных.

А. В. Коршунов называет систему сообществ в социальных сетях «виртуальным отражением части человеческого социума» [Коршунов 2015], подчеркивая тем самым тот факт, что виртуальные и реальные сообщества располагают одинаковыми признаками. В целом, виртуальная социальная сеть строится по такому же принципу, как и социальная сеть, и обе эти сети имеют много схожих характеристик. Главными отличиями являются роль социального статуса пользователя: «социальный статус пользователей Интернета, участвующих в каком-либо обсуждении, перестает быть фактором, определяющим их поведение» [Мельникова, Яковлев 2014], а также отсутствие территориальной ограниченности и реального места. Говоря простыми словами, в виртуальном мире все равны и собраны со всех уголков земли. Далее мы рассмотрим несколько классификаций признаков реальных и виртуальных сетей. Современный российский социолог С. С. Фролов отмечает три самых распространенных вида реальных социальных сообществ – социальные группы, квазигруппы и социальные круги.

С.С. Фролов определяет социальную группу как «совокупность индивидов, взаимодействующих определенным образом на основе разделяемых ожиданий каждого члена группы в отношении других» [Фролов 2001], выделяя два условия для существования социальной группы:

- взаимодействие индивидов;
- оправдание ожиданий индивидов по отношению к поведению других индивидов в группе.

К специфическим признакам социальной группы он относит [Фролов 2001]:

- четко определенное членство в группе, ощущение принадлежности и границ группы, ее места в какой-либо организации;
- групповое сознание, касающееся признания правильности и уместности социальных норм, характерных для данной группы;
- четкая постановка целей;
- взаимозависимость членов группы, возникающая в результате совместной деятельности;
- способность индивидов действовать в составе группы, решать общие задачи;
- конформность к существующим нормам, признание лидеров;
- наличие четко определенных групповых ролей.

Стоит отметить, что в бытовой речи понятие «социальная группа» в виртуальной среде отличается от того, что им описывается в социологии. Под этим термином (обычно в сокращенной форме «соц. группа») понимается любое сообщество, созданное на платформе той или иной социальной сети.

Невыполнение какого-либо условия из тех, что перечисляет С. С. Фролов свидетельствует о том, что сообщество представляет из себя уже разновидность социальной группы: квазигруппу или социальный круг. Квазигруппу нельзя считать полноценной социальной группой ввиду следующего набора признаков:

- спонтанность образования (пришедшие на мероприятие участники не имеют никаких связей – собираются незнакомые друг другу люди);

- неустойчивость взаимосвязей (перед участниками нет никаких обязательств, не установлены правила, возможен выход из группы без последствий);

- отсутствие разнообразия во взаимодействиях (пришедшие на концерт танцуют и подпевают, митингующие несут транспаранты и требуют);

- кратковременность совместных действий (до тех пор, пока все ее члены занимаются одним делом: во время проведения митинга все митингующие – члены квазигруппы; в последствии она может перейти в социальную группу при появлении лидеров, установке правил, разделения общих идей и др.).

Социальный круг – это переходный этап от общественных взаимоотношений между людьми, объединенными общими интересами, к полноценной социальной группе. Он возникает для обмена информацией между его членами и предполагает несколько большую связь между людьми, чем квазигруппа. Возникать он может осознанно или неосознанно, но главное его отличие от квазигруппы в том, что представители квазигруппы встречаются лишь единожды во время совершения общего действия, а представители социального круга встречаются на более-менее постоянной основе (например, коллектив на работе, аудитория тренинга). Функция социальных кругов заключается в обмене информацией. В них нет реальной активности, нет общности разделяемых идей и нет тесных устойчивых отношений.

Данное разделение может быть свойственно и для виртуальных сообществ, но с некоторыми отличительными признаками:

- преобладание письменной речи (в большинстве случаев свободной от установленных норм языка);

- территориальная и временная неограниченность (участники группы могут быть собраны со всего мира и общаться вне зависимости от временных рамок – общение с пользователем в режиме оффлайн);

- анонимность (даже если пользователь реализует коммуникацию под настоящим именем, располагает фотографиями своей внешности и сведениями анкетного характера, много информации о нем все равно не известно и, таким образом, отсутствует возможность адекватного восприятия его личности);

- «раздвоение личности» (в виртуальном пространстве пользователь может выдумать себе характер, так как чувствует свободу от каких-либо социальных или моральных рамок, которые он не может нарушать в реальном мире);

- для всех видов групп не обязательно взаимодействие вне виртуальной среды.

Также для виртуального информационного общества есть и другой перечень характеристик. Социальный философ Э. К. Погорский выделяет следующие характеристики [Погорский 2012]:

- информация и коммуникационные процессы определяют в обществе доминирующую картину мира и основные социальные практики, которые выступают как источник и механизмы социокультурных изменений и диалога культур;

- сетевые социальные взаимодействия становятся важнее иерархических взаимодействий, поскольку обеспечивают автономность личности и социальных общностей - под влиянием все более увеличивающейся значимости социально-сетевых способов организации общественной жизни сети становятся более эффективными конструктами общества, чем иерархия;

- виртуальный мир становится неотъемлемой частью жизни человека;

- пространство и время перестают быть существенными для коммуникации;

- инновационность становится важнее поддержания культурных образцов - позиции «консерваторов» начинают ослабевать, расчищая дорогу новым идеям и инновациям, берущим начало из молодежной среды;

- виртуальные связи не менее значимы для личности или группы, чем живое общение;

- автономные системы взаимодействия и коммуникации становятся неотъемлемой чертой социальных групп, их атрибутом и источником формирования.

К сожалению, на данный момент не существует подробной классификации сообществ в виртуальных социальных сетях, но чаще всего (по крайней мере в иностранной литературе) встречается разделение на communities of practice, CoP – термин, введенный Дж. Лэйв и Э. Венгером, и communities of interest, CoI.

Реальное CoP обозначает группу людей, вовлеченных в совместную деятельность. В таком сообществе доминирует коллективная идентификация, и целью является совместный труд и создание артефактов [Henri, Pudelko 2003]. Такой подход приводит к упорядочиванию знаний. Его признаками являются:

- активное участие участников группы (не рассматривается пассивное участие);

- обмен информацией/советами;

- обеспечение поддержкой соучастников;

- наличие компетентного лидера хотя бы с минимальной базой знаний (или группы лидеров для координации и модерации).

Виртуальное СоР позволяет пассивное восприятие информации и отсутствие реального места сбора. С одной стороны, это дает преимущества (больше количество участников, следовательно, больший объем информации и большая продуктивность); с другой стороны, усложняется задача упорядоченности знаний, лидерам сложнее контролировать ход процесса.

Ф. Анри и Б. Пуделько (2003) считают, что реальное СоІ представляет собой сообщество, объединяющее людей определенного круга интереса/интересов. Временных рамок у такого сообщества тоже нет – оно может быть как долгосрочным, так и краткосрочным. В отличие от СоР, участникам не обязательно учиться чему-либо, а также не обязательно проявлять активность – участники не ждут друг от друга систематического обмена информацией и мнениями. Также, участникам свойственна индивидуальность; отсутствует задача создания артефактов. Такое сообщество располагает следующими признаками [Henri, Pudelko 2003]:

- обмен информацией и обсуждение информации, связанной с интересом;
- обязательно разделение интереса с соучастниками;
- не обязательно присутствие квалифицированного лидера.

Цель реального СоІ – обеспечить участникам пространство для обмена информацией об интересе. Для виртуального СоІ характерны все те же признаки, за исключением наличия реального места.

Помимо социальных сообществ на просторах виртуальной социальной сети существуют и другие сообщества. Например, коммерческие сообщества, где присутствует интерес, но нет никакого группового взаимодействия. Сообщество рассчитано на рекламу своих товаров или услуг или может являться полноценной торговой площадкой (можно связаться с контактным

лицом, посмотреть фотографии товара, оформить заказ, уточнить детали и др.). Также есть доска объявлений. Такие сообщества могут быть как контролируемые (посты на стену может выставлять только администратор, участники могут предлагать новости, и они будут находиться в разделе «предложенное»), так и неконтролируемые (кто угодно может выкладывать посты на стену, информация никем не проверяется). Такие неконтролируемые открытые стены чаще всего собирают огромное количество спама. Но они, как правило, тоже имеют определенный интерес (поиск работы, б/у вещи, фриланс вакансии). То есть любой паблик вне зависимости от своей категории имеет интерес, за которым следуют пользователи.

Универсальный алгоритм кластеризации виртуальных сообществ по интересам на сегодня отсутствует. «ВКонтакте» предлагает свой список главных интересов сообществ (которые могут делиться на более специфичные интересы) в разделе «популярные сообщества», который состоит из 13 категорий: новости, спорт, музыка, блогеры, радио и

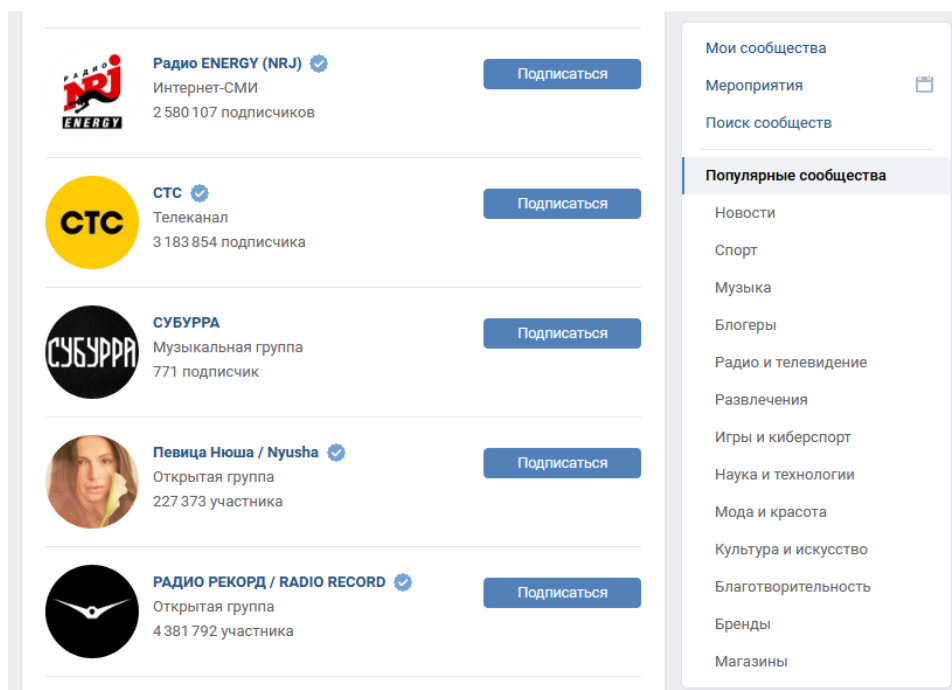


Рисунок 1 – Скриншот – Категории интересов в ВКонтакте.

телевидение, развлечения, игры и киберспорт, наука и технологии, мода и красота, культура и искусство, благотворительность, бренды, магазины. Такое деление не обязательно полностью соответствует действительному интересу, но в любом случае упрощает навигацию по имеющейся огромной базе сообществ, растущей с каждым днем в связи с появлением новых пользователей, новых тенденций и новых интересов.

Наибольшей популярностью пользуются развлекательные паблики, наполненные юмором и иронией по отношению к действительности. Пользователей привлекает простой и доступный каждому контент: чем проще, тем больше вероятность репоста. В целом, тенденция к простоте свойственна и для реальной жизни. У популярных интересов (художественные фильмы, легкая литература, российские сериалы на отечественных каналах) большая аудитория. Более того, большинство среднестатистических пользователей привлекает «картинка» к постам: обычно длинные тексты без визуального или звукового сопровождения многие обходят стороной (на эту тему в виртуальной среде даже есть отдельная шутка: «Если вы все же читаете пост без картинки...»; «Спасибо, что дочитали пост до конца»). Поэтому контент в виртуальной сети зачастую содержит больше графических элементов, нежели текста.

Особенно популярно и привлекательно помещать текст внутри картинки, и такие посты собирают большее количество лайков. Оставлять комментарии также намного проще посредством эмодзи, стикеров или всем известных мемов, нежели описывать свои эмоции. Тут дело, возможно, не столько в избегании слов, сколько в экономии времени. Для этой цели не так давно были введены голосовые сообщения, которые раньше были присущи только инстант-мессенджерам (WhatsApp, Viber, Telegram и др.). Теперь они активно используются на платформе «ВКонтакте». Если говорить о мемах, то их использование, скорее, свидетельствует о том, что пользователь «в курсе»

актуального юмора и стремится к одобрению и поощрению со стороны соучастников, к общности с группой.

Далее подробнее разберем лингвистические особенности виртуальной социальной сети.

1.2. Лингвистические признаки сообществ в социальной сети

Если до этого мы говорили о том, что признаки виртуальных сообществ не глобально отличаются от реальных и обладают большим количеством сходств, то про разницу в языках этих сообществ так сказать уже невозможно. Виртуальное информационно-коммуникационное пространство предстает как место реализации нового, отличного от реально существующего языка и рассматривается как новая сфера функционирования языка. Так как это пространство преимущественно письменного языка, принципиальную значимость имеет визуальная информация. При этом письменная речь, свойственная для реального мира, имеет другой ряд нетипичных признаков. Рассмотрим различия письменной речи двух пространств, где признаки реальной речи описаны М.Р. Львовым [Львов 2000]:

Реальное пространство	Виртуальное пространство
Главная функция – зарождение и передача новой мысли	Главная функция - коммуникативная
Отсутствуют невербальные средства	Присутствуют невербальные средства
Особая значимость орфографии и пунктуации	Отсутствие норм, свободное использование языка
Порождает текст	Порождает и текст, и дискурс

Таблица 1 – Письменная речь виртуального и реального пространств

Язык виртуального общения наполнен многими элементами, несуществующими в реальной коммуникации, то есть, он является поликодовым – вербально-визуальным – состоящим из многих элементов разного характера [Матусевич 2016]: мультимедия (изображения, аудиозаписи, видеозаписи, gif-изображения), графические эмодзи (стикеры, эмодзи и эмотиконы (смайлики)), шрифтовое разнообразие (зачеркнутый текст – больше присущ блогowym платформам; написание «капсом», «лесенка Маяковского»), тэггинг (хэштеги и мэншоны), универсальная

акронимия («прив», «го», «irl», «thx») и др. Рассмотрим подробнее некоторые из них, типичные для социальных сетей.

Первый широко используемый набор эмодзи (с яп. – изображение, знак), также известные как смайлики, был разработан японским интерфейс-дизайнером Сигэтакой Курита в 1998-1999 году и состоял из 172 символов размером 12×12 пикселей [Wikipedia]. Маленькие картинки, обозначающие человеческие эмоции, набирали популярность, развивались по функционалу и приобретали все новые и новые формы и значения. На сегодняшний день существуют различные словари эмодзи, а используя Яндекс Переводчик возможно перевести текст на и с языка эмодзи.

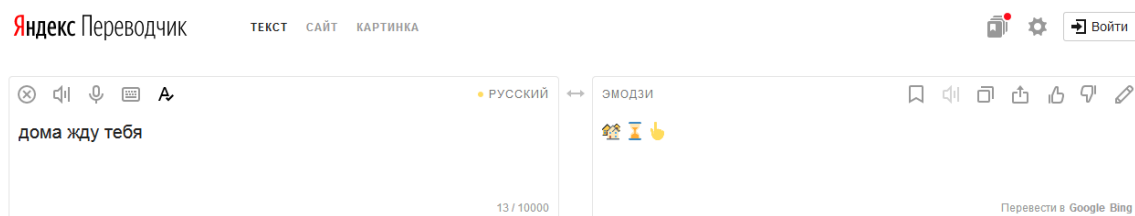


Рисунок 2 – Скриншот – перевод на язык эмодзи в Яндекс Переводчике.

На сегодняшний день пользователи виртуальных площадок для коммуникации обладают огромным спектром смайликов в виде людей, животных, различных предметов и всевозможных иконок. Самыми популярными остаются желтые круглые рожицы (впервые введенные американским коммерческим художником Харви Боллом в 1963 году для поддержания боевого духа сотрудников страховой компании [Wikipedia, Эмодзи]), отображающие в разной степени гиперболизированные человеческие эмоции. Если вспомнить времена ICQ и QIP, то большей популярностью обладали анимированные смайлики. Сегодня же они в большинстве своем статичны. Зачастую смайлики сопровождают чуть ли не каждое сообщение пользователей (есть версия, что такая особенность в большей степени присуща пользовательницам [Wolf 2000]), но в рамках современного мира это звучит несколько по-сексистски и слишком

субъективно), и один и тот же смайлик может быть повторен несколько раз в целях гиперболизации эмоций, при этом в реальной жизни пользователь не обязательно испытывает те эмоции, что он обозначил. Более того, не обязательно использовать эмодзи, можно воспользоваться обычной скобочкой «)» или «(» для упрощенного обозначения радости и грусти. Такая графическая реализация эмоций называется эмотиконами. Причем, для иностранных пользователей обязательно наличие «глаз» – «:»), иногда и «носа» – «:-)», что было больше типично для ранней эпохи эмотиконов. Обычная скобка свойственна для рунета, и таким образом можно догадаться о происхождении собеседника.

Смайлики обладают двумя характерными чертами: 1) узнаваемость и понятность – за основу берется широко известное явление или предмет (солнце, люди, флаги, общепринятые иконки), а также они отражают реальные человеческие эмоции; 2) преобладающая единичность – как правило смайлики представлены одним предметом, но для подчеркивания отношений между предметами (эмодзи с семьей, хлопающие ладони) могут использоваться несколько разных или одинаковых элементов.

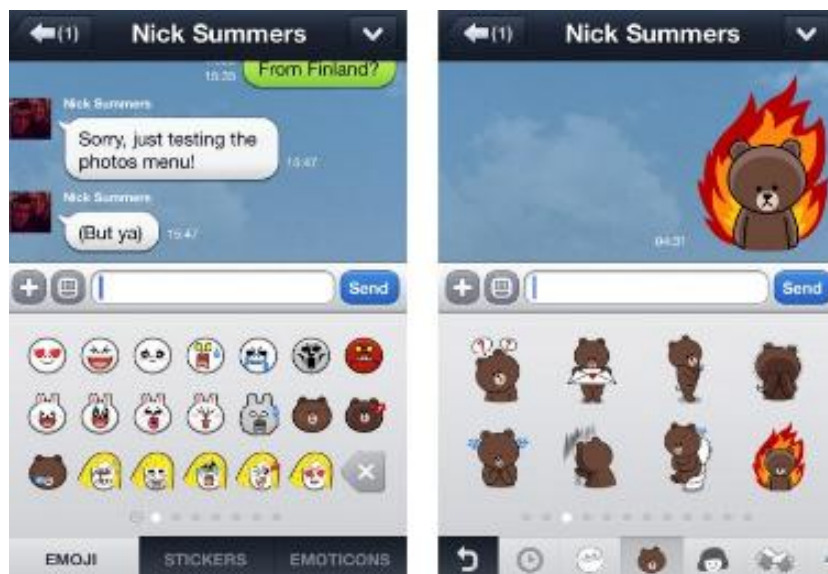


Рисунок 3 – Скриншот – первые стикеры в приложении Line.

Стикеры появились после эмодзи. По своей природе они более детализированы, являются большими по размеру, но несут ту же функцию, что и смайлики. Впервые были разработаны на платформе корейского приложения мгновенных сообщений Line и в 2013 появились для пользования на платформе социальной сети Facebook [Russell 2013].

Стикеры также доступны для любых платформ виртуальной коммуникации и обладают огромным спектром выражения не только эмоций, но и действий и состояний, часто сопровождаемые словами («болею», «жду», «уже бегу»). Их отличие в том, что они отсылаются единичной картинкой, в то время как смайликов можно отправить в одном сообщении огромным количеством, как одинаковых, так и разных. Также стикеры могут брать за основу мемы – медиаобъекты, создаваемые электронными средствами коммуникации, фразы, концепции или занятия, как правило, остроумные и иронические, спонтанно приобретающие популярность, распространяясь в Интернете разнообразными способами [Wikipedia, Мем].

Также большой популярностью и ценностью для анализа и изучения виртуальной коммуникации обладает хэштег (hashtag) – тег с метаданными в виде слова или фразы без пробелов либо с нижним подчеркиванием для обозначения и облегченного поиска темы, перед которым стоит знак решетки «#». Впервые был предложен Крисом Мессина в 2007 на платформе микроблоггинга Twitter и описан как "thing for nerds" (вещь для задротов) [Wikipedia, Hashtag]. Однако хэштег мгновенно разлетелся по всему Твиттеру, а после и по другим виртуальным платформам. Интересно, что хэштег пробирается и в реальную коммуникацию, когда индивид произносит некое высказывание и в конце добавляет слово «хэштег» и обозначает тематику высказывания (*У меня сегодня болит голова, текут сопли, и вообще я не соображаю ничего – хэштег болею*). Феномен хэштега связан с еще одним глобальным явлением – фолксономией (folk + taxonomy) или коллективному тэггингу. Термин фолксономия был введен Томасом Вандер

Волом. Он дает ей следующее определение – «the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval» [Vander Wal 2007] («Результат персонального свободного тэггинга информации и объектов (чего-угодно с URL) в целях возврата к информации» - перевод наш, В.А.). С ее помощью категоризируется и резюмируется онлайн контент. Во «ВКонтакте» многие сообщества используют хэштеги для обозначения тематики поста, а также для удобной навигации по тематике постов сообщества: нажимая на хэштег, пользователь получает список постов, отмеченных данным тегом. Для того чтобы тэги не перемешивались с информацией с других сообществ, их специализируют, придумывая свое дополнительное обозначение. Это может быть, например, вариация названия паблика (#МирРун_магия – сообщество «Мир Рун», #mb_demon – сообщество «monster boyfriend»). Также вместо хэштегов могут фигурировать мэншоны (mentions), отсылающие уже не к тематике, а непосредственно к сообществу или пользователю, где вместо знака решетки используется символ at «@» (@somebody). При этом при фолксономии может быть и слияние хэштега и мэншона (#art@blacksad_public). Онлайн сервис Hashtagify.me по ключевому слову выдает до 30 связанных с ним актуальных популярных хэштегов на платформах Twitter и Instagram. «Today our free Hashtags Finder allows you to search among 12,408,848 Twitter and Instagram hashtags and pick the best ones for your needs based on their popularity, relationships, languages, Twitter influencers and many other metrics» [Hashtagify.me] («На сегодняшний день бесплатный поиск по хэштегам позволяет вам найти среди 12,408,848 хэштегов Твиттера и Инстаграма и выбрать самые подходящие для вас на основе популярности, отношений, языка, влиятельных пользователей Твиттера и многих других метрик» - перевод наш, В.А.).

Благодаря акронимии, на просторах виртуального пространства зародилось большое количество универсальных сокращений, и некоторые из

них сложны для распознавания, особенно взятые из английского языка (напр. «ирл» - «irl» - in real life). Стремление к сокращению объясняется экономией времени, ввиду чего акронимы из языка виртуальной коммуникации пробираются в реальную («лол» - «lol» - laughing out loud, «спс» - спасибо, «норм» - нормально, «хз» - не знаю (взято от грубого выражения)).

Ранее отмечалось, что лингвистические признаки (наименование участников и стиль речи) относятся к наследуемым признакам сообщества. Язык социальных сообществ зачастую меняет поведение участников и влияет на способ обмена информацией. Внутри общего для всех пользователей сети языка рождаются специфические языковые ответвления. Индивид, принадлежащий к определенному виртуальному сообществу, часто использует в своей речи то, что уже слышал от других ее участников. Так индивид подчеркивает свою принадлежность к другим участникам и их общему интересу, солидарность и заинтересованность. Этот процесс ретрансляции происходит постоянно и приводит к образованию «коллективного» в языке группы, что является признаком принадлежности говорящего к группе, т.е. что определяет интересы сообщества. Е.В. Михалькова выделяет следующие лингвистические признаки сообщества [Михалькова 2017]:

- ключевые слова (могут включать хэштеги);
- дериваты (напр. «замурчательный» от «мур»);
- уникальные лексемы (характерные только для участников данного сообщества/социальной группы, могут совпадать с существующими словами, но иметь другой смысл – «укотовление», «раздавать ключи»);
- именованные сущности (наименование участников, имена лидеров, значимых личностей, места сбора; могут включать мэншоны).

Данные лингвистические признаки в полном своем составе преимущественны для социальных групп, но при этом могут и присутствовать в речи меньших групп – квазигрупп и социальных кругов. Если говорить о классификации на CoI и CoP, эти признаки также будут для них характерны. Даже доска объявлений может иметь свой специфический лексикон. То есть, наличие этих признаков зависит не только от типа сообщества, но и от проявляемого этим сообществом интереса. Лексические единицы отражают интерес, следовательно, интерес в виртуальном пространстве не может без них существовать.

Большинство виртуальных сообществ имеет следующую структуру: короткое высказывание на тему того или иного предмета, обычно сопровождающееся нелингвистическим контентом (изображения, музыкальные дорожки, gif-изображения, видео) и комментариями. Так создается лингвистическая база, привлекающая не только новых пользователей, но и исследователей, которые в свою очередь используют ее в контент-моделях для классификации интересов пользователей.

Е.В. Михалькова вводит понятие «главный интерес пользователя» - MaI (main interest), которое предполагает некую сферу, обозначаемую существительным или любой другой субстантивированной частью речи (кино, комиксы, музыка и др.) [Михалькова 2017]. В данной работе, главный интерес мы рассматриваем с двух перспектив: 1) как доминирующий интерес сообщества в парадигме интересов (напр. «виниловые вечеринки – ретро музыка – музыка», где музыка – главный интерес); 2) как зонтичный термин (umbrella interest), объединяющий интересы (напр. литература, музыка, кино, ИЗО = искусство). При этом для определения зонта обязательно выявление доминирующих интересов сообществ, то есть перспектива 2 не может существовать без перспективы 1, но в обратную сторону это условие не действует.

Задачу кластеризации сообществ, близких по интересам, можно свести к задаче классификации по набору лексических единиц, обозначающих эти интересы. Многие интересы пересекаются между собой наличием одинаковых лексических единиц (напр. сообщества NHL и NBA: матч, игра, игрок, конференция, плей-офф и др.), и таким образом попадают под один главный интерес (зонтичный – спорт). Но при этом есть сообщества с несколькими интересами, и здесь задача усложняется, но решается при определении доминирующего интереса на основе лингвистических признаков (подробнее описано во второй главе).

1.3. Тезаурус на основе интересов пользователей

Для облегчения поиска, извлечения и упорядочивания информации создаются специальные системы индексации – тезаурусы.

Тезаурус представляет собой словарь-справочник или тематический словарь, содержащий в себе полный вокабуляр для индексирования и извлечения информации [Srinivasan]. Н. В. Лукашевич определяет тезаурус как «словарь, в котором слова и словосочетания с близкими значениями сгруппированы в единицы, называемые понятиями, концептами или дескрипторами, и в котором явно (в виде отношений, иерархии) указываются семантические отношения между этими понятиями» [Лукашевич 2011]. Дескриптор – единица тезауруса.

Первым тезаурусом в современном смысле слова стал английский Тезаурус Роже [Roget 1852]. Составлен британским лексикографом Питером Марком Роже около 1805 г. и опубликован в 1852 г. Структура Тезауруса Роже включает шесть основных семантических категорий, внутри которых построены древовидные иерархии подкатегорий до 8 уровней вложенности [Академик, Тезаурус]. Данный тезаурус является общеязыковым.

Land 342

N. land, earth, ground, dry land, terra firma
continent, mainland, peninsula, chersonese[Fr], delta; tongue of land, neck of land; isthmus, oasis; promontory &c. (projection) 250; highland &c. (height) 206.
coast, shore, scar, strand, beach; playa; bank, lea; seaboard, seaside, seabank, seacoast, seabeach[obs3]; ironbound coast; loom of the land; derelict; innings; alluvium , alluvion[obs3]; ancon.
riverbank, river bank, levee
soil, glebe, clay, loam, marl, cledge, chalk, gravel, mold, subsoil, clod, clot; rock, crag.
V. land, come to land, set foot on the soil, set foot on dry land; come ashore, go ashore, debark
Adj. earthy, continental, midland, coastal, littoral, riparian,; alluvial; terrene &c. (world) 318; landed, predial, territorial; geophilous; ripicolous
Adv. ashore; on shore, on land

Рисунок 4 – Фрагмент словарной статьи тезауруса Роже.

Тезаурусы, особенно в электронном формате, являются одним из действенных инструментов для описания отдельных предметных областей и призваны способствовать пониманию в общении и взаимодействии лиц, связанных одной дисциплиной или профессией. Одним из самых известных тезаурусов в сфере компьютерной лингвистики является компьютерный тезаурус WordNet, появившийся в свободном доступе в 1995 г., представляющий собой общеязыковой тезаурус.

В отличие от толкового словаря, тезаурус позволяет выявить смысл не только с помощью определения, но и посредством соотнесения слова с другими понятиями и их группами, благодаря чему может использоваться для наполнения баз знаний систем искусственного интеллекта.

Вал. А. и Вл. А. Луковы отмечают, что понятие тезауруса в таких гуманитарных науках как социология, культурология, филология, психология, стало применяться недавно и со значением, которое не совпадает с теми, что закрепились в лингвистике и информатике: «оно маркирует ментальные структуры, придающие смысл обыденным действиям людей и их сообществ, но кроме этого предопределяющие самые различные отклонения от обыденности и оказывающие воздействие, возможно — решающее, на весь комплекс социальных структур, социальных и культурных институтов и социокультурных процессов» [Луков, Луков 2008, 17].

Структура тезауруса определяется решаемыми задачами, функциональными требованиями и спецификой лексического наполнения языка предметной области. По Н. В. Лукашевич, цель создания тезауруса – «найти хорошие, компактные слова и фразы для описания основных тем документов, сведя синонимы и квазисинонимы к дескрипторам тезауруса» [Лукашевич 2011]. Иначе говоря, тезаурусный подход позволяет упорядочить терминологию в отдельных сферах и науках.

Э. К. Погорский рассматривает тезаурус как систему концептов, как «полный систематизированный свод освоенных социальным субъектом знаний, существенных для его ориентации в окружающей среде и саморазвитая» [Погорский, 2012]. Тезаурусы, по его мнению, сочетают две характеристики:

- полнота информации, необходимая для ориентации субъекта в социокультурной среде;
- ценностная составляющая такой ориентации – существенность того знания, которое составляет тезаурус;
- систематичность – тезаурус представляется системой концептов.

Ю. Н. Караулов рассматривает два основных типа тезаурусов: общеязыковые и информационно-поисковые. Отмечается, что «принципиальной разницы между общеязыковым и информационно-поисковым тезаурусом нет: тезаурус является лексическим инструментом информационно-поисковых систем. Он состоит из контролируемого, но изменяемого словаря терминов, между которыми указаны смысловые связи» [Караулов 1981].

Общеязыковые тезаурусы, как правило, входят в число ономаσιологических словарей, т.е. в отличие от алфавитных словарей, которые посвящены словам как единицам речи, они представляют связь от понятия к слову, размещая слова по их смысловой близости, облегчают достижение ближайшей к намеренному смыслу лексики. С этой точки зрения нет существенных различий между тезаурусом и идеографическим словарём [Лафтими 2011].

Специальный информационно-поисковый тезаурус имеет большое применение в информационном поиске в различных областях науки и техники. Тезаурус является совокупностью нормализованных и

развивающихся терминов специальной области науки, которые связаны друг с другом семантическим способом.

И. Лафтими выделяет следующие функции информационно-поискового тезауруса [Лафтими 2011]:

- как средство перевода с естественного языка на информационно-поисковой язык при индексировании информационных документов и запросов;

- как источник сведений о полезных для поиска отношениях между дескрипторами:

- как терминологическое пособие при решении различных научно-информационных задач (в частности, предоставляет информацию о составе лексики, используемой при индексировании документов).

Н. И. Кондаков выделяет немного другие типы тезаурусов [Кондаков 1975, 587]:

- лингвистический тезаурус – словарь дескрипторов естественного языка, систематизированных в соответствии с принятой классификационной системой;

- статистический тезаурус – информационно-поисковой словарь, содержащий дескрипторы текстов по определенной тематике и сгруппированных в словарные статьи на основе частоты их встречаемости.

Однако по своим принципам они отражают ту же суть, что и общеязыковые и информационно-поисковые тезаурусы.

По языковой направленности тезаурусы бывают одноязычными и многоязычными. Многоязычный тезаурус содержит термины из нескольких естественных языков и представляет эквивалентные по смыслу понятия в каждом из них.

По своему составу тезаурусы делятся на тезаурусы, все единицы которых являются дескрипторами (предпочтительными терминами) и тезаурусы, в состав которых входят как дескрипторы, так и аскрипторы (непредпочтительные термины) [Савкин 2014].

П. Сринивасан разделяет построение тезауруса на основе текстов на три этапа [Srinivasan]: 1) construction of vocabulary (построение вокабуляра); 2) similarity computations between terms (вычисление схожести терминов); 3) organization of vocabulary (организация вокабуляра).

Построение вокабуляра предполагает выбор лексических единиц и их нормализацию. Слова и фразы могут браться из любой части текстового документа – названия, введения, полного текста. При составлении информационно-поискового тезауруса интересов пользователей социальной сети дескрипторы могут отображаться в следующих текстовых пространствах сообщества (подробнее рассмотрены в пункте 2.1):

- 1) в названии сообщества;
- 2) в описании сообщества;
- 3) в тексте постов на стене сообщества;
- 4) в комментариях к постам;
- 5) в разделе «обсуждения».

Пространства 1 и 2 облегчают задачу определения интереса сообщества. В отсутствие первых двух пунктов (сообщество с (абсолютно) несвязанным с интересом названием и отсутствием описания/наличием непонятого и опять же несвязанного с интересом описания) анализируется текст из пространств 3, 4 и 5. При отсутствии пунктов 1, 2, 3 (на стене постятся изображения (напр. комиксы), аудио и видеозаписи без текстового (или с совсем малым и недостаточным для выявления интереса объемом) сопровождения)

задействуются пункты 4 и 5. В отсутствие всех 5 пространств лингвистический анализ интересов становится недоступным. Исключением в таком случае может быть присутствие дескрипторов на изобразительном пространстве – текст на выложенных изображениях или текст видео-контента (как устная, так и письменная речь), но при этом будет невозможен автоматический поиск – только сбор данных вручную.

При машинном отборе фраз используется, например, алгоритм Choueika Procedure [Srinivasan], который можно применить и к ручной обработке текста:

1. обозначить дозволенный лимит фразы (напр. от 2 до 5 слов);
2. выстроить лист потенциальных фраз, подходящих под лимит;
3. удалить фразы, которые начинаются или заканчиваются на стоп-слова;
4. удалить фразы, которые содержат большое количество нетривиальных слов;
5. оценить потенциал подвыражений и избавиться от неподходящих;
6. попытаться соединить меньшие выражения в одно большое и более значимое.

Нормализация терминов подразумевает избавление от стоп-слов и лемматизацию лексики. Для существительных это именительный падеж, единственное число; для глаголов это форма инфинитива.

Вычисление схожести терминов предполагает значимые статистические ассоциации между терминами. Так, при кластеризации интересов в пределах тезауруса будут выявлены связи между интересами, а также на основе повторяющихся в разных тематических сообществах терминов будет легче определить объединяющую их лексику. Например, сообщество ЛГБТ-

знакомств связывает два главных интереса: ЛГБТ+ и виртуальные знакомства.

ИГРЫ	СПОРТ	РОЛЕВЫЕ ИГРЫ
игра		
играть		<u>ролевая по мультфильму</u>
игрок	NHL	<u>Аватар (A)</u>
команда	<u>StanleyCup (ТН)</u>	<u>АТЛА (ТН)</u>
матч	Бишоп	<u>Legend_of_Aang (ТН)</u>
	Блюз	<u>Legend_of_Korra (ТН)</u>
ВИДЕОИГРЫ	Бостон	<u>аватар</u>
	<u>Брюинз</u>	<u>Аанг</u>
Counter Strike CHAMPIONSCLUB	гол	<u>Корра</u>
<u>Swop (SV)</u>	Даллас	<u>Азула</u>
<u>Steam</u>	забивать гол	<u>ЛоА</u>
админ	<u>игра</u>	<u>ЛоК</u>
администрация	<u>игрок</u>	Земля
банить	Каролина	Огонь
<u>игра</u>	<u>конференция</u>	Вода
<u>играть</u>	<u>матч</u>	Воздух
<u>игрок</u>	обыгрывать	<u>кроссовер</u>
карта	<u>плей.офф (SV)</u>	маг
моделька	<u>победа</u>	магия
ник	серия	<u>ролевая игра</u>
<u>сервак (SV)</u>	финал	<u>Сато</u>
сервер	шайба	Тай
сервера CHAMPIONSCLUB	<u>Шаркс</u>	<u>уткочереха</u>
<u>скрин</u>		

Рисунок 5 – Скриншот – Пример MaI «ИГРЫ» из тезауруса интересов.

Последний шаг – организация вокабуляра – возможно, самый сложный, так как предполагает упорядочивание полученного лингвистического материала в виде иерархии на основе выделенных в предыдущем шаге пересечений дескрипторов. Для экспериментального тезауруса интересов эта иерархия будет выглядеть следующим образом: главный интерес – интерес – тематика сообщества. Например, главный интерес (umbrella interest) ИГРЫ включает такие доминирующие интересы сообществ как видеоигры, спорт, ролевые игры.

Тезаурусу также можно присвоить систему обозначений, описывающих отношения терминов. Одной из таких систем является Zthes, выделяющая следующие типы отношений [Барахнин 2007]:

- BT – связь с родительским термином – с термином более широкого смысла;
- NT – связь с дочерним термином – более узкого смысла;
- USE – связь с термином, который используется вместо этого;
- UF – взаимно-обратная связь USE;
- RT – связь, определяющая связанный по смыслу термин;
- LE – связь между лингвистически эквивалентными терминами;
- FE – полностью тождественные термины.

Похожие отношения можно найти в системе тезауруса INSPEC [Srinivasan], который включает дескрипторы таких областей, как физика, электро-инженерия, электроника, компьютерные технологии. Он распределяет терминам отношения NT, BT, USE, UF, RT и отличается двумя другими типами:

- TT – top term (главный термин, зонт);
- CC – classification code (код классификации).

Для разработанного тезауруса интересов мы используем иную систему отношений с элементами системы Zthes. Это гибридная форма, более подходящая для реализации поставленных целей (подробно описана во второй главе).

Важно понимать, что составление одного огромного тезауруса интересов всех сообществ виртуального пространства социальной сети пока невозможно ввиду еще несуществующих автоматически выявляющих интересы приложений, а также постоянно расширяющегося в объеме (создание новых аккаунтов, сообществ) пространства. Любой тезаурус не может быть полным ввиду скорости развития культуры. Тем не менее,

попытки анализа интересов пользователей интернет-сообществ, организованного в форме тезауруса дескрипторов интересов этих сообществ, приближают к возможности понимания картины мира пользователей виртуальных социальных сетей, к возможности упростить поиск нужного среди бесконечного потока информации, а также более точного определения темы потенциально интересных информационных текстов, к принятию которых открыт пользователь, и авторитетов, чье мнение важно для данного пользователя и влияет на его поведенческие установки.

Выводы по Главе 1

В Главе 1 настоящего исследования обосновывается теоретический аппарат исследования. Центральным является вопрос о признаках и особенностях виртуального информационно-коммуникационного пространства. Так было выделено 2 классификации социальных сообществ в целом: 1) социальная группа, квазигруппа и социальный круг; 2) сообщества практики и сообщества по интересам. Несмотря на возможность разделения на данные категории и виртуальных, и реальных сообществ, их признаки отличаются. К признакам виртуального сообщества относятся преобладание письменной речи, анонимность, «раздвоение личности», территориальная и временная неограниченность, необязательное взаимодействие вне виртуальной среды.

Далее была дана характеристика лингвистических особенностей виртуального языка, где главными отличиями являются анонимность, свобода от норм языка, территориальная и временная независимость, наличие хэштегов и мэншонов, наличие графических невербальных элементов – эмодзи, эмотиконы, стикеры, мемы. Важным компонентом виртуального информационного пространства также выступает фолксономия – народный теггинг, который позволяет упорядочивать не всю, но достаточно большую часть имеющейся и появляющейся информации.

Тезаурус – это концептуальный словарь-справочник. Для построения тезауруса необходимо понимать его цели и сферы применения. Построение информационно-поискового тезауруса включает три этапа: 1) построение вокабуляра – отбор необходимой лексики и нормализация; 2) вычисление схожести терминов – выявление пересекающихся дескрипторов; 3) организация вокабуляра – построение иерархических отношений. Для разных целей могут быть использованы разные схемы отношений, как существующие, так и собственно сгенерированные. В полученном тезаурусе интересов пользователей социальных сетей иерархия построена по

следующему принципу: главный интерес – доминирующий интерес – направленность сообщества. Пространство для сбора дескрипторов в виртуальном сообществе располагается в следующих текстовых пространствах сообщества: 1) в названии сообщества; 2) в описании сообщества; 3) в тексте постов на стене сообщества; 4) в комментариях к постам; 5) в разделе «обсуждения». Дескрипторы могут быть выражены лексическими единицами любого языка (преимущественно русского).

Тезаурусный метод организации всех существующих интересов в виртуальном пространстве пока невозможен, но попытки создания такового выведут поиск информации на новый уровень.

ГЛАВА 2. ТЕЗАУРУСНЫЙ МЕТОД ОРГАНИЗАЦИИ ИНТЕРЕСОВ

2.1. Анализ интересов

Анализ интересов – это некоторый набор задач, которые ставятся в маркетинге, криминалистике и других сферах деятельности. Суть всех задач сводится обычно к кластеризации сообществ или выявлению сообщества из массы других на основе каких-либо признаков социальной активности. Эти признаки и обозначаются словом «интерес».

Как уже было отмечено, в данной работе, нас интересует главный интерес с двух перспектив: 1) как доминирующий интерес сообщества в парадигме интересов; 2) как зонтичный термин (*umbrella interest*).

Для того, чтобы определить доминирующий интерес сообщества, можно ввести термин «дескриптор интереса», который будет означать лексический маркер, характерный для того или иного интереса (на любом языке). Дескрипторы интересов работают в пределах трех следующих компонентов, в комбинациях или по отдельности: субъект, предикат и объект.

Для поиска дескрипторов можно выделить следующие этапы:

1. выделение ключевых слов;
2. определение частоты выделенных ключевых слов;
3. выделение типичных предикатных конструкций (субъект, предикат, объект);
4. выделение предикатов для ключевых слов;
5. определение тональности ключевых слов и предикатов – сентиментный анализ.

Сентиментный анализ (*sentiment analysis*) – направление компьютерной лингвистики, посвященное анализу тональности высказывания. Его принцип

– определение мнения пользователя по отношению к объекту интереса. Тональность может быть представлена на шкале «негативный – позитивный» либо в виде определения эмоции говорящего (классификация эмоций также может быть разной). Как правило, лексическими маркерами тональности могут выступать все значимые части речи (чаще прилагательные и глаголы). Также эмоции могут проявляться невербально посредством смайликов, стикеров, изображений, gif-изображений (гифки), видео- и аудиоматериалами. Ранее было отмечено, что дескрипторами для тезауруса выступают ключевые понятия, однако при определении интереса важно отметить эмоциональные дескрипторы, которые вербально отражаются через прилагательные (ламповый – уютный), наречия (грустно, жизненно), специфическую лексику (жиза, кул, топ) и глагольные выражения (разбивать сердечко).

Дескрипторы интереса отображаются в текстовых пространствах сообщества, которые уже были отмечены в главе 1 (см. раздел 1.3), могут быть выражены не только русским языком, который преобладает ввиду большего количества русскоязычных пользователей, но и любым другим языком (напр. английским); рассмотрим их подробнее:

1) в названии сообщества («Greenpeace России»);

2) в описании сообщества («Greenpeace — это движение, объединяющее людей, которые хотят сохранить природу и мир на планете» - Greenpeace, сохранить природу);

3) в тексте постов на стене сообщества («Сегодня, 24 мая, проходят международные климатические забастовки #FridaysForFuture. Глобальное изменение климата — наша общая проблема» - климатические забастовки, #FridaysForFuture, глобальное изменение климата);

4) в комментариях к постам («вот вам весьма реальный ответ, где за одну ночь допустимые нормы загрязнения воздуха просто взяли и увеличили в несколько раз» - загрязнение воздуха);

5) в разделе «обсуждения» («О том кто такой волонтер Гринпис и как им стать написано здесь...») - волонтер, Гринпис).

Как уже отмечалось ранее, не каждое текстовое поле из вышеперечисленных может нести в себе достаточно информации или же может вообще отсутствовать. Сообщества, где есть подробное описание, но нет большой смысловой нагрузки в остальных текстовых пространствах, не пригодны для извлечения дескрипторов.

В данном исследовании выделяются следующие лингвистические признаки интереса, установленные в ходе эксперимента:

- в синтаксических отношениях интерес в основном выступает в роли объекта (смотреть *кино*, слушать *музыку*);

- интерес частотен – часто присутствует в текстовом пространстве (при очистке текста от стоп-слов);

- интерес выражается полнозначными словами (напр. кино, музыка, комиксы, руны);

- интерес выражается такими частями речи, как существительное (поэзия), глагол (танцевать), отглагольное существительное (рисование), а также глагольными словосочетаниями (писать *музыку*, организовывать *вечеринку*) и именной группой (замок с привидениями, известные люди, испанские сериалы).

- интерес также определяется эмоциональными маркерами (sentiment analysis).

В зависимости от типа контента в сообществе, некоторые признаки могут отсутствовать. Так, например, в сообществе с отсутствием текста на стене, но присутствием описания и/или комментариев интерес не обязательно будет частотным, но указанные лексические единицы, несущие информационную нагрузку об интересе, будут включены в тезаурус как дескрипторы.

Еще одна важная деталь для выявления интереса – это присутствующие в сообществе наименования. На материале проанализированных текстов, собранных с виртуальных социальных сообществ, можно выделить следующие категории наименований:

- имена значимых личностей или персонажей (для участников сообщества «Twenty One Pilots» это имена участников группы – Тайлер Джозеф и Дшош Дан; для участников сообщества «Гвин Пикс» это имена персонажей – агент Купер и Лора Палмер, а также имя режиссера – Дэвид Линч); сюда также входят имена богов (для сообщества «Мир Рун» - скандинавский бог Тор);

- имена «админов» и значимых участников сообщества (для сообщества «Страна Игр» это имена журналистов, также пишущих посты для паблика - Егор Просвирин и др. – и имя главного редактора - Константин Говорун);

- наименования команд (для участников сообщества «NHL» это названия хоккейных команд – Шаркс, Бостон Брюинз и др.);

- наименования брендов (для сообщества «РадиоВидео.рф» это бренды раций: «Baofeng», «Kenwood», «Аргут»);

- наименования значимых произведений – фильмов, сериалов, музыкальных и литературных произведений и др. (для участников сообщества «дочери десятой музыки» значимыми являются кинокартины «Джентльмен Джек» и «Убивая Еву»);

- наименования мест сбора и мероприятий (для участников сообщества антикафе «Каледонский Лес» это помещения антикафе на двух станциях метро Санкт-Петербурга – «Лес на Садовой» и «Лес на Чкаловской»; для сообщества «Beat Film Festival» это названия мероприятий «Beat Film Festival», «Каннский кинофестиваль», фестиваль «Вудсток»);

- наименования значимых организаций («ядром» сообщества «Blizzard Entertainment» является компания Blizzard);

- наименование участников сообщества (участники сообщества антикафе «Республика Кошек» называются «республиканцы»; участники социальной группы «Сыроедение - Я ФРУКТОРИАНЕЦ» называют себя «сыроежки»).

Метод взвешивания частотности терминов основывается на идее о том, что в исследуемом текстовом материале контекстная информация об интересах или главных темах проявляется с большей частотностью. Но его использование в чистом виде недостаточно для выявления интересов, так как в социальных сетях встречается большое количество информации, не описывающей интересы пользователя. Так, например, группа исследователей Jeongin Kim, Dongjin Choi, Byeongkyu Ko, Eunji Lee и Pankoo Kim (2014) предлагают добавить в анализ интересов подсчет «лайков».

Также одним из важных маркеров интересов является хэштег. Хэштегами, как правило, отмечают значимые тематики информации постов. По хэштегам сообщества можно определить его направленность (однако, не каждое сообщество ими пользуется): например, если посмотреть у сообщества «Nordic Design» встречающиеся хэштеги #madeinfinland, #nordicarchitecture, #swedishdesign, #finnishdesign, то сразу становится понятно, что речь в данном сообществе идет о стилях и дизайнерских решениях скандинавских стран. Хэштеги также облегчают задачу определения интереса у сообществ, которые посвящены изображениям

(мемы, скриншоты из фильмов/сериалов, фотографии, «гифки»), видеозаписям или аудиозаписям (которые также могут сопровождаться статичными и gif-изображениями), но в которых при этом нет никаких дескрипторов интересов ни в одном из текстовых пространств. Например, у молодого сообщества (первая запись от 12 мая 2019 года) «your thoughts .» нет никаких дескрипторов, кроме хэштегов: #cartoons #adventuretime #art #animation и др. По хэштегам можно понять, что сообщество «постит» скриншоты из анимационных сериалов и мультфильмов.

Определение доминирующего интереса сообщества затрудняется еще и тем фактом, что анализируемое сообщество может быть посвящено нескольким интересам. Например, что считать за доминирующий интерес в сообществе «ЛГБТ \знакомства\ 16+»: ЛГБТ+ или виртуальные знакомства? В такой ситуации нужно отталкиваться от того, что конкретно в этом сообществе ищут пользователи: ищут ли они информацию о ЛГБТ+ или же они ищут знакомства. Более того, можно опираться на преобладающие дескрипторы интереса. В паблике «ЛГБТ \знакомства\ 16+» выделены следующие дескрипторы (фрагмент из полученного тезауруса – Приложение 1):

актив

анкета

беседа

внешность

возраст

гулять

девушка

друг

дружба

знакомство

ЛГБТ

ЛГБТ_Анонимно

ЛГБТ_Би

ЛГБТ_Лесби

любовь

общаться

общение

отношения

парень

романтика

Дескрипторов интереса «виртуальные знакомства» (анкета, беседа, знакомство, любовь, общаться, общение, отношения) значительно больше, что указывает на доминирование этого интереса над интересом «ЛГБТ+».

Однако лексические единицы не всегда могут отобразить доминирующий интерес. Например, сообщество «Логомашина: дизайн логотипов и фирстилей» содержит большое количество лексики, присущей доминирующему интересу «ИЗО», но при этом данное сообщество ориентировано на предоставление коммерческих услуг по разработке логотипов и фирменных стилей:

logomachine.ru

айдентика

дизайн

Логомашина

логотип

портфолио

посмотреть портфолио

презентация

проект

продукт

разрабатывать логотип

разрабатывать айдентику

стиль

фирменный стиль

На доминирующий интерес «услуги» здесь указывает описание сообщества: «Делаем логотипы для клиентов со всего мира и пишем жгучие посты. Наш сайт и портфолио — logomachine.ru. Позвоните нам, вас проконсультирует менеджер по продукту Василий — ... Напишите нам, вам поможет менеджер клиентского сервиса Сергей» [id:logomachine]. В описании говорится об услуге по разработке логотипов на заказ, указывается на наличие клиентов со всего мира, есть информация о предоставляющем услуги сайте и о наличии портфолио, есть контактная информация и официальные представители, готовые проконсультировать клиента по продукту и оказать необходимую помощь. То есть здесь важна не столько частотность дескриптора и не столько однозначность ключевых слов, сколько концепция и цели сообщества, которые нельзя не учитывать, так как анализ проводится вручную и позволяет обращать внимание на такие детали.

Для того чтобы указать связь сообщества не только с доминирующим интересом, но и с интересом из другой группы, в тезаурусе обозначены тематические пересечения, основанные на анализе каждого сообщества на наличие одинаковых, синонимичных или присущих другой тематике значимых лексических единиц. Раздел называется «анализ лингвистических пересечений» и представлен в Приложении 2.

В ходе анализа интересов 95 сообществ было выделено 12 главных интересов (*umbrella interest*) и 21 доминирующий интерес сообществ:

- искусство (А) (музыка, литература, ИЗО, кино);

- рынок (**Р**) (товары, поиск работы, услуги, антикафе, магазины);
- наука (**Н**) (природа, психология, иностранные языки, история);
- игры (**И**) (видеоигры, ролевые игры, спорт);
- здоровье (**З**) (правильное питание, медицина, ментальное здоровье, самоубеждение);
- активизм (**АКТ**) (защита людей, защита животных);
- еда (**Е**);
- ЛГБТ+ (**Л**);
- сэйфспейс (**SS**) – безопасное место (от англ. safe space);
- животные (**Ж**);
- виртуальные знакомства (**ВЗ**);
- религия (**РЕЛ**).

«Сэйфспейс» подразумевает безопасное пространство, обычно необходимое пользователям, притесняемым обществом (не обязательное условие). Для более точного его понимания были проанализированы еще несколько сообществ без включения в тезаурус. В целом это сообщества, где пользователь чувствует себя безопасно и/или уютно. Так, сообщество «stop hurting yourself» предлагает его участникам безопасную, поддерживающую атмосферу, направленную на помощь людям, склонным к самоистязанию (селф-харму): «Паблик создан на добровольных и некоммерческих началах для людей, которые занимаются селф-хармом, с целью поддержки, взаимопомощи, принятия и возможном уменьшении и/или замещении практики селф-харма на практики, направленные на самоадвокацию, принятие, поддержку и развитие личности» [id: stophurtingyourself]. Такие сообщества предоставляют людям возможность быть понятыми,

«защищают» их интересы, считают проблемы участников значимыми (что очень важно), тем самым создавая безопасную виртуальную площадку. Конкретно для данного интереса присуще небольшое количество дескрипторов, например: «безопасное пространство», «помощь», «поддержка», «принятие». Однако он включает в себя большое количество дескрипторов интересов (АКТ) и (З) (в основном доминирующего интереса «ментальное здоровье»): феминизм, сексизм, насилие, суицидальные мысли, селф-харм (от англ. self-harm), психологическая помощь и др.

Самым многочисленным по количеству виртуальных сообществ получилась категория «искусство» - включает в себя 26 (27,4%) сообществ 4 категорий: музыка, литература, изо и кино.

39 сообществ (41,05%) посвящены нескольким интересам, имеют от 1 до 3 связей и отнесены к доминирующему интересу, которому больше уделяется внимания в сообществе. В результате анализа лексических пересечений интересов была получена следующая схема связей главных интересов между собой (Приложение 2):

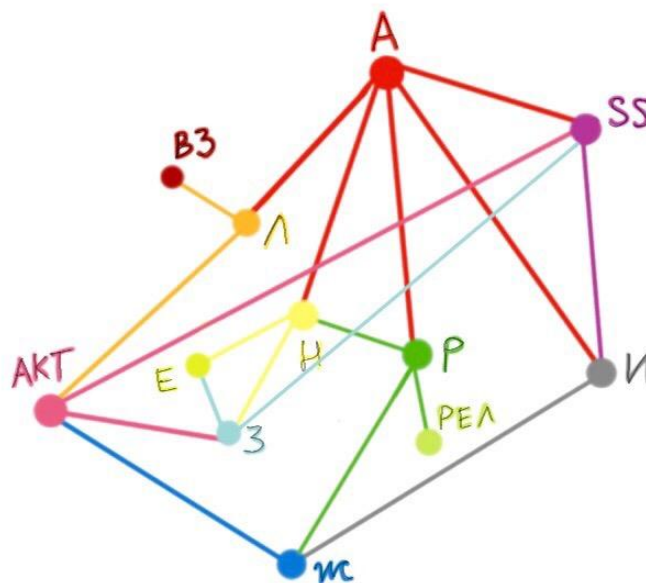


Рисунок 6 – Схема лексических пересечений главных интересов.

В результате все 12 главных интересов по отдельности имеют разные лексические пересечения (от 1 до 5). Большая доля пересечений выпадает на интерес «искусство» (А) – 5 связей. Далее идут «активизм» (АКТ), «наука» (Н), «рынок» (Р), «здоровье» (З) и «сэйфспейс» (SS) – по 4 связи. Затем «игры» (И), «ЛГБТ+» (Л), «животные» (Ж) – по 3 связи. 2 связи имеет интерес «еда» (Е). По 1 связи имеют интересы «религия» (РЕЛ) и «виртуальные знакомства» (ВЗ). На основе данной иерархии составлено содержание тезауруса.

Рассмотрим, например, список дескрипторов сообщества «PLAY LIKE A GIRL» из интереса (SS) – безопасное женское развлекательно-игровое пространство про-фем направленности:

art@playlikeagirl

DragonAge@playlikeagirl

kassandra

Life is Strange

overwatch@playlikeagirl

админка

геймерская культура

геймерша

героиня

женское пространство

марвел

мейл-фри

мизогиния

паладинка

персонажка

сексизм

феминитивы

Здесь присутствуют 3 лексические связи: **(И)** - DragonAge@playlikeagirl, Life is Strange, геймерская культура, геймерша и др; **(А)** – сериалы, марвел, art; **(АКТ)** – сексизм, мизогиния, феминитивы. Несмотря на преобладание лексики игровой направленности, главный интерес данного сообщества – безопасное пространство (safe-space) для девушек-геймерш.

Для сообщества из группы **(SS)** «lovely days» – уютное виртуальное пространство – также характерна связь с интересом **(А)** (его дескрипторы подчеркнуты):

*аниме**вп (SV)**душевная группа**манга**милые посты**музыка**сериалы**твиттер**уютные посиделки**фильм*

Рассмотрим теперь дескрипторы того же интереса **(SS)** из сообщества «stop hurting yourself»:

*stop_hurting_youself**tw**бить**боль*

больно

делать больно

истерика

кровь

ЛГБТКИА+

порез

практика селф-харма

пси-помощь

резать

романтизация селф-харма

саморазрушение

селф-харм

селфхарм

суицидальные мысли

физическое насилие

шрамы

Здесь присутствует связь с главным интересом **(З)**: истерика, пси-помощь, саморазрушение, селф-харм/селфхарм, суицидальные мысли. В результате интерес (SS) имеет 4 лексические связи: **(И)**, **(А)**, **(АКТ)**, **(З)**. По такому принципу присутствия у сообществ внутри главного интереса пересекающихся дескрипторов была получена схема лексических пересечений главных интересов.

Для некоторых интересов можно выделить лексические единицы, встречающиеся в большинстве объединенных им сообществ. Например, для главного интереса «игры» выделены общие для доминирующих интересов (видеоигры, ролевые игры, спорт) лексические единицы: игра, играть, игрок, команда, матч. В тезаурусе они обозначены в начале раздела интереса (см.

рис. 5). Также после кластеров сообществ (от 2х сообществ) обозначены повторяющиеся лексические единицы с меньшим появлением в сообществах. Так, у кластера «история» из главного интереса «наука» выделены следующие повторяющиеся лексические единицы: история, кровь, преступление, убийство. Однако у интереса «сэйфспейс», которое включает 3 сообщества, нет совпадающих лексических единиц.

2.2. Составление тезауруса интересов пользователей

В исследовании тезаурус интересов пользователей социальных сетей – эксперимент по организации лингвистических признаков интереса в некую систему и иерархию. По структуре полученный тезаурус является информационно-поисковым и многоязычным (включает лексические единицы русского, английского, испанского языков), состоит только из дескрипторов, функционирует как пособие лингвистических признаков интересов пользователей.

Для каждого сообщества обозначено до 20 наиболее частых, значимых дескрипторов, отражающих некоторую совокупность знаний и концептов об интересе, которыми владеет пользователь. Значимость дескрипторов определялась на основе принципа их расположения (в одном из 5 текстовых пространств). Самым значимым выступают название и описание сообщества (при наличии такового), далее идут тексты постов, затем обсуждения, и, наконец, комментарии. Данный вопрос считается спорным и субъективным, но такая иерархия была взята ввиду следующих факторов:

- название и описание сообщества являются его «лицом» - это первое, на что смотрит пользователь, чтобы понять, «о чем здесь идет речь»;

- тексты постов – это суть сообщества; в них описывается и обговаривается та информация, что отражает интерес сообщества и его пользователей; присутствуют в большей части активных сообществ;

- раздел «обсуждения» включает информацию, важную для понимания организации сообщества (правила), а также предоставляет возможность админам налаживать связь с участниками (пожелания участников, FAQ, помощь участникам, пространство для общения между участниками и др.);

- в комментариях нет структурированного общения в рамках одного интереса – пользователи, которые оставляют комментарии, располагают

разными наборами знаний; преимущественно рассчитаны на выражения мнения и эмоций.

Значимость пространств меняется при условиях разных комбинаций наличия текстовых пространств в сообществе. Например, если в сообществе нет текстов в постах и/или нет раздела «обсуждения», значимыми единицами будут выступать комментарии участников.

Так как в тезаурусе представлена лексика виртуального языка, включающая хэштеги, акронимы, сленг и различные наименования, внутри парадигмы дескрипторов есть свои связи. На основе системы отношений терминов Zthes и системы тезауруса INSPEC, в тезаурусе для дескрипторов были обозначены следующие связи:

- SYN – synonym – синоним: синонимичными обозначены лексические единицы, связанные с интересом (напр. виниловые вечеринки – пати);

- FE – full equivalent – полный эквивалент: обычно обозначает эквиваленты на разных языках (напр. виниловая вечеринка – vinyl party); также обозначает вариант написания (напр. эмо – имо; Блэксэд – Блэксэд);

- EM – emotional marker – эмоциональный маркер: элемент сентиментного анализа (напр. для интереса «перевод песен emo»: «вырывать сердечко», «грустно», «жиза», «разбивать сердечко»);

- SV – specific vocabulary – специфическая лексика: используется в рамках данного интереса (напр. для сообщества «Республика Кошек»: «замурчательный», «республиканцы», «укотовление»);

- TH – thematic hashtag – тематический хэштег (напр. для сообщества «Russian Travel Geek»: #russiantravelgeek, #scienceandtravel).

В скобках с интересом сообщества обозначена связь по лингвистическим признакам с другими (от 1 до 3) главными интересами

(напр. у сообщества «PLAY LIKE A GIRL» из главного интереса «сэйфспейс» есть 3 связи – **(И) (А) (АКТ)**). Схема связей всех групп интересов между собой представлена на рисунке 6 (см. раздел 2.1).

Для некоторых дескрипторов необходимо было указать маркеры эмоционального отношения к словам, воспринимаемым подавляющим большинством пользователей иначе:

- neg (negative) – в рамках интереса слово с нейтральной или позитивной коннотацией значением приобретает негативную коннотацию (напр. для сообщества «Рёбра Евы», которое посвящено борьбе с гендерной дискриминацией, слова «мужчина» (нейтральное) и «сексуальность» (позитивное) принимают негативную коннотацию ввиду борьбы за права женщин и борьбы против сексизма и гендерных стереотипов);

- pos (positive) – в рамках интереса слово с нейтральной или негативной коннотацией приобретает позитивную коннотацию (напр. для сообщества «Soc-Fem», которое посвящено социалистическому феминизму, слово «аборт» (негативное) приобретает позитивную коннотацию – участницы борются за право женщинам делать аборт, особенно ввиду новых законов о запрете абортов в США).

Все условные обозначения в тезаурусе указаны в скобках рядом с лексическими единицами, к которым они относятся. Рассмотрим в качестве наглядного примера список дескрипторов к сообществу «Neologeek» из главного интереса **(А)**, посвященное виниловым вечеринкам:

new wave

retro wave

synth party

vinyl party (FE)

билет

винил

виниловая пластинка

движ (SV)

для настоящих меломанов

жаркая ночь (EM)

крутить винил

на виниле

неологики (SV)

по-неологиковски (SV)

отрываться

пати (SYN)

проходка (SV)

синти-поп

танцевать

теплая вечеринка (EM)

фотки

Перейдем к строению тезауруса (Приложение 1). Главный интерес прописан заглавными буквами, жирным шрифтом и выделен маркером серого цвета (напр. **АКТИВИЗМ**). Далее доминирующие интересы (если таковые выделяются) выделены заглавными буквами и жирным шрифтом (напр. **ЗАЩИТА ЖИВОТНЫХ**). Каждое сообщество, входящее в кластер того или иного интереса, включает название сообщества и его краткое описание при необходимости; выделено жирным шрифтом (напр. **Greenpeace России**). Повторяющиеся лексические единицы расположены списком после кластеров (повторяющиеся лексические единицы); такие дескрипторы внутри кластера выделены нижним подчеркиванием.

Некоторые главные интересы включают кластеры сообществ, объединенные одним доминирующим интересом. Так, интерес «наука» включает в себя 4 кластера:

- природа («The Batrachospermum Magazine» – развлекательный научно-популярный журнал; «Russian Travel Geek» – научно-популярные экспедиции с молодыми учёными);

- психология («Когнитивный патимейкер» – когнитивистика и мемы; «Саморазвитие»);

- иностранные языки («Я – переводчик» – информация для переводчиков; «english yo» – английский язык 18+; «испанский язык»; «ЕГЭ по английскому»; учебник «New English File 3rd edition»);

- история («Древний мир: история и факты»; «Мир Маньяков и Серийных Убийц»; «История: пытки, казни, телесные истязания»; «ЕГЭ ОГЭ | История»).

Структура тезауруса построена на основе иерархии по количеству имеющихся у главного интереса связей (рис. 6). Разделы дескрипторов идут в порядке по убыванию – от большего по количеству лингвистических пересечений интереса (искусство) к меньшему (религия).

Содержание тезауруса интересов без указания страниц:

1. ИСКУССТВО

- музыка
- литература
- изо
- кино

2. РЫНОК

- товары
- поиск работы

- услуги
- антикафе
- магазины

3. НАУКА

- природа
- психология
- иностранные языки
- история

4. АКТИВИЗМ

- защита людей
- защита природы

5. ЗДОРОВЬЕ

- правильное питание
- медицина
- ментальное здоровье

6. СЭЙФСЕЙС

7. ИГРЫ

- видеоигры
- ролевые игры
- спорт

8. ЛГБТ+

9. ЖИВОТНЫЕ

10. ЕДА

11. ВИРТУАЛЬНЫЕ ЗНАКОМСТВА

12. РЕЛИГИЯ

Дескрипторы каждого сообщества расположены в алфавитном порядке (сначала английские и испанские, при наличии). Английские лексические единицы чаще проявляются в тематических хэштегах. Интересен также тот факт, что многие русскоязычные сообщества носят английское название

(напр. «Bronze» – визуальная культура; «Russian Travel Geek» – научно-популярные экспедиции с молодыми учёными). Некоторые сообщества «развивают» иностранные интересы, ввиду чего используют вокабуляр иностранного языка, который далее может перейти в активный лексикон пользователя, затем превращается в слово, написанное транслитом, входит в естественный язык и теряет свои «корни» в сознании говорящих. Например, в сообществе «Indie Music» используется большое количество английских лексических единиц, описывающих жанры музыки:

Alternative

Alternative-pop

Electronic

Electropop

Experimental

Hip-Hop

Indie (FE)

indie-pop

Industrial

Pop

Punk

Rock

Synth-Pop

дебютный альбом

музыкант

новый альбом

плейлист

Лексические единицы hip-hop, synth-pop, indie, indie-pop и electropop имеют русскоязычную форму написания и произношения: хип-хоп, синт-поп/синти-поп, инди, инди-поп, электропоп. Слово «плейлист» используется

в сообществе только в русскоязычной форме и адекватно воспринимается в естественном языке, хотя оно тоже перешло в речь от английского playlist, но в такой форме используется реже (чаще как хэштег или его элемент).

Испанский язык встречается в сообществе «Испанские сериалы и фильмы» для указания оригинальных имен актеров и актрис и названий фильмов:

Blanca Suárez

blancasuarez (ТН)

elinternado (ТН)

Laguna Negra

lagunanegra (ТН)

mariocasas (ТН)

актер

актриса

испанские актрисы

испанские сериалы

испанский кинематограф

на языке оригинала

сезон

сериал

фильм

Черная Лагуна

Выводы по Главе 2

В главе 2 были описаны две составляющие эксперимента: анализ интересов и оформление полученных интересов в тезаурус. Под анализом интересов понимается выделение из сообщества интереса посредством лингвистического анализа – определения дескрипторов. Были выделены следующие лингвистические признаки интереса:

- в синтаксических отношениях интерес в основном выступает в роли объекта;

- интерес частотен – часто присутствует в текстовом пространстве (при очистке текста от стоп-слов);

- интерес выражается полными словами;

- интерес выражается такими частями речи, как существительное, глагол, отглагольное существительное, а также глагольными словосочетаниями и именной группой;

- интерес также определяется эмоциональными маркерами (sentiment analysis);

Еще одна важная деталь для выявления интереса – это присутствующие в сообществе наименования. Были выделены следующие категории наименований:

- имена значимых личностей или персонажей;

- имена «админов» и значимых участников сообщества;

- наименования команд;

- наименования брендов;

- наименования значимых произведений – фильмов, сериалов, музыкальных и литературных произведений и др.;

- наименования мест сбора и мероприятий;
- наименования значимых организаций;
- наименование участников сообщества.

Также одним из важных маркеров интересов является хэштег.

Однако, для точного определения интереса сообщества важны не только лингвистические признаки, но и концепция и цели сообщества.

При анализе 95 моно- и мультязыковых сообществ было получено 12 главных интересов, и также был выделен 21 доминирующий интерес.

Также был проведен анализ лингвистических пересечений, в результате которого выделено 39 сообществ (41,05%), которые имеют от 1 до 3 связей и отнесены к доминирующему интересу.

По структуре полученный тезаурус является информационно-поисковым и многоязычным, состоит только из дескрипторов, функционирует как пособие лингвистических признаков интересов пользователей.

Для каждого сообщества обозначено до 20 наиболее частых, значимых дескрипторов. Значимость дескрипторов определялась на основе принципа их расположения (в одном из 5 текстовых пространств). Дескрипторы каждого сообщества расположены в алфавитном порядке. Была обозначена система отношений внутри дескрипторов, условные обозначения которых указаны в скобках рядом с лексическими единицами. Строение тезауруса основано на иерархии по количеству лексических связей (анализ лингвистических пересечений) – от большего к меньшему.

ЗАКЛЮЧЕНИЕ

Конечный результат эксперимента по определению и организации интересов виртуальных социальных сообществ представлен в виде тезауруса. Достигнута цель установления иерархии между главными интересами посредством анализа лингвистических пересечений, организованная в виде схемы (рис. 6). Выделено 12 главных интересов – зонтов.

Основные выводы:

1. Признаки реальных и социальных сетей значительно отличаются, однако строятся на общих основных принципах; к признакам виртуального сообщества относятся преобладание письменной речи, анонимность, «раздвоение личности», территориальная и временная неограниченность, необязательное взаимодействие вне виртуальной среды, преобладание сетевых взаимоотношений над иерархией, ориентация на инновационность.

2. Язык виртуального сообщества специфичен: свобода от норм языка, территориальная и временная независимость, наличие хэштегов и мэншенов, наличие графических невербальных элементов – эмодзи, эмотиконы, стикеры, мемы.

3. Важным компонентом виртуального информационного пространства является фолксономия – позволяет упорядочивать не всю, но достаточно большую часть имеющейся и появляющейся информации.

4. Тезаурусный метод организации всех существующих интересов в виртуальном пространстве пока невозможен, но попытки создания такового выведут поиск информации на новый уровень.

5. Для поиска дескрипторов выделены следующие этапы: 1) выделение ключевых слов; 2) определение частоты выделенных ключевых слов; 3) выделение типичных предикатных конструкций (субъект, предикат, объект); 4) выделение предикатов для ключевых слов; 5) sentimentный анализ.

6. Выделены следующие лингвистические признаки интереса, установленные в ходе эксперимента: 1) в синтаксических отношениях интерес в основном выступает в роли объекта; 2) интерес частотен; 3) интерес выражается полными словами; 4) интерес выражается такими частями речи, как существительное, глагол, отглагольное существительное, а также глагольными словосочетаниями и именной группой; 5) интерес также определяется эмоциональными маркерами (sentiment analysis); в зависимости от типа контента в сообществе, некоторые признаки могут отсутствовать.

7. Одним из важных маркеров интересов является хэштег – ими отмечают значимые тематики информации постов; помогают определить направленность сообщества.

8. Определение доминирующего интереса сообщества затрудняется присутствием в нем нескольких интересов. Данная проблема решается преобладающими дескрипторами. Однако не всегда лексические единицы отображают доминирующий интерес, тогда важно учитывать концепции и цели сообщества, которые обозначены в описании сообщества.

9. Значимость дескрипторов можно определить на основе принципа их расположения (в одном из 5 текстовых пространств). Самым значимым выступают название и описание сообщества (при наличии такового), далее идут тексты постов, затем обсуждения, и, наконец, комментарии. Значимость пространств меняется при условиях разных комбинаций наличия текстовых пространств в сообществе.

10. Наименования являются важными дескрипторами интереса, так как несут в себе большую смысловую нагрузку; выделен следующий перечень наименований анализируемых сообществ:

- имена значимых личностей или персонажей;
- имена «админов» и значимых участников сообщества;

- наименования команд;
- наименования брендов;
- наименования значимых произведений – фильмов, сериалов, музыкальных и литературных произведений и др.;
- наименования мест сбора и мероприятий;
- наименования значимых организаций;
- наименование участников сообщества.

11. Многие русскоязычные сообщества носят английское название. Некоторые сообщества «развивают» иностранные интересы, ввиду чего используют вокабуляр иностранного языка, который далее может перейти в активный лексикон пользователя, затем превращается в слово, написанное транслитом, входит в естественный язык и теряет свои «корни» в сознании говорящих.

Проведенное исследование, разумеется, не является конечной точкой ввиду постоянного развития и модификаций виртуального пространства и коммуникации в его пределах.

СПИСОК ИСТОЧНИКОВ И ЛИТЕРАТУРЫ

1. Alicia Wolf. Expression Online: Gender Differences and Emoticon Use. CYBERPSYCHOLOGY & BEHAVIOR Volume 3, Number 5, 2000 Mary Ann Liebert, Inc.
2. Cortes C., Pregibon D., Volinsky C. Communities of interest // Proceedings of the 4th International Conference on Advances in Intelligent Data Analysis (IDA'01). Springer. 2001. С. 105–114.
3. France Henri, Béatrice Pudelko. Understanding and analysing activity and learning in virtual communities. Journal of Computer Assisted Learning, Wiley, 2003, 19, С. 474-487.
4. Hashtags search engine «Hashtagify.me». URL: <https://hashtagify.me/>
5. Hazra Imran, Aditi Sharan. Thesaurus and query expansion. International Journal of Computer science & Information Technology (IJCSIT), Vol 1, No 2, November 2009.
6. ISO 2788-1986. Documentation – Guidelines for the establishment and development of monolingual thesauri. Ed. 2.
7. Jeongin Kim, Dongjin Choi, Byeongkyu Ko, Eunji Lee, and Pankoo Kim. Extracting User Interests on Facebook. Research article. International Journal of Distributed Sensor Networks, Volume 2014, Article ID 146967, 5 с.
8. Mariam Adedoyin-Olowe, Mohamed Medhat Gaber and Frederic Stahl. A Survey of Data Mining Techniques for Social Network Analysis. School of Computing Science and Digital Media, Robert Gordon University. URL: <https://jdmdh.episciences.org/18/pdf>
9. Martínez-Cámara E., Martín-Valdivia M. T., Ureña-López L. A., Montejo-Ráeza A. Sentiment analysis in Twitter // Natural Language Engineering. Vol.20. Issue 1. January 2014. С. 1-28.

10. Michelson M., Macskassy S.A. Discovering users' topics of interest on twitter: a first Look. 2010, Toronto, Ontario, Canada. URL: <http://research.c.s.rutgers.edu/~sofmac/paper/and2010/michelson-and2010.pdf>.
11. Mikhalkova Elena, Ganzherli Nadezhda, Karyakin Yuri. A comparative analysis of social network pages by interests of their followers. Arxiv preprint. 2017. URL: <https://arxiv.org/abs/1707.05481v2>
12. Mikhalkova E., Karyakin Y., Glukhikh I. (2018) Large Scale Retrieval of Social Network Pages by Interests of Their Followers. In: Shi Y. et al. (eds) Computational Science – ICCS 2018. ICCS 2018. Lecture Notes in Computer Science, vol 10860. Springer, Cham.
13. Russell, Jon (2013-07-12). "Stickers: From Japanese Craze to Global Messaging Phenomenon" URL: <https://thenextweb.com/asia/2013/07/12/stickers/>
14. Teng C. Y. and Chen H. H. Detection of bloggers' interests: Using textual, temporal, and interactive features. In Proc. Of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, 2006.
15. Thomas Vander Wal. Folksonomy. URL: <http://www.vanderwal.net/folksonomy.html>
16. Srinivasan P. Thesaurus construction. University of Iowa. URL: <http://orion.lcg.ufrj.br/Dr.Dobbs/books/book5/chap09.htm>
17. Xin Li, Guo L., Zhao Y.I. Tag-based social interest discovery // Proceeding of the 17th international conference on World Wide Web. 2008. С. 675-684.
18. Величко В., Волошин П., Свитла С. Автоматизированное создание тезауруса терминов предметной области для локальных поисковых систем. XVth International Conference "Knowledge-Dialogue-Solution" KDS-2 2009, Kyiv, Ukraine, October, 2009.
19. Градюшко А. А. Потенциал социальных медиа в структуре инновационной коммуникации / Медиа- и межкультурная коммуникация в европейском контексте: материалы Международной научно-практической конференции. ФГАОУ ВПО «Северо-

- Кавказский федеральный университет», Институт социально-экономических и гуманитарных исследований Южного научного центра РАН, 2014. – С. 81–85.
20. Галямина Е. Ю. Лингвистический анализ хештегов Твиттера. Современный Русский язык в интернете / ред. Я. Э. Ахапкина, Е. В. Рахилина. — М.: Языки славянской культуры, 2014. С. 13-22.
21. ГОСТ 7.24-2007. Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый многоязычный. Состав, структура и основные требования к построению (введен в действие с 1 июля 2008 г.).
22. ГОСТ 7.25-2001. Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления (введен в действие с 1 июля 2002 г.).
23. Ермаков А. Е. Автоматизация онтологического инжиниринга в системах извлечения знаний из текста. Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2008, Бекасово, 2008.
24. Загорюлько Ю. А., Боровикова О. И. Построение многоязычного тезауруса предметной области средствами технологии создания порталов научных знаний. Знания-Онтологии-Теории (ЗОНТ-11).
25. Игошев Б. М., Рубина Л. Я. Гражданская активность студенческой молодежи: концептуальный словарь. Политическая лингвистика 4 (34), 2010. С. 41-49.
26. Интернет-энциклопедия «Википедия». URL: <http://ru.wikipedia.org>
27. Караулов, Ю. Н. Лингвистическое конструирование и тезаурус литературного языка [Текст] / Ю. Н. Караулов - Наука, Москва, 1981.- с.148-150.
28. Кондаков Н. И. Логический словарь-справочник / ред. Д. П. Горский. М.: Наука 1975. 721 с.

29. Коршунов А. В. Задачи и методы определения атрибутов пользователей социальных сетей. Труды Института системного программирования РАН. С. 439-445.
30. Коршунов А. В. Исследование структуры сообществ пользователей в графах онлайн-социальных сетей: [Диссертация]: Москва, 2015.
31. Лафтими И. Общая характеристика и критерии классификации словаря тезаурусного типа // Молодой ученый. — 2011. — №12. Т.1. — С. 252-255.
32. Лесников С. В. Тезаурус как отражение системности языка. Вестник Челябинского государственного университета. 2011. № 28 (243). Филология. Искусствоведение. Вып. 59. С. 52-61.
33. Лукашевич Н. В. Тезаурусы в задачах информационного поиска – М.: Издательство МГУ, 2011.
34. Луков В. А. Тезаурусы и сетевое общество. – М., 2016. Электронный архив НГУ. URL: <https://nsu.ru/xmlui/handle/nsu/8967>
35. Луков Вал. А., Луков Вл. А. (2008) Тезаурусы: Субъектная организация гуманитарного знания. М.: Изд-во Национального ин-та бизнеса.
36. Львов М. Р. Основы теории речи. – М., 2000.
37. Майорова А. М. Виртуальные знакомства как форма социальной коммуникации. [Автореферат дипломной работы]: Саратов, 2017.
38. Мельникова М. С., Яковлев И. П. Понятие «Социальная сеть» в социологических теориях и интернет-практиках. Вестник СПбГУ. Сер. 9. 2014. Вып. 1.
39. Михалькова Е. В. Система автоматического анализа принадлежности пользователей к социальной группе: [Магистерская диссертация]: Тюмень, 2015.
40. Михалькова Е. В., Кульчицкий Д. М., Уржумцев О. Д., Муромцев Д. И., Закомолдин И. С. Дискурс социально активных молодежных сообществ г. Тюмени (по данным на 2012 г.): Монография / под ред. Е. В. Михальковой. Тюмень: Титул, 2012. 112 с.

41. Морозова А. Тематическая классификация сообществ в социальной сети «ВКонтакте» как нового средства массовой информации. [Электронный ресурс]: URL: <http://elib.bsu.by/bitstream/123456789/171543/1/160-166.pdf>
42. Омельченко Е. Молодежные культуры и субкультуры. М.: Ин-т социологии РАН, 2000. 262 с.
43. Погорский Э. К. Инновационность молодежи как источник социокультурных изменений на этапе становления информационного общества: [Автореферат диссертации]: Москва, 2012.
44. Рогушина Ю. В. Разработка онтологической модели информационной потребности пользователя при семантическом поиске. Онтология проектирования 2 (12), 2014. С. 60-82.
45. Савкин Д. А. Особенности составления русско-испанского и испано-русского тезаурусов «Горный и пешеходный туризм»: [Дипломная работа]: Ростов-на-Дону, 2014.
46. Самойленко Л. В. Фатические средства в речи пользователей компьютерной сетью (на материале жанра чата): [Автореферат диссертации]: Астрахань, 2010.
47. Селезнев Р. С., Скрипак Е. И. Социальные сети как феномен информационного общества и специфика социальных связей в их среде. Вестник КемГУ 2013 № 2 (54) Т. 3.
48. Сидорова Е. А. Многоцелевая словарная подсистема извлечения предметной лексики // Труды международной конференции Диалог'2008 «Компьютерная лингвистика и интеллектуальные технологии». М.: РГГУ, 2008. Вып. 7 (14). С. 475-481.
49. Сомс Н. Л., Добров А. В., Доброва А. Е. Использование средств лингвистической обработки текстов в системе мониторинга информационных ресурсов по пользовательским предпочтениям. Материалы XVII Всероссийской объединенной конференции

- «Интернет и современное общество» IMS-2014, Санкт-Петербург, 2014.
50. Социальная сеть «ВКонтакте». URL: <https://vk.com>.
51. Степаненко А. А., Шиляев К. С., Резанова З. И. Атрибуция профессиональных интересов пользователей социальной сети «вконтакте» на основе текстов тематических групп и персональных страниц. Вестник Томского государственного университета. Филология. 2018. № 52.
52. Тезаурусный анализ мировой культуры: Сб. науч. трудов. Вып. 2 / Под общ. ред. Вл. А. Лукова. — М.: Изд-во Моск. гуманит. ун-та, 2005. 92 с.
53. Фролов С. С. Социология. Учебник. Для высших учебных заведений. М.: Наука, 1994. 256 с.

ТЕЗАУРУС ИНТЕРЕСОВ ПОЛЬЗОВАТЕЛЕЙ

УСЛОВНЫЕ ОБОЗНАЧЕНИЯ

SYN – синоним

FE – полный эквивалент

EM – эмоциональный маркер

SV – специфическая лексика

TH – тематический хэштег

neg – в рамках интереса слово
приобретает негативную
коннотацию

pos – в рамках интереса слово
приобретает позитивную
коннотацию

P – рынок

A – искусство

I – игры

J – животные

AKT – активизм

L – ЛГБТ+

SS – сэйфспейс

N – наука

E – еда

Z – здоровье

REL – религия

СОДЕРЖАНИЕ

1. ИСКУССТВО – стр. 73

- музыка

- литература

- изо

- кино

2. РЫНОК – стр. 76

- товары

- поиск работы

- услуги

- антикафе

- магазины

3. НАУКА – стр. 78

- природа

- психология

- иностранные языки

- история

4. АКТИВИЗМ – стр. 80

- защита людей

- защита природы

5. ЗДОРОВЬЕ – стр. 82

- правильное питание

- медицина

- ментальное здоровье

6. СЭЙФСПЕЙС – стр. 83

7. ИГРЫ – стр. 83

- видеоигры

- ролевые игры

- спорт

8. ЛГБТ+ - стр. 85

9. ЖИВОТНЫЕ – стр. 86

10. ЕДА – стр. 86

11. ВИРТУАЛЬНЫЕ

ЗНАКОМСТВА – стр. 87

12. РЕЛИГИЯ – стр. 88

ИСКУССТВО**МУЗЫКА****эмо панк**

melodic hardcore

post-hardcore

screamo

имо бандаимо панк (FE)

мастеринг (SV)

мидвест

микстейп (SV)

панк-хардкор

скримо

хардкор

эмо панк сцена

emo translations – перевод**песен emo (H)**емо

emo translation (FE)

emotional

вырывать сердечко (EM)

грустно (EM)

жиза (EM)

жизненно (EM)

имо

можно перевести как

название отсылается

отредактированный переводперевод

песня описывает

песня отсылается

помощь в перевод

разбивать сердечко (EM)

тексты песен

трудный подростокbedroom popmidwest emo

Ваня

вечеринка

готовиться к концерту

группа Трудный Подросток

имо группаимо трек

классные футболки

мерч

новый трек

тусовка

футболка трудный подросток

**Neologeek – виниловые
вечеринки (P)**

new wave

retro wave

synth party

vinyl party (FE)

билет

винил

виниловая пластинка

движ (SV)

для настоящих меломанов

жаркая ночь (EM)

крутить винил

на виниле

неологика (SV)

по-неологиковски (SV)

отрываться

пати (SYN)

проходка (SV)

синти-поп

танцевать

теплая вечеринка (EM)фотки**Indie Music – инди музыка**

Alternative

Alternative-pop

Electronic

Electropop

Experimental

Hip-Hop

Indie (FE)indie-pop

Industrial

Pop

Punk

Rock

Synth-Pop

дебютный альбом

музыкант

новый альбом

плейлист

**Инди Рок Indie Rock – инди-
рок музыка**Indie_Bio (TH)Indie_Playlist (TH)Indie_Photo (TH)альбомальтернативная группа

альтернативный рок

американская рок-группавыпускать альбомгруппа

инди-рок

инди-рок-группа

рок

рок-группа**Twenty One Pilots**

Cliqheart (SV, TH)

Jennaajoseph (TH)

Jordandun (TH)

Trench

Tylertjoseph (TH)

Брэд

Брэндон

группа twenty one pilots

Дженна

Джим

Джордан

Джош

интервью

перевод

музыка

студия

Тайлер

фото**повторяющиеся****лексические единицы:**

emo

indie

pop

альбом

вечеринка

группа

имо

перевод

фото

ЛИТЕРАТУРА**монологи Луны**

белый

грустить

звезда

Земля

Когда ты Луна

космос

Луна
люди
небо
облака
светиться
ярко светить

**абстрактные истории
штукатурки**

автор
дом
жизнь
зло
история
комент_ИШ (ТН)
краска
обои
пятно
рисовать
художник
штукатурка

**Подслушано - чужие
истории из жизни**

анонимно
воспоминание
вспоминать
детство
друг
жизненная ситуация
жизнь
люди
мама
муж
откровение
подруга
Подслушано_детство@overhe
ar (ТН)
Подслушано_провал@overhea
r (ТН)
работа
родители
свадьба
секрет

**самиздат Батенька, да вы
трансформер (Н)**

автор
архив
архивбатеньки (ТН)
батенькакурсмеламеда (ТН)

герой
журналист
интервью
исследование
история
писатель
подкаст
публиковать
рубрика
самиздат
студенты Британской высшей
школы дизайна
Теодор Глаголев
читатель

**помощь писателям и
сценаристам**

автор
акт
герой
глава
детектив
издательство
история
кульминация
нон-фикшн
персонаж
писатель
писать
повесть
полезный материал
произведение
рассказ
соавтор
сцена
сценарий
характеры
читатель

комикс Blacksad (И)
art@blacksad_public (ТН)
Blacksad
Blacksad: Under the Skin
Grandville
Блэксэд (FE)
Блэксэд (FE)
детектив
игра
комикс
кот
новый том

оригинал
перевод
персонаж
раскадровка
том
трейлер игры
фигурка
шестой том

**повторяющиеся
лексические единицы:**

автор
герой
история
персонаж
писатель
читатель

ИЗО

**Могилат – портреты
Петербургских улиц**

атмосфера
атмосферный двор
Виктор Могилат
город
двор
композиция
Петербург
природа
прогулка
улица
фото
фотограф
фотопаблик

**Bronze – визуальная
культура**

Harper's Bazaar
Life
Magnum
Vogue
Анатолий Егров
Виногранд
выставка
журнал
запечатлеть
изображение
искусство
кадр
мастер

мода
 портрет
 Ричард Аведон
фотограф
фотография
 Школа Академической
 Фотографии

**FTRM – современное
 фотоискусство**

Selected Works
 Sofia Savinova
атмосфера
 аутентичные пейзажи
 колоритный фотопроjekt
 красота
 пейзаж
 пейзажные фотографии
 снимок
 стрит-фотографии
 трэвел-фотографии
фотограф
фотография
 фотопроjekt

**Posterino – искусство
 плакатов**

Alexandro Valcarcel
 Anna Kulachek
 Ben Arfur
 Dima Shiryayev
 Everyday Practice Studio
 Javier Jaén
 Lasca Studio
 Posterino
 posters
 афиша
дизайнер
иллюстратор
плакат
 постер
 художник

Nordic Design

danishdesign (TH)
 finnishdesign (TH)
 ИКЕА
 madeinfinland (TH)
 nordicarchitecture (TH)
 nordicinterior (TH)
 swedishdesign (TH)

архитектор
 бренд
 Виркалла
дизайн
дизайнер
 здание
 интерьер
искусство
 мебель
 северный дизайн (FE)
 скандинавский стиль
 скандинавскийдизайн (TH)

Pollskill – графический

дизайн
 PollSkill.com
 айдентика
 брендинг
 графический дизайн
дизайн
дизайнер
иллюстратор
 иллюстрация
 логотип
плакат
 упаковка
 фирменный стиль

**your thoughts . – скриншоты
 из мультфильмов**

90style (TH)
 adventure_time (TH)
 animation (TH)
 art (TH)
 cartoons (TH)
 esthetics (TH)
 ladybugandcatnoir (TH)
 sponge_bob (TH)
 ladybug_and_catnoir (TH)
 your_thought (TH)
 spongebob (TH)
 yourthought (TH)

**monster boyfriend – арты
 монстро-мальчиков (И)**

Dungeons and Dragons
 mb_alien (TH)
 mb_amphibian (TH)
 mb_demon (TH)
 mb_furry (TH)
 mb_games (TH)

mb_good_boy (TH)
 Star vs. the Forces of Evil
 World of Warcraft
 инопланетяне
 ксенофилия
 мифические монстры
 выдуманные расы
 монстро-мальчик
 персонаж
 призраки
 роботы
 тератофилия
 фендом (SV)
 экзофилия

**повторяющиеся
 лексические единицы:**

атмосфера
 дизайн
 дизайнер
 иллюстратор
 искусство
 плакат
 фотограф
 фотография

КИНО

CLIQUE – поп-культура

кино
 дебютный трейлер фильма
 Дейнерис
 Игра Престолов
 Квентин Тарантино
 Малéfисента: Владычица
 тьмы
 новая серия
 новый Бэтмен
 новый сезон
показ фильма
 поп-культурное оружие
сезон
 серия
 сиквел
 Том Холланд
 трейлер
фильм
 фотосессия
 Черное зеркало
 Эмилия Кларк

**Beat Film Festival —
фестиваль документального
кино (P)**

A Dog Called Money
Beat Film Festival
Coming of Age
PJ Harvey
Берлинский
фейсконтрольщик
билет
Вудсток
гала-премьера
главный герой
документальный фильм
Дэвид Боуи
Каннский фестиваль
кино
кинокритик
кинотеатр
показ
режиссер
фестиваль
фильм

Твин Пикс | Twin Peaks

DavidLynch (TH)
PeggyLipton (TH)
TwinActors (TH)
TwinPeaks (TH)
агент Купер
Дэвид Линч
Красная Комната
Лора Палмер
Марк Фрост
Мисс Твин Пикс 2019
сериал
Твин Пикс
фильм
Хэнк Дженнингс
чертовски хороший кофе

**ИНОЕКИНО – прокат кино
(P)**

Аврора
Андрей Тарковский
билет
выход в прокат
главный герой
документальный фильм
Жертвоприношение
журнал Искусство Кино

Иноекино
искусство
кадр
Каннский кинофестиваль
кино
кинотеатр
кинофестиваль
музыка
режиссер
премьера
сеанс
фильм

**Испанские сериалы и
фильмы**

Blanca Suárez
blancasuarez (TH)
elinternado (TH)
Laguna Negra
lagunanegra (TH)
maricocas (TH)
актер
актриса
испанские актрисы
испанские сериалы
испанский кинематограф
на языке оригинала
сезон
сериал
фильм
Черная Лагуна

**повторяющиеся
лексические единицы:**

главный герой
Каннский
кино
кинотеатр
показ
режиссер
сезон
сериал
фильм

РЫНОК

ТОВАРЫ

**РадиоВидео.рф - рации и
аксессуары**

Baofeng
Kenwood
TurboSky
Vector
аккумулятор
Аргут
портативная радиостанция
приобретать
радиостанция
рация
связь

музыкальная барахолка (A)

в отличном состоянии
гитара
гитарный предусилитель
документы в комплекте
за символическую цену
комбик (SV)
обмен
отдавать в комплекте
пересыл
продам
скол
состояние устройства
торг
трещина
усилитель
чехол в подарок
электрогитара

косплей барахолка (A)

cosplaysecond_Москва (TH)
cosplaysecond_СПб (TH)
аниме
бант
кимоно
косплей
костюм
лейс (SV)
линзы
парик
пересыл (TH)
персонаж
платье

пошив
продажа (ТН)
размер
ушки

ключи Steam (И)

dropkeys.ru
Steam
давать ключи
ключи
ключи от админа
получать ключи
промокод
раздача ключей
Стим

ПОИСК РАБОТЫ

вакансия
график работы
заработная плата
зп
искать в команду
коллектив
навык
оплата
опыт работы
портфолио
работа
работать
рабочий день
резюме
тз
условия

УСЛУГИ

фотосессии в Красноярске (А)

альбом
девушка-модель
есть желающие
записываться на фотосессию
парк
парк Горького
Сады мечты
фото
фотография
фотопрогулка
фотосессия

аренда помещения Место

Действия
аренда
атмосфера
запись
календарь событий
мастер-класс
мероприятие
место действия
МК
ПАЛЬМА
площадка
стоимость
творческое пространство

камерный кинотеатр Out Cinema (А)

donation
Out Cinema
авторское кино
анимация
аниме
афиша
билет
визуальный код
вкусный чай
зал в аренду
кино
кинотеатр
лекция
на языке оригинала
режиссер
с русскими субтитрами
сеанс
скидка студентам
фильм

архитектурные экскурсии в Москве (А)

МГИ_календарь (ТН)
авангард
ар-деко
архитектор
архитектура
архитектурный
башня
билет
дворец
здание
история
крыша

мастер-класс
Москва
музей
прогулка
продолжительность
экскурсии
станция метро
экскурсия

дизайн логотипов и фирстилей Логомашина (А)

logomachine.ru
айдентика (SV)
дизайн
Логомашина
логотип
портфолио
посмотреть портфолио
презентация
продукт
проект
разрабатывать
логотип
разрабатывать айдентику
стиль
фирменный стиль

АНТИКАФЕ

антикафе Каледонский Лес (А) (РЕЛ)

donation
Бельтайн
волшебное пространство
волшебство
домик лесной ведьмы
каледонский
кинопоказ
колдунство
ламповый (ЕМ)
лекция
Лес на Садовой (SYN)
Лес на Чкаловской (SYN)
магия
мастер-класс
мероприятие
мультфильм
настольные игры
обитель городских фей
показывать в оригинале
чай

котокафе Республика**Кошек (Ж)**

catcafe (TH) (FE)

CatsRepublic (TH)

Выставка

добровольные пожертвования

замурчательный (EM) (SV)

запись

кот

котик

котокафе

кошка

мяу

на Литейном

на Якубовича

РеспубликаКошек (TH)

республиканцы (SV)

СанктПетербург

уководление (SV)

МАГАЗИНЫ**магазин одежды ASOS**

ASOS

бренд

брюки

джинсы

коллекция

костюм

кроссовки

модный

принт

рубашка

магазин винтажной одежды**Фрик Фрак**

freakfrak (TH) (FE)

freakfrakshop (TH)

oldschool (TH)

retro (TH)

vintage (TH)

бархатный

блузка

брюки

винтажный

винтажныймагазин (TH)

винтажныймаркет (TH)

гранж

одежда

принт

ретро

фрикфрак (TH)

книжный магазин**Подписные Издания (А)**

автор

бестселлер

биография

главный герой

жанр

журнал

издание

иллюстрации

интервью

книга

купить с доставкой

литература

новинка

писатель

Подписные Издания

сборник

читатель

повторяющиеся**лексические единицы:**

donation/пожертвования

билет

брюки

в оригинале

лекция

мастер-класс

мероприятие

портфолио

принт

чай

НАУКА**ПРИРОДА****The Batrachospermum****Magazine –****развлекательный научно-популярный журнал (Ж)**

Батрахоспермум (FE)

биолог

биология

биоразнообразие

ген

животное

зараза

зверь

исследование

муравей

насекомое

научная история

паук

позвоночные

природа

птица

слизь

статья

ученые

эволюция

Russian Travel Geek –**научно-популярные****экспедиции с молодыми учёными (Ж)**

RTG

RTG Camp

russiantravelgeek (TH)

scienceandtravel (TH)

Байкал

Берег медведей

вулкан

геология

гид

гляциологическая станция

животное

заповедник

Земля

Камчатка

ледник

лекция

маршрут

наука

научно-популярная
экспедиция
научный заказ
экспедиция
ПСИХОЛОГИЯ

**Когнитивный патимейкер –
когнитивистика и мемы**
выборка
Выготский
исследование
клиническая группа
когнитивист
когнитивная психология
Лурия
мем
наука
научник (SV)
опросник
продавать ковры
психоанализ
психолог
психология
психотерапия
студент
Торндайк
эксперимент

Саморазвитие (З)

внутренний мир
внутренний рост
воспитание детей
знание
зона комфорта
незаурядность
новая жизнь
ошибки родителей
позитив
психолог
психосоматика
совет
счастье
удовольствие
успех

повторяющиеся лексические единицы:

животное
исследование
наука
психолог

ИНОСТРАННЫЕ ЯЗЫКИ

**Я – переводчик –
информация для
переводчиков (Р)**
английский
вакансия переводчика
вариант
испанский
итальянский
как переводится
коллега переводчик
контекст
направление
перевод документов
переводчик
переводчик на выставку
письменный перевод
технический перевод
устный перевод

english yo – английский язык 18+

balls
language
yopra (SV) (TH)
американский
в английском
в значении
выражение
запоминать
значение
использовать выражение
на английском языке
объяснять
означать
онлайн-курс
перевод
переводить
сленг
язык
яйца

испанский язык в испанском языке

глагол
грамматика
запоминать
изучать испанский
испания
испанская раскладка

на испанском
разговорные выражения
сериал
слово
улучшить свой испанский
учебник
фразы на испанском
ЕГЭ по английскому
adjective
conditional mood
correct answer
grammar
practice
practice time
test
ЕГЭ
задание
письмо
повторять
подборка лексики
проверка задания
разбор задания
степени сравнения
прилагательных
устойчивые выражения
эссе

учебник New English File 3rd edition

additional materials
books
English File 3rd Edition
Advanced
English File 3rd Edition
Intermediate
English File 3rd Edition Pre-
intermediate
English File 3rd Edition Upper-
Intermediate
iChecker
iTutor
new archive package
student's Book
Teacher's Book
Tests
Workbook
ZIP
дополнительный материал
книги
новый архив

повторяющиеся лексические единицы:

английский
выражение
испанский
перевод
язык

ИСТОРИЯ

Древний мир: история и факты

Архимед
бог
богиня
гладиатор
гробница
доспехи
Древняя Греция
Египет
император
история
мифология
Рем
римляне
Ромул
саркофаг
Сократ
фараон
хеттский язык
царь
Эвклид

Мир Маньяков и Серийных

Убийц

женщина
животные
каннибал
кости
кровь
массовый убийца
насилие
преступление
причина смерти
психопат
расследование
серийный убийца
смерть
социопат
труп
убийство

убийца

История: пытки, казни, телесные истязания

виселица
гильотина
жертва
инквизиция
история
кровь
наказание
обезглавливание
орудие пытки
осужденный
палач
преступление
преступник
приговор
пытка
смертная казнь
убийство
четвертование
электрический стул

ЕГЭ ОГЭ | История

анализ
ЕГЭ
Задание
задания с документами
задание с картой
историческое сочинение
история
история в схемах
история в таблицах
История России
Орлов
подготовка к ЕГЭ
политика
Россия
Русь
справочник
экономика

повторяющиеся лексические единицы:

история
кровь
преступление
убийство

АКТИВИЗМ

ЗАЩИТА ЛЮДЕЙ

trans voice - поддержка для транс* и лгбт+ персон (Л)
information@trans_voice (ТН)
transguys@trans_voice (ТН)
гендер
гендерная идентичность
грудь
день видимости
дисфория
ЛГБТ+
мисгендерить (SV)
насилие
небинарные люди
переход
поддержка
поддержка@trans_voice (ТН)
помощь
транс*
трансгендерность
трансгендерные люди
трансфобия

Рёбра Евы – борьба с гендерной дискриминацией

Александра Олейник
Вирджиния Вулф
женщина
культура
лекция
Маргарита Мурзинова
миф о красоте
мужчина (neg)
насилие
пространство Ребра Евы
психологиня
Рёбра Евы
самооборона
сексуальность (neg)
стереотип
феминизм
феминистка
Харриет Филлипс
чаепитие

женская власть - феминизм

гендерный стереотип
декриминализация
женщина

мужик (neg)
 мужчина (neg)
насилие
патриархат (neg)
 поддерживать феминизм
 профем
сексизм
 сексист
 сексуальное насилие
 суицид
 фемвертайзинг (SV)
феминизм
феминистка
 феминитив

Чашечку феминизма и печеньку, пожалуйста –

феминизм и котки
 odonata@od_fem (ТН)
 psy@od_fem (ТН)
 абьюз
 безопасность
борьба
борьба с патриархатом
гендер
 гендерные стереотипы
 домашнее насилие
женщина
 жертва
 защищать себя
насилие
 отношения
сексизм
 травма
 утренние_киски (ТН)
феминизм

Soc-Fem –

социалистический феминизм, 18+ (Л)
 Аборт (pos)
 беременность
борьба
 вазэктомия
 гомосексуал
 гомофобия
 девочка
 девушка
 дискриминация
женщина
защита от притеснений

инцест
насилие
 Нэнси Роман
 права женщин
 ПравоНаАборт (ТН)
 РепродуктивныеПрава (ТН)
 стерилизация

ЗАЩИТА ЖИВОТНЫХ

Greenpeace России (Ж)

Greenpeace
 активист
 Арктика
 белуха
 беречь природу
борьба
 волонтер
 Гринпис (FE)
 дикая природа
 забастовка
 загрязнение
защита лесов
 ЗащитимЛес (ТН)
 китовая тюрьма
 лесные пожарные
 нерестовые водоемы
 нивкакиесвалки (ТН)
 освободить животных
 остановиогонь (ТН)
 СпасемЛосося (ТН)

коты в добрые руки (Ж)

в дар
 вдар (ТН)
 вдобрыеруки (ТН)
 ветпаспорт
 девочка
животное
 искать дом
 кастрирован
 кот
 котенок
 котик
 кошечка
 кошка
 куратор
 малыш
 мальчик
 отдавать
 передержка (SV)

повторяющиеся

лексические единицы:

борьба
 гендер
 женщина
 животное
 защита
 мужчина
 насилие
 патриархат
 сексизм
 стереотип
 феминизм
 феминистка

ЗДОРОВЬЕ**ПРАВИЛЬНОЕ ПИТАНИЕ****СМТ - Научный подход (Е)
(Н)**

аутофагия
БАД
белок
Борис Цацулин
вес
ГМО
голодание
гормон
диета
добавка
жир
здоровье
исследование
лекарство
питание
продукт
физиология
фитнес
ЭКГ

**Сыроедение - Я
ФРУКТОРИАНЕЦ –
здоровый образ жизни (Е)
(А)**

банан
голодание
животные
здоровое питание
здоровье
йога
моно-фруктоед
орех
очищение души и тела
смузи
сыроедение
сыроедний
сыроежка (SV)
сырой
убийство животных
ферментированный чай
фрукт

МЕДИЦИНА**ДНК – здоровье и медицина
(Е) (Н)**

депрессия
здоровье
ЗОЖ
медицина
мозг
наука
питание
питание@dnahealth (TH)
полезный перекус
психика
психология
рак
расстройство
смерть
факты@dnahealth (TH)
физиология

МЕНТАЛЬНОЕ ЗДОРОВЬЕ**Face Your Disorder**

бороться с расстройством
врач
депрессия
консультация
лгбт-френдли
мозг
отклонение
паника
пациент
прокрастинация
психика
психическое расстройство
психология
СДВГ
состояние
страх
стресс
терапия
тревога
эмоция

**pretty eerie – ментальное
здоровье и феминизм (А)**

dirai@peerie (TH)
антидепрессант
бодипозитив
бупропион
депрессия
загон

зож
лекарство
лечиться
менталка (SV)
наркотики
побочка (SV)
психиатр
психотерапия
расстройство
рпп
триггер (SV)
феминизм
феминистка

**Новая жизнь – позитивное
мышление**

бороться со стрессом
быть счастливым
внутренний мир
дофамин
здоровый образ жизни
личностный рост
мозг
монах
наслаждаться жизнью
неприятные эмоции
позитивная энергия
просветление
радость
счастливый человек
счастье

повторяющиеся**лексические единицы:**

голодание
депрессия
здоровье
зож/здоровый образ жизни
мозг
питание
психология
расстройство
стресс
эмоция

СЭЙФСЕЙС

**PLAY LIKE A GIRL –
безопасное женское
развлекательно-игровое
пространство про-фем
направленности (И) (А)
(АКТ)**
art@playlikeagirl (ТН)
DragonAge@playlikeagirl (ТН)
kassandra (ТН)
Life is Strange
overwatch@playlikeagirl (ТН)
админка
геймерская культура
геймерша
героиня
женское пространство
марвел
мейл-фри (SV)
мизогиния
паладинка
персонажка
сексизм
феминитивы

**lovely days – уютное
виртуальное пространство
(А)**
аниме
вп (SV)
душевная группа
манга
милые посты
музыка
сериалы
твиттер
уютные посиделки
фильм

**stop hurting yourself –
поддержка людей, которые
занимаются селф-хармом (З)**
stop_hurting_youself (ТН)
tw (SV)
бить
боль
больно (ЕМ)
делать больно
истерика
кровь

ЛГБТКИА+
порез
практика селф-харма
пси-помощь (SV)
резать
романтизация селф-харма
(neg)
саморазрушение (SYN)
селф-харм
селфхарм (FE)
суицидальные мысли
физическое насилие
шрамы

ИГРЫ

игра
играть
игрок
команда
матч

ВИДЕОИГРЫ

**Counter Strike
CHAMPIONSCLUB**
awp (SV)
Steam
админ
администрация
банить
игра
играть
игрок
карта
моделька
ник
сервак (SV)
сервер
сервера CHAMPIONSCLUB
скрин
стим

**игровой журнал Страна Игр
(А)**
видеоигра
Егор Просвирин
журнал
игра
играть
колонка
команда СИ
Константин Говорун
переиздание
подкаст
проходить игру
рецензия
сайт Gameland.ru
СИ_архив (ТН)
статья
Страна игр
файтинг (SV)

Blizzard Entertainment
Blizzard
Hearthstone

Overwatch
World of Warcraft
Альянс (SV)
вселенные Blizzard
игра
Король-лич (SV)
Орда (SV)
орк
офис Blizzard
статуя
фигурка

новости PlayStation

PS4 (TH)
Days Gone
DLC
Dreams
God of War
PlayStation
PlayStation 4
PS
Sony
геймплей (SV)
игра
разработчик
релиз игры
ремейк
сюжетка (SV)
трейлер игры

игровые новости Hypewave

EGS (TH)
Epic Games
esports@hypewave (TH)
PC (TH)
Red Dead Redemption 2
Rockstar Games
Steam (TH)
Ubisoft
анонс
геймплей (SV)
игра
игрок
разработчик
релиз
тизерить (SV)
эксклюзивность

FIFA19

FIFA (TH)
FUT19 (TH)

болеть за
гол
играть
игрок
команда
легенда
лига
матч
Месси
спортивная видеоигра
турнир
фифа (FE)
футбол
чемпион

повторяющиеся лексические единицы:

Steam
игра
игрок
играть
видеоигра
релиз
геймплей
разработчик

РОЛЕВЫЕ ИГРЫ

ролевая по мультфильму

Аватар
ATLA (TH)
Legend_of_Aang (TH)
Legend_of_Korra (TH)
Аанг
аватар
Азула
анимационный сериал
Вода
Воздух
Земля
Корра
кроссовер
ЛоА
ЛоК
маг
магия
мультфильм
Огонь
ролевая игра
сериал

Коты-Воители | Племя Темных Ущелий – ролевая (Ж)

Архимед
Валун
воитель
глашатай
двуногие
игра
кот
котенок
коты воители
Мокроноса
Племя Темных Ущелий
ролевая
старейшина
хвост
целитель
чудище
шерсть

повторяющиеся лексические единицы:

игра
ролевая

СПОРТ

NHL
StanleyCup (TH)
Бишоп
Блюз
Бостон
Брюинз
гол
Даллас
забивать гол
игра
игрок
Каролина
конференция
матч
обыгрывать
плей офф (SV)
победа
серия
финал
шайба
Шаркс

NBA

NBA
 NBA News (TH)
 NBA Playoffs (TH)
 Raptors (TH)
 Stephen Curry (TH)
 блок-шот (SV)
 Голден Стэйт
драфт (SV)
 защитник
игра
играть
игрок
команда
конференция
 Лейкерс
матч
НБА (FE)
плей-офф
 разыгрывающий
 трехочковый

New-York Knicks

ESPN
 stats@knicksnation (TH)
 Бэррет
драфт (SV)
 Дэвис
 Зайон
игра
играть
игрок
команда
конференция
матч
НБА
 Никс (FE)
 Нью Йорк
 Нью-Йорк Никс (FE)
 Пеликанс
победа
 сезон
 Уильямсон
 форвард (SV)

повторяющиеся

лексические единицы:

драфт
 игра
 играть
 игрок
 команда

конференция
 матч
 НБА
 плей-офф
 победа

ЛГБТ+

**дочери десятой музы –
 женщины, которые любят
 женщин, 18+ (А) (АКТ)**

eve
 lesbian
 sappho
 афродита
 вирджиния вульф
 героиня
 джентльмен джек
 ева
 жена
 женщина
 лесбиянка
 любить женщин
 любовь
 сапфо
 убивая еву
 феминизм
 энн листер

ЖИВОТНЫЕ

кот
котик
питомец
порода
хозяин

МЫСЛИ КОТОВ

британец
жмяк
КОТ
КОТИК
кошачий список дел
кошачья тусовка
лапка
мейн кун
ПИТОМЕЦ
помогать человеку
ПОРОДА
свой человек
ХОЗЯИН
царь
человек
шотландский

ПИТОМЦЫ

животное
КОТ
КОТИК
кошка
пес
песик
ПИТОМЕЦ
питомцы рассказывают (ТН)
ПОРОДА
собака
совет
уход за питомцами
ХОЗЯИН

ЕДА

авокадо
зелень
лук
масло
молоко
овоши

вегетарианская еда (З)

авокадо
белок
веганство@just_veg (ТН)
зелень
йогурт
капуста
лук
овоши
орех
полезно_знать@just_veg (ТН)
растительное молоко
рис
смузи
соевый соус
сыроедение
сыроедить
тофу
фасоль
яблоко
яйцо

веганские рецепты

govegan (ТН)
veganfood (ТН)
авокадо
веганский рецепт (ТН)
имбирь
капуста
картофель
кокосовая мука
кокосовое молоко
лук
нут
овоши
оливковое масло
орех
петрушка
растительное масло
семена
хумус

диетические рецепты (З)

авокадо
Б/Ж/У
БЖУ
диета
диетический рецепт
зелень
калории
КБЖУ
кефир
курица
лук
мед
натуральный йогурт
овоши
оливковое масло
перец
похудение
ппголовногомозга (ТН)
правильное питание
яйцо

Cook Good - лучшие рецепты

блинчики
голубцы
зелень
какао
капуста
картофель
картошка
крем
курица
кусочки
листья салата
масло
молоко
мука
мясо
начинка
перец
посолить
разрыхлитель
яйцо

повторяющиеся**лексические единицы:**

авокадо
йогурт
капуста
картофель

курица
масло
молоко
овощи
орех
перец
рецепт
яйцо
яйцо

ВИРТУАЛЬНЫЕ ЗНАКОМСТВА

анкета
возраст
гулять
друг
знакомство
общение
отношения

молодежные знакомства

story_datings

sd_short (ТН)
sd_живувсети (ТН)
sd_Москва (ТН)
sd_неважно (ТН)

анкета
возраст
гулять
друг
знакомиться
знакомство
искать друзей
искать отношений
компания
общение
отношения
собеседник

молодежные знакомства

Existential datings

eМосква (ТН)
анкета
возраст
встреча
гулять
друг
eИнтернет (ТН)
eПетербург (ТН)
знакомство
интерес
общение
отношения
прогулка
собеседник
увлечение

Знакомства СПб

без вредных привычек
возраст
гулять

девушка
дети
знакомиться
знакомство
мужчина
общаться
общение
отношения
поиск друзей
рост
семья
серьезные отношения
создание отношений

ЛГБТ знакомства 16+ (Л)

актив (SV)
анкета
беседа
внешность
возраст
гулять
девушка
друг
дружба
знакомство
ЛГБТ
ЛГБТ_Анонимно (ТН)
ЛГБТ_Би (ТН)
ЛГБТ_Лесби (ТН)
любовь
общаться
общение
отношения
парень
романтика

повторяющиеся

лексические единицы:

анкета
возраст
гулять
девушка
друг
знакомиться
знакомство
общаться
общение
отношения

РЕЛИГИЯ

гадание
 расклад
 энергия

Дикое Неизвестное Таро

гадание
 Дикое Неизвестное Таро
 жезлы
 Императрица
 карта
 кверент (SV)
 колода
 консультация
 мечи
 младший аркан
 пентакли
расклад
 старший аркан
 таро
 таролог
 туз
 чаши
 Шут
 эзотерика
энергия

Мир Рун

боги
гадание
 карма
 Кеназ
 лунный день
 МирРун_магия (ТН)
 МирРун_Руны (ТН)
 перевернутая руна
расклад
 ритуал
 руна
 суеверие
 Тор
 Турисаз
 Уруз
 энергетика
энергия

ПРИЛОЖЕНИЕ 2 – АНАЛИЗ ЛИНГВИСТИЧЕСКИХ ПЕРЕСЕЧЕНИЙ

	искусство	рынок	активизм	игры	животные	лгбт+	сэйфспейс	наука	еда	здоровье	религия
перевод песен емо (A)								translation перевод перевести			
виниловые вечеринки (A)		билет проходка									
самиздат Батенька, да вы трансформер (A)								исследование студенты			
комикс Blacksad (A)				игра трейлер игры							
monster boyfriend – арты монстро-мальчиков (A)				Dungeons and Dragons World of Warcraft games							
Beat Film Festival – фестиваль документального кино (A)		билет фестиваль показ									
ИНОЕКИНО – прокат кино (A)		Аврора билет сеанс выход в прокат									
музыкальная барахолка (P)	гитара электрогитара										
косплей барахолка (P)	аниме косплей										
ключи Steam (P)	steam/стим ключи										
фотосессии в Красноярске (P)	фото фотосессия фотопрогулка фотография										
камерный кинотеатр Out Cinema (P)	donation афиша билет зал в аренду сеанс										
архитектурные экскурсии в Москве (P)	архитектура архитектурный архитектор										
дизайн логотипов и фирстилей Логомашина (P)	айдентика дизайн логотип стиль										
антикафе Каледонский Лес (P)	кинопоказ мультфильм										волшебство колдунство магия Бельтайн
котакафе Республика Кошек (P)					кот кошка котик мяу						
книжный магазин Подписные Издания (P)	автор биография главный герой литература										

ПРИЛОЖЕНИЕ 2 – АНАЛИЗ ЛИНГВИСТИЧЕСКИХ ПЕРЕСЕЧЕНИЙ

The Batrachospermum Magazine – развлекательный научно-популярный журнал (Н)					животное зверь муравей птица					
Russian Travel Geek – научно-популярные экспедиции с молодыми учёными (Н)					берег медведей животное заповедник					
Саморазвитие (Н)									зона комфорта психолог психосоматика	
Я – переводчик – информация для переводчиков (Н)		вакансия переводчика переводчик на выставку								
игровой журнал Страна Игр (И)	журнал колонка рецензия статья									
ролевая по мультфильму Аватар (И)	анимационный сериал сериал мультфильм									
Коты-Воители Племя Темных Ущелий – ролевая (И)					кот котенок хвост шерсть					
ЛГБТ знакомства 16+ (ВЗ)						ЛГБТ ЛГБТ_Анонимно ЛГБТ_Би ЛГБТ_Лесби				
СМТ - Научный подход (З)								БАД белок питание диета продукт	исследование физиология ГМО	
Сыроедение - Я ФРУКТОРИАНЕЦ – здоровый образ жизни (З)			животные убийство животных					банан здоровое питание орех смузи		
ДНК – здоровье и медицина (З)							наука психика психология физиология			
pretty eerie – ментальное здоровье и феминизм (З)			бодипозитив феминистка феминизм							
вегетарианская еда (Е)									полезно_знать сыроедение сыроедить	
диетические рецепты (Е)									ппголовногомозга правильное питание диета	

ПРИЛОЖЕНИЕ 2 – АНАЛИЗ ЛИНГВИСТИЧЕСКИХ ПЕРЕСЕЧЕНИЙ

дочери десятой музы – женщины, которые любят женщин, 18+ (Л)	sappho сапфо вирджиния вулф джентельмен джек		феминизм						
trans voice - поддержка для транс* и лгбт+ персон (АКТ)						ЛГБТ+ небинарные люди транс трансгендерность			
Soc-Fem – социалистический феминизм, 18+ (АКТ)						гомосексуал гомофобия			
Greenpeace России (АКТ)					белуха китовая тюрьма животные				
коты в добрые руки (АКТ)					кот котенок кошечка котик				
PLAY LIKE A GIRL – безопасное женское развлекательно-игровое пространство про-фем направленности (SS)	сериалы марвел art		сексизм мизогиния феминитивы	Life is Strange геймерская культура геймерша					
lovely days – уютное виртуальное пространство (SS)	сериалы музыка аниме фильмы манга								
stop hurting yourself – поддержка людей, которые занимаются селф-хармом (SS)									истерика пси-помощь саморазрушение, селф-харм/селфхарм суицидальные мысли