



УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

ЛЕКТА:
профессиональный
КОНТЕНТ-анализ за ОДИН
ДЕНЬ

Алексей Рюмин

Оглавление

ЛЕКТА:	1
профессиональный контент-анализ за один день.....	1
ЛЕКТА — качественно-количественный контент-анализ. Уникальный маркетинговый инструмент для российского рынка.	3
Что же такое ЛЕКТА?.....	4
Программа ЛЕКТА выполняет следующий ряд функций:.....	4
Глава первая. Сбор массива текстов и разбивка на фрагменты.....	6
Глава вторая. Словарь.....	13
Глава третья. Исходная матрица контент-анализа.....	21
Глава четвёртая. Факторный анализ в Statistica.	24
Глава пятая. Обработка матриц факторного анализа в Excel.	33
Часть шестая. Интерпретация матриц факторного анализа и написание отчёта.	39
Структура будущей аналитической записки.....	39
Аналитическая записка по результатам контент-анализа.....	42

ЛЕКТА — качественно-количественный контент-анализ. Уникальный маркетинговый инструмент для российского рынка.

Здравствуйтесь!

С удовольствием представляю вам свою первую книгу о практике контент-анализа, о программе ЛЕКТА — мощном инструменте маркетингового и социологического анализа. Мне посчастливилось в течении 3-х лет активно использовать этот продукт для реализации маркетинговых и социологических исследований методом многомерного контент-анализа, а также в течении 2-х лет преподавать этот курс студентам социологам и социальным работникам в Нижегородском Университете имени Лобачевского. Буду очень рад поделиться опытом использования этой замечательной программы, разработанной на кафедре прикладной социологии в ННГУ!



Не хочу отпугнуть вас, уважаемый читатель, однако несколько предостерегу от преждевременных надежд на получения универсального и идеального инструмента контент-анализа, лёгкого в освоении, самостоятельно решающего абсолютно все исследовательские задачи и т.д. и т.п. Этот инструмент действительно очень хорош, эффективен, удобен, вместе с тем, его использование потребует от Вас усилий в освоении его функционала, времени и просто желания обучаться и развиваться в малопопулярной сегодня в России сфере контент-анализа. Понять, для чего конкретно необходима работа над каждым из этапов замера Вы, как показывает практика, сможете только по завершении всего исследования, вплоть до написания последнего слова в аналитической записке по результатам всей работы. Если Вы готовы пройти этот путь — вперёд. Я же постараюсь объяснить работу в ЛЕКТА простым языком, останавливаясь на каждой из значимых деталей. Имея неплохой опыт обучения работе в программе, мне

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

хорошо известно, какие моменты могут вызвать основные трудности в работе с инструментом — на них буду останавливаться подробно. Эта книга собрана из публикаций на моём блоге, то есть с данными материалами вы легко можете ознакомиться на блоге о контент-анализе. Ссылка внизу каждой страницы книги. На блоге вы сможете ознакомиться с комментариями читателей, задать вопрос в том же формате, написать мне через форму обратной связи, если у вас есть какие-либо предложения по поводу сотрудничества или предложения по блогу и т.д., ознакомиться с другими материалами на блоге. Пишу я с ошибками и стиль мой не из лучших :-). В этой связи прошу вас обращать внимание на содержание, а не на форму изложения. Я практик, а не писатель всё же...

Итак, поехали!

Что же такое ЛЕКТА?

Лекта — лексико-семантический текстовый анализатор. Предназначен для проведения количественного и качественного контент-анализа текста. Для работы может быть использован, главным образом, маркетологами, социологами практиками, PR-специалистами, специалистами по рекламе, политтехнологами и представителями многих других профессий, нуждающихся в получении информации и продукте, услуге, публичной персоне, социальном явлении, процессе с помощью анализа имеющейся в распоряжении или потенциально доступной вербальной информации, выраженной в текстовом формате.

Ключевой принцип работы программы заключается в определении доминирующих в информационном пространстве изучаемого предмета микротем, после чего следует качественная интерпретация составляющих идентифицированных тематических блоков. Такая идентификация блоков может быть легко и быстро произведена независимо от объёма текстового массива. Хотя, стоит оговориться, что при работе с массивами на 50 000 — 200 000 страниц текста вам придётся запастись терпением. Можно сходить погулять на пол дня или поспать. Компьютер в этот момент лучше не трогать. Ощущение такое, что он вот вот задымится. Происходит это при построении исходных матриц контент-анализа. Но о них позже.

Программа ЛЕКТА выполняет следующий ряд функций:

1. Разбиение текстов на фрагменты с опорой на число слов либо на вспомогательные символы, вставленные в текст, сигнализирующие о необходимости разбивки текста в определённых местах. По умолчанию это символы ##SL
2. Составление словаря контент-анализа, опирающегося на принцип частотности либо на построенный категориальный аппарат. Включение всех словоформ выбранных лексем в словарь.
3. Составление матрицы контент-анализа, отражающей встречаемость каждой из лексем в каждом из фрагментов текста
4. Проведение факторного анализа, позволяющего сгруппировать единицы счёта и

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

фрагменты текста в тематические группы.

5. Экспорт матрицы контент-анализа в dat файл, с которым можно работать как в Excel, так и выгрузить, например, в Statistica.

Вот такие основные функции выполняет эта программа. Также для её освоения рекомендую Вам ознакомиться со статьями о качественном контент-анализе, располагающимися на блоге.

Официальным сайтом программы является сайт научно-исследовательского социологического центра НИСОЦ. По поводу приобретения программы можете обращаться по контактам, указанным на сайте [НИСОЦ](#).

Теперь давайте начнём непосредственное знакомство с программой — научимся разбивать тексты на отдельные части, используя функционал ЛЕКТА.

Глава первая. Сбор массива текстов и разбивка на фрагменты

Сейчас мы с вами поговорим о двух аспектах исследования, проводимых методом контент-анализа:

1. О наборе текстового массива
2. О загрузке и первоначальной механической обработке массива в программе ЛЕКТА

Хочу сразу сказать — здесь я хотел бы резюмировать свой опыт работы с программой ЛЕКТА. Я обладаю неплохим опытом работы в ней (простите, если повторяюсь). Это более 30 замеров, проведённых лично мной и руководство более чем 100 студенческими исследованиями. За это время мне удалось достаточно хорошо узнать, КАК ИМЕННО СТОИТ РАБОТАТЬ с программой ЛЕКТА. Продукт это уникальный, и он позволяет провести анализ текстового массива ОТ НАЧАЛА И ДО КОНЦА, при условии наличия самого текстового контента и адекватно работающего МОЗГА 😊 Ещё замечу — для России этот продукт уникален, и без всяких сомнений, достоин вашего времени для его освоения.

Итак, «поле».

Для сбора текстового массива для проведения контент-анализа в программе ЛЕКТА (или в ином пакете) используются разные подходы и они сильно варьируются в зависимости от Вашего

- бюджета
- задач
- исследовательских предпочтений

Причины для вариаций могут быть и иными. В нашем конкретном примере мы с вами будем анализировать глубинные интервью. Приводить здесь гайд (план интервью, включающий набор тем и подтем для разговора с респондентом), я не стану. О составлении гайда и о том, чем он собственно является, я напишу на блоге. В данном случае, порывшись в недрах своего компьютера, я нашёл именно этот массив. Почему бы не поработать именно с ним? 😊 Однако, сразу оговорюсь — глубинные интервью — бесконечно удобная и информативная работа. Сами интервью берутся на основе гайда — чёткой структуры, что в свою очередь помогает нам при работе в ЛЕКТА получить отдельные сегменты информационного пространства отношения к услуге, продукту, социальному явлению, процессу, политической персоне и т.д. и т.п. Вместе с тем, Вы можете работать и с другими массивами — главное понимать алгоритм работы с программой. Снова подчеркну — понять Вы его сможете полностью лишь после проведения исследования от начал и до конца. Так всегда было в рамках моей педагогической работы. Надеюсь, серия этих статей, поможет Вам разобраться с алгоритмом сразу же после прочтения. Причин так не думать не вижу — читайте всю серию и разберётесь, после чего приступайте к работе над своим собственным замером. Будут вопросы — пишите — с удовольствием отвечу. Чтобы было понятнее в перспективе хочу записать видеоуроки и выложить их на блоге, но пока эта работа ещё только задумывается.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

Учитывая то, что не у всех есть в распоряжении набор глубинных интервью (10-20 суммарным объёмом 50-100 страниц будет уже достаточно для проведения контент-анализа в ЛЕКТА), я хотел бы вкратце остановиться на подборе иных массивов текста и подготовке их к работе. К слову, тексты могут быть на русском, английском, немецком языках (с другими массивами, увы, не работал).

1. **Публикации в СМИ.** Наиболее важными аспектами при работе с этим видом контента являются единство жанра и единство объёма. Если Вы набираете публикации из СМИ в электронном формате, не смешивайте информационные заметки на 5-7 строк и развёрнутые публикации на 10-15 страниц текста. Также не совмещайте в массиве текстов статьи из СМИ и книги. Объём публикаций должен быть приблизительно равен. Также не надо соединять в едином массиве статьи юмористического, художественного, научного, юридического или иного жанра. Выберете уже что-то одно 😊 По опыту скажу — в ЛЕКТА лучше всего работать с публицистикой — СМИ, по тематике конгруэнтными массовой культуре, содержащими и эмоциональную и рациональную составляющую.

Для загрузки, например, статей, подойдёт поисковик типа google, или же поиск по интернет изданию. Однако, есть и более удобные, хотя и платные сервисы, типа EASTVIEW, с помощью которого Вы можете скачивать одновременно по 20 статей на нужную тему. Так за 2-3 часа работы можно набрать контента на 1000-1500 статей на интересующую Вас тему + собрать удобный реестр материала в виде таблицы, что позволит отдельно поработать и с ним. Есть и аналоги — гугл Вам в помощь 😊

2. **Фокус группы.** Тут всё так же как и интервью. Идеальный вариант для контент-анализа в ЛЕКТА

3. **Интернет-переписка.** Опять же, лучше её строить по заранее продуманному гайду и, опять же, это прекрасный массив для работы.

4. **Интернет-форумы.** Здесь придётся основательно поработать. Прежде всего вам надо скачать массив текста с форума. Если объём значителен, то вручную копировать текст общения на форуме нецелесообразно. Лучше воспользоваться специализированным софтом. Это, например, программа Teleport. Далее вам потребуется конвертировать HTML файлы в TXT, например, в программе HTMLasTEXT и потом, объединить файлы для дальнейшей обработки (удаления сохранившихся html тегов, рекламы и т.д.), например, в программе Advanced File Joiner. Работу с ними в будущем опишу на блоге. Аспектов тут масса, в частности разбиение текстов на элементы, соответствующие самим сообщениям на форуме (для этого можно использовать функцию разбиения на фрагменты ##SL в ЛЕКТА). Если не хотите с ними разбираться — в форме обратной связи на блоге предлагайте варианты сотрудничества 😊

5. **Блоги.** При внушительном объёме блога (блогов) аналогично предыдущему пункту.

6. **Нормативно правовые акты** — очень удобный вариант работы, учитывая прозрачность семантик лексем, однако, даёт сравнительно мало материала для работы.

7. **Художественная литература.** Не пробовал: экспериментируйте да обрящите 😊

Идём дальше.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

На руках мы имеем несколько десятков (сотен, тысяч, десятков тысяч, сотен тысяч и т.д.) статей. Для того, чтобы нам было удобнее работать в дальнейшем нам было бы неплохо составить реестр всех текстовых документов (просто перечень материалов в формате таблицы с указанием некоторых данных о них — дата публикации, наименование СМИ. Либо пола респондента, его возраста. Либо адреса сайта, с которого взята статья, её автора (блоггера, например) и т.д.). Вот так такой реестр может выглядеть:

№	Код файла	интервьюер	респондент	пол респондента	возраст респондент	профессия респондента	Комментарии
1	Ivanov_Evstigneev_m_20_geolog	Иванов	Евстигнеев	м	20	геолог	
2	Petrov_Morozov_m_25_sociolog	Петров	Морозов	м	25	социолог	
3	Sidirov_Smirnova_Z_45_marketolog	Сидоров	Смирнова	ж	45	маркетолог	
4	Ivanov_Zaycev_m_62_poliseman	Иванов	Зайцев	м	62	полицейский	
5	Petrov_Kuznecov_m_17_geolog	Петров	Кузнецов	м	17	геолог	
6	Sidorov_Shtolc_m_23_sociolog	Сидоров	Штольц	м	23	социолог	
7	Ivanov_Ribakova_Z_35_marketolog	Иванов	Рыбакова	ж	35	маркетолог	
8	Petrov_Shompolov_m_34_policeyskiy	Петров	Шомполов	м	34	полицейский	
9	Sidorov_Rrilov_m_19_biznesman	Сидоров	Крылов	м	19	бизнесмен	
10	Ivanov_Yashina_z_40_pozarniy	Иванов	Яшина	ж	40	пожарный	

Соответственно, при работе с другими массивами текста, параметры для реестра могут отличаться. Подчеркну самое важное — чем больше их будет — тем лучше.

Как Вы видите, среди прочих параметров здесь присутствует параметр КОД — здесь закодированы значения всех параметров, присутствующих в таблице. Этим кодом и будут называться наши документы: каждый текстовый документ (а мы их будем сохранять именно в формате txt, нам надо будет назвать этим кодом. То есть Вы открываете блокнот — вставляете в него текст документа и сохраняете его под этим названием (код=название документа). Если документов больше 1000, например, то работа становится весьма кропотливой. Это возможно, если вы, например, анализируете статьи из СМИ и скачали их ОЧЕНЬ МНОГО. Тут для создания кодов Вам на помощь придут формулы EXCEL (например, =СЦЕПИТЬ и =ЗАМЕНИТЬ — для объединения значения параметров в ячейках и удаления лишнего).

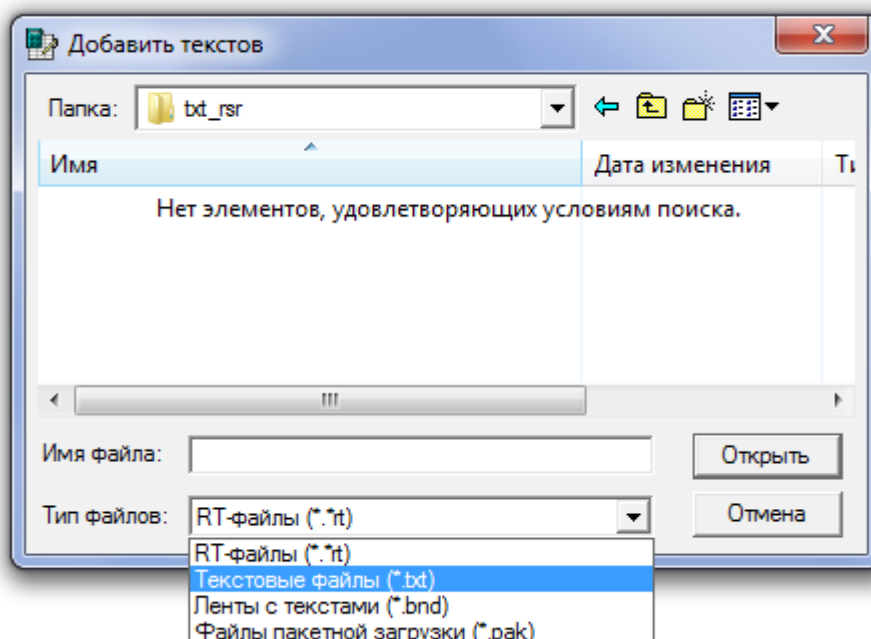
Итак, мы с Вами получили набор текстов в txt формате. Что-то типа этого:

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

Имя	Дата изменения	Тип файла	Размер
03BRN73M.txt	11.01.2011 11:55	Текстовый докум...	4 КБ
03BRN74M.txt	11.01.2011 11:58	Текстовый докум...	5 КБ
3Agp19w.txt	11.01.2011 10:48	Текстовый докум...	3 КБ
3Agp22M.txt	11.01.2011 10:57	Текстовый докум...	3 КБ
3Agp33M.txt	11.01.2011 11:01	Текстовый докум...	4 КБ
3Agp43w.txt	11.01.2011 11:03	Текстовый докум...	3 КБ
3Agp45w.txt	11.01.2011 11:05	Текстовый докум...	3 КБ
3BLTKV60W.txt	11.01.2011 11:32	Текстовый докум...	12 КБ
3BLTKV62M.txt	11.01.2011 11:36	Текстовый докум...	12 КБ
3BLTKV63W.txt	11.01.2011 11:39	Текстовый докум...	19 КБ
3BLTKV70M.txt	11.01.2011 11:42	Текстовый докум...	5 КБ
3BLTKV72W.txt	11.01.2011 11:46	Текстовый докум...	20 КБ
3Brs 18m.txt	11.01.2011 15:18	Текстовый докум...	3 КБ
3Brs21W .txt	11.01.2011 12:00	Текстовый докум...	3 КБ
3Brs33M.txt	11.01.2011 15:08	Текстовый докум...	2 КБ
3Brs40m .txt	11.01.2011 15:12	Текстовый докум...	4 КБ
3Brs54M .txt	11.01.2011 15:14	Текстовый докум...	3 КБ
3CHKN 23W.txt	11.01.2011 15:30	Текстовый докум...	4 КБ
3CHKN21W.txt	11.01.2011 15:22	Текстовый докум...	4 КБ
3CHKN23W.txt	11.01.2011 15:23	Текстовый докум...	4 КБ
3CHKN24W.txt	11.01.2011 15:27	Текстовый докум...	4 КБ
3CHKN25W.txt	11.01.2011 15:28	Текстовый докум...	4 КБ
3Dem24m.txt	11.01.2011 15:35	Текстовый докум...	3 КБ
3Dem34w.txt	11.01.2011 15:39	Текстовый докум...	2 КБ
3Dem39w.txt	11.01.2011 15:42	Текстовый докум...	2 КБ

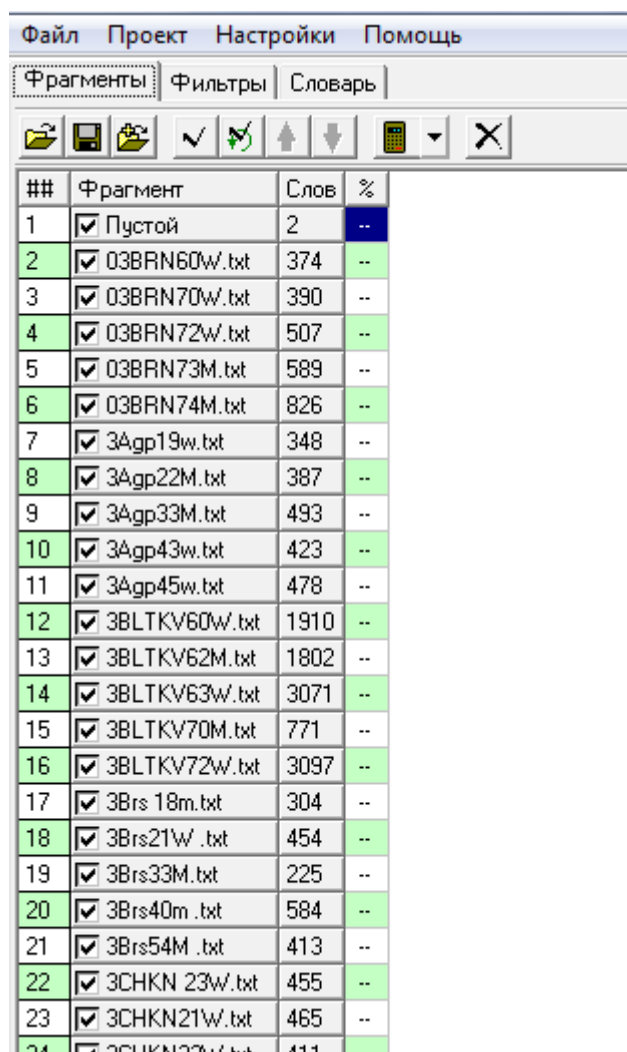
Элементов: 250

Загрузим их в программу ЛЕКТА. Для этого нажмём на кнопку «добавить текстов». Выберем тип файлов TXT



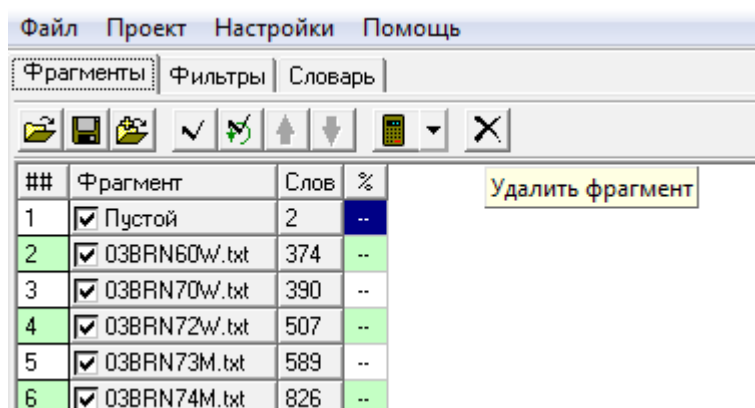
УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

Теперь выберем все интересующие нас файлы в нужной директории диска, например, через функцию Ctrl A. Нажимает ОТКРЫТЬ. Файлы загрузились.



##	Фрагмент	Слов	%
1	<input checked="" type="checkbox"/> Пустой	2	--
2	<input checked="" type="checkbox"/> 03BRN60w.txt	374	--
3	<input checked="" type="checkbox"/> 03BRN70w.txt	390	--
4	<input checked="" type="checkbox"/> 03BRN72w.txt	507	--
5	<input checked="" type="checkbox"/> 03BRN73M.txt	589	--
6	<input checked="" type="checkbox"/> 03BRN74M.txt	826	--
7	<input checked="" type="checkbox"/> 3Agp19w.txt	348	--
8	<input checked="" type="checkbox"/> 3Agp22M.txt	387	--
9	<input checked="" type="checkbox"/> 3Agp33M.txt	493	--
10	<input checked="" type="checkbox"/> 3Agp43w.txt	423	--
11	<input checked="" type="checkbox"/> 3Agp45w.txt	478	--
12	<input checked="" type="checkbox"/> 3BLTKV60w.txt	1910	--
13	<input checked="" type="checkbox"/> 3BLTKV62M.txt	1802	--
14	<input checked="" type="checkbox"/> 3BLTKV63w.txt	3071	--
15	<input checked="" type="checkbox"/> 3BLTKV70M.txt	771	--
16	<input checked="" type="checkbox"/> 3BLTKV72w.txt	3097	--
17	<input checked="" type="checkbox"/> 3Brs 18m.txt	304	--
18	<input checked="" type="checkbox"/> 3Brs21w .txt	454	--
19	<input checked="" type="checkbox"/> 3Brs33M.txt	225	--
20	<input checked="" type="checkbox"/> 3Brs40m .txt	584	--
21	<input checked="" type="checkbox"/> 3Brs54M .txt	413	--
22	<input checked="" type="checkbox"/> 3CHKN 23w.txt	455	--
23	<input checked="" type="checkbox"/> 3CHKN21w.txt	465	--
24	<input checked="" type="checkbox"/> 3CHKN23w.txt	411	--

Сразу удалим «пустой файл», формирующийся системой по умолчанию. Нажмём на кнопочку с изображением крестика.



##	Фрагмент	Слов	%
1	<input checked="" type="checkbox"/> Пустой	2	--
2	<input checked="" type="checkbox"/> 03BRN60w.txt	374	--
3	<input checked="" type="checkbox"/> 03BRN70w.txt	390	--
4	<input checked="" type="checkbox"/> 03BRN72w.txt	507	--
5	<input checked="" type="checkbox"/> 03BRN73M.txt	589	--
6	<input checked="" type="checkbox"/> 03BRN74M.txt	826	--

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

Мы видим, что ЛЕКТА показывает все наши статьи, отображая коды, созданные нами, а также число слов в тексте. Нажав на код, мы увидим справа текст файла:

The screenshot shows the LECTA software interface. On the left, there is a table with columns: ##, Фрагмент, Слов, and %. The table lists 25 fragments, with fragment 2 (03BRN60w.txt) selected. On the right, a preview window shows the text of the selected file, which is an interview transcript. The text includes questions and answers about living in Nizhny Novgorod and comparing it to Moscow.

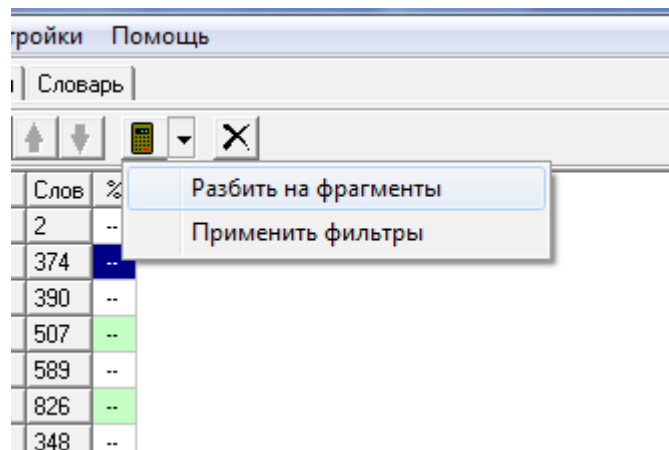
##	Фрагмент	Слов	%
1	✓ Пустой	2	--
2	✓ 03BRN60w.txt	374	--
3	✓ 03BRN70w.txt	390	--
4	✓ 03BRN72w.txt	507	--
5	✓ 03BRN73M.txt	589	--
6	✓ 03BRN74M.txt	826	--
7	✓ 3App19w.txt	348	--
8	✓ 3App22M.txt	387	--
9	✓ 3App33M.txt	493	--
10	✓ 3App43w.txt	423	--
11	✓ 3App45w.txt	478	--
12	✓ 3BLTKV60W.txt	1910	--
13	✓ 3BLTKV62M.txt	1802	--
14	✓ 3BLTKV63W.txt	3071	--
15	✓ 3BLTKV70M.txt	771	--
16	✓ 3BLTKV72W.txt	3097	--
17	✓ 3Brs18m.txt	304	--
18	✓ 3Brs21W.txt	454	--
19	✓ 3Brs33M.txt	225	--
20	✓ 3Brs40m.txt	584	--
21	✓ 3Brs54M.txt	413	--
22	✓ 3CHKN23W.txt	455	--
23	✓ 3CHKN21W.txt	465	--
24	✓ 3CHKN23W.txt	411	--
25	✓ 3CHKN24W.txt	523	--

Сейчас наша задача заключается в том, чтобы разбить тексты на приблизительно равные фрагменты. ЭТО ОЧЕНЬ ВАЖНО ДЛЯ ПРОВЕДЕНИЯ ДАЛЬНЕЙШЕГО АНАЛИЗА. Если фрагменты будут сильно варьироваться по величине, то сегментировать информационное пространство будет просто невозможно. Итак, нажимаем на кнопку ОБРАБОТКА и выбираем РАЗБИТЬ НА ФРАГМЕНТЫ.

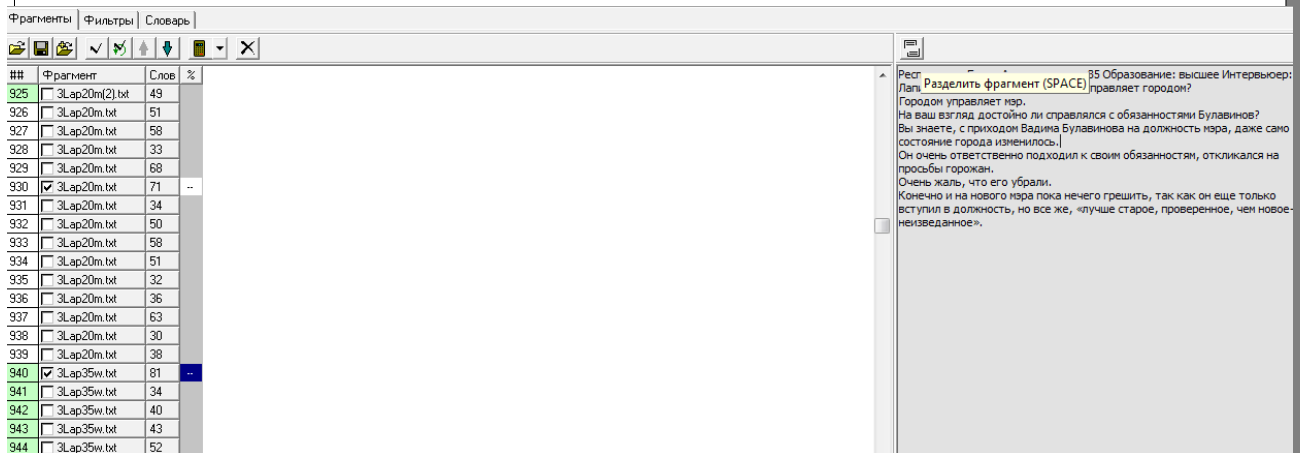
The screenshot shows the LECTA software interface with the 'Обработка' (Processing) button highlighted. The table of fragments is visible, and the 'Обработка' button is located to the right of the table.

##	Фрагмент	Слов	%
1	✓ Пустой	2	--
2	✓ 03BRN60w.txt	374	--
3	✓ 03BRN70w.txt	390	--
4	✓ 03BRN72w.txt	507	--
5	✓ 03BRN73M.txt	589	--
6	✓ 03BRN74M.txt	826	--

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.



Ставим галочку на РАЗБИТЬ ПО РАЗМЕРУ и выбираем размер: 50 +/- 20. Нажимаем ОК. Всё. Наши тексты разбились по размеру в рамках указанного диапазона. Однако разбились, возможно, и не все, и нам надо закончить разбиение вручную (программа не всегда может справиться с этим самостоятельно). Для ручной разбивки находим фрагменты, не разбившиеся на части: они отмечены знаком «-». Кликаем по ним в правой части программы приблизительно по середине фрагмента и желательно в конце абзаца или предложения (так, чтобы разделить мысли высказывания) и нажимаем пробел либо кнопку выше:



Так поступаем с фрагментами, которые явно сильно выходят за рамки указанного диапазона числа слов.

Далее поговорим о КЛЮЧЕВОМ этапе работы — составлении словаря контент-анализа.

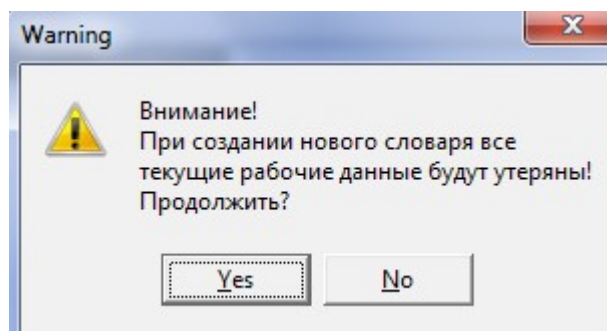
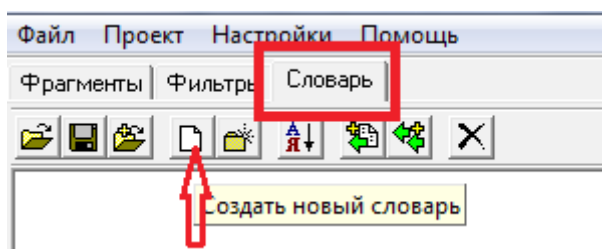
Глава вторая. Словарь

Этап создания словаря очень ответственный и достаточно специфический — грамотное составление словаря требует опыта такой работы. Составляя словарь впервые, исследователь может допустить ошибки, которые помешают дальнейшей работе над замером.

Итак, что же такое словарь контент-анализа?

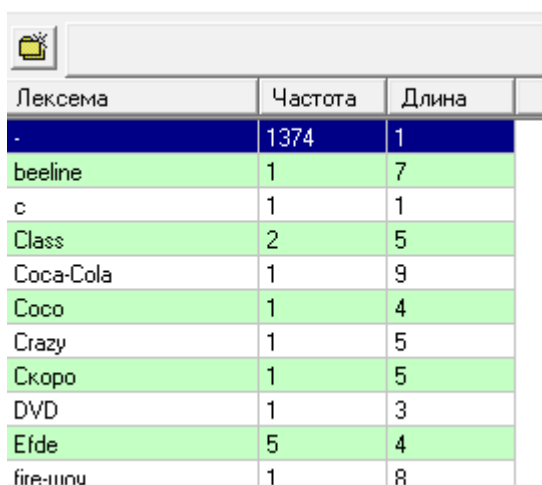
Под словарём в данном случае мы будем понимать набор слов и их словоформ, отражающих те или иные микротемы, микросюжеты, формирующие ключевое тематическое наполнение информационного пространства, представленное массивом текста. В нашем случае, в случае с работой в ЛЕКТА нас будут интересовать те слова, которые наиболее часто встречаются в тексте, а также обладают чёткой семантикой. Слова, входящие в словарь, должны быть конкретны и высокочастотны — это 2 наиболее важных требования к единицам счёта в рамках контент-анализа в ЛЕКТА.

Перейдём в интерфейс ЛЕКТА на закладку СЛОВАРЬ. В появившемся диалоговом окне-предупреждении о возможности потери текущих данных кликаем YES.



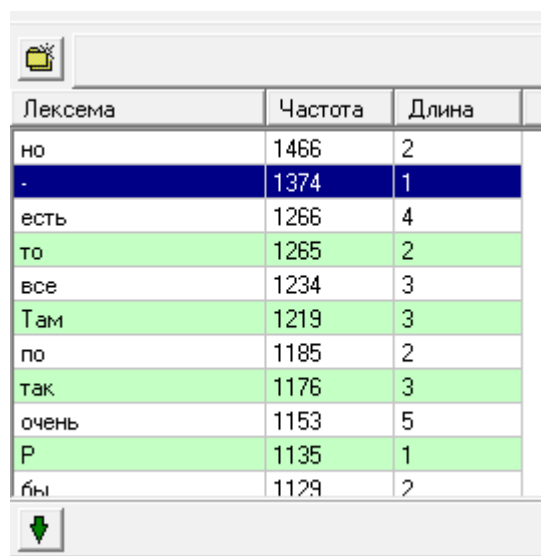
По умолчанию все слова массива, отображаемые справа, отсортированы по алфавиту. Вначале идёт латиница, потом кириллица.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.




Лексема	Частота	Длина
-	1374	1
beeline	1	7
с	1	1
Class	2	5
Coca-Cola	1	9
Coco	1	4
Crazy	1	5
Скоро	1	5
DVD	1	3
Efde	5	4
fire-шину	1	8

Наша задача на этом этапе заключается в том, чтобы выбрать наиболее часто встречающиеся слова массива. Для этого нам надо кликнуть по заголовку столбца ЧАСТОТА. Теперь наверху оказались слова, чаще всего встречающиеся в наших текстах. Разумеется, в первую очередь это предлоги, союзы и другие неинформативные части речи.



Лексема	Частота	Длина
но	1466	2
-	1374	1
есть	1266	4
то	1265	2
все	1234	3
Там	1219	3
по	1185	2
так	1176	3
очень	1153	5
Р	1135	1
бы	1129	2

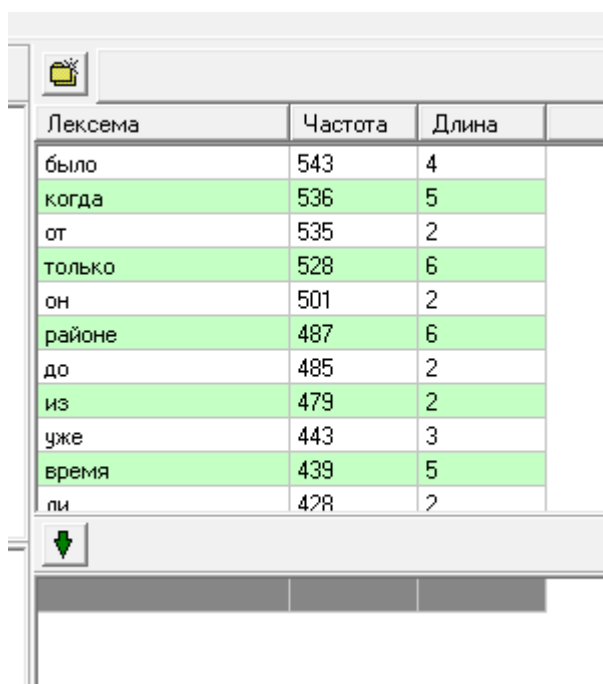


Приступим к выбору слов — единиц счёта контент-анализа. Напомню, что слова должны соответствовать 2-м критериям — они должны часто встречаться в тексте, и они должны иметь относительно чёткое значение. Станет понятно на примере: пропуская все слова, не имеющие точного значения (союзы, предлоги и т.д.) и опускаясь ниже, доходим до слов, обладающих более чёткой семантикой. Нас интересуют слова, которые могут быть употреблены с одинаковым или близкими значениями в разных контекстах. Так на рисунке ниже мы видим часто встречающееся слово ВРЕМЯ и также часто встречающееся слово РАЙОНЕ. Слово ВРЕМЯ явно не имеет чёткого значения — может быть использовано в

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

разных контекстах с разными значениями (трудное время, время от времени, во время и т.д.).

Зная, что наши тексты о социально-экономической действительности города Нижнего Новгорода, мы можем сделать вывод о том, что слово РАЙОН имеет достаточно определённую семантику и используется в массиве текстов чаще всего в единственном значении — синонимично ТЕРРИТОРИИ. Для того, чтобы сделать своеобразную закладку этого слова надо кликнуть по нему и нажать пробел. Закладка отобразится выше. Таким же образом, постепенно опускаясь вниз, (двигаясь к менее частотным лексемам) мы добавляем закладки и к другим единицам счёта. Число таких закладок может быть любым. Оптимально от 70 до 150. Однако я работал со словарём и в 299 единиц счёта, но это был контент-анализ 57 000 страниц текста. Ещё отмечу — двигаться вниз лучше используя стрелочку ВНИЗ справа — так переходы будут плавными и частотность будет снижаться постепенно. И ещё — не бойтесь сделать одинаковые закладки — работа со словоформами пойдёт далее.



Лексема	Частота	Длина
было	543	4
когда	536	5
от	535	2
только	528	6
он	501	2
районе	487	6
до	485	2
из	479	2
уже	443	3
время	439	5
ли	428	2

Итак, допустим мы насобирали около 100 закладок. Я, для примера пока набрал только 15-20. Частота не должна быть ниже 8-12 употреблений слова (суммарно со словоформами, но об этом ниже). Выглядит это примерно так:

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

Лексема	Частота	Длина
год	90	3
хожу	90	4
чаще	90	4
могут	90	5
находится	90	9
Правда	89	6
центре	89	6
квартиру	89	8
метро	88	5
теперь	88	6
пб	86	2

Перейдём к работе с словоформами — кликнем по заголовку столбца ЛЕКСЕМА. Наши слова отсортировались по алфавиту. Кликнем по самой первой закладке слева, с которой и начнём работу. И вот мы видим все словоформы слова РАЙОНЕ — в моём примере это первая закладка. Принцип работы на этом этапе заключается в том, чтобы создать папки, в которые мы поместим все словоформы слов, вошедших в закладки. Для этого кликнем по слову РАЙОНЕ в закладке и нажмём на кнопку выше — СОЗДАТЬ ИМЕНОВАННЫЕ ПАПКИ ИЗ ВЫДЕЛЕННОГО.

Лексема	Частота	Длина
районе	487	6
района	95	6
районам	11	7
районами	8	8
районах	46	7
районной	1	8
районные	1	8
районов	45	7
районом	18	7
район-то	1	8

Слева появляется папка с названием РАЙОНЕ и числом слов (пока только слов «районе») в ней. Кликаем по ней, тем самым как бы заходя в неё. Кликаем в нижнем дубле перечня папок (всего их 2 — так работать удобнее — в верхней части Вы

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

работаете с общим словарём и закладками, а в нижней с отдельными папками). Теперь нажимаем на кнопочку ВНИЗ, перемещая лексемы из общего словаря в папку:

Файл Проект Настройки Помощь

Фрагменты | Фильтры | Словарь

районе | Нижний | здоровье | Новгород | Пробки | це

Лексема	Частота	Длина
райисполкома	2	12
района	95	6
районам	11	7
районами	8	8
районах	46	7
районной	1	8
районные	1	8
районов	45	7
районом	18	7
район-то	1	8
райони	18	6

↓

рай	Перенести выделенное	6
район	370	5
-Район	2	6

Так нам надо перенести все лексемы из верхней части экрана в нижнюю — из общего словаря в конкретную папку. Подчеркну, что слова должны иметь единую семантику, ведь эту папку мы будем использовать в дальнейшем как единую единицу счёта. То есть слова могут иметь и разную семантику, и нам нужно включать в словарь максимально близкие по значению слова. Обычно это бывают все словоформы, но встречаются и исключения, особенно при работе с большими массивами текста.

Если Вы ошиблись и включили в папку лишнее, неконгруэнтное семантике слово, просто нажмете на него в микрословаре папки и далее используйте кнопку ВВЕРХ — слово вернётся в базовый общий словарь.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

раковине	1	8
раковины	1	8
ракушки	1	7
раллистом	1	9
рамках	3	6
райкой	1	5
↑		
рай	6	
район	5	
-Район	2	6
района	95	6
районам	11	7
районами	8	8
районах	46	7
районной	1	8
районные	1	8
районов	45	7
районом	18	7
район-то	1	8
району	18	6
райончик	4	8
райончики	1	9
районы	120	8
Райцентр	1	8

Ещё хотел бы обратить Ваше внимание на то, что работу можно ускорить, хотя при этом, возможно, она будет сделана не столько качественно. Обведя несколько слов в базовом словаре, зажав левую клавишу мышки, система сама определит их общую часть, которую мы можем включить в словарь. Например, это будет слово «РАЙОН*» и все слова, начинающиеся с РАЙОН попадут в папку. При этом уже не обязательно отдельно включать в словарь все словоформы.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

Лексема	Частота	Длина
районной	1	8
районные	1	8
районов	45	7
районом	18	7
район-то	1	8
району	18	6
райончик	4	8
райончики	1	9
районы	126	6
райсовета	1	9
Райсовет	1	8

районе	107	8
район	370	7
-Район	2	6
района	95	6
районам	11	7

Переходи к следующей закладке и действуем по аналогии: создаём новую папку и переносим в неё словоформы лексемы из базового словаря. Так получаем 70-150 папок. Сейчас нам необходимо избавиться от дублей — похожих папок, объединив их. Для этого выберем, например, папки «Горький» и «горького». Учитывая то, что это старое название Нижнего Новгорода, можно предположить что слова имеют единую семантику. Используя как и ранее кнопки со стрелкой для перемещения лексем, соберём все лексемы в единую папку.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

Файл Проект Настройки Помощь

Фрагменты | Фильтры | Словарь

Лексема	Частота	Длина
Горький	53	7
Горьком	6	7
Горьким	6	7
горького	58	8

Перенести выделенное 8

Пустую же папку просто удалим — нам она не нужна.

Файл Проект Настройки Помощь

Фрагменты | Фильтры | Словарь

Удалить выделенное

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

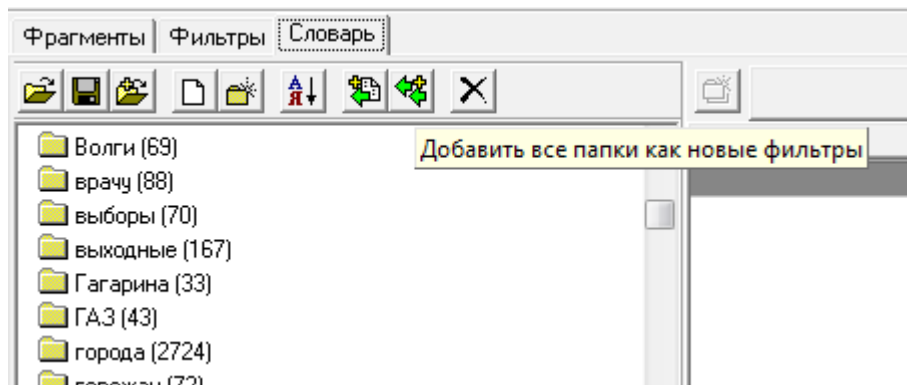
Ещё раз проверим все папки на наличие дублей. Сохраним проект (меню Проект/Сохранить проект). Да, забыл сказать — словари можно отдельно загружать и отдельно сохранять без сохранения всего проекта. Это может быть актуально, если Вы решили применить словарь к другому массиву текстов.

Далее поговорим о применении словаря к фрагментам текста.

Глава третья. Исходная матрица контент-анализа

Сейчас я расскажу вам о следующем этапе проведения контент-анализа в ЛЕКТА — применении словаря к фрагментам текстового массива и получении базовой матрицы контент-анализа. С матрицей контент-анализа можно продолжать работать не только в ЛЕКТА, но и в таких статистических пакетах, как SPSS, Statistica и иных.

Итак, приступим к работе с фильтрами — нашими папками, содержащими наборы лексем. Они станут переменными при дальнейшей работе по проведению факторизации матрицы контент-анализа. Для того чтобы создать единый набор таких фильтров нажмём на кнопку «добавить все папки как новые фильтры»



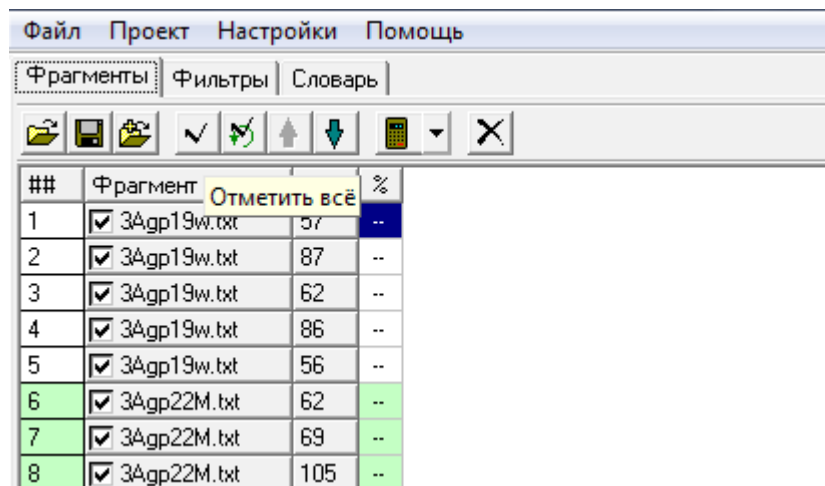
Переходим на закладку ФИЛЬТРЫ и видим развёрнутые папки с лексемами. Сразу же нажимаем на «крестик», чтобы удалить первый пустой фильтр, формируемый системой до добавления фильтров пользователями.

##	Фильтр	Лекс.	%	Удалить фильтр					
1	<input checked="" type="checkbox"/> Пустой	1	--						
2	<input checked="" type="checkbox"/> аварии	7	--	аварии	аварий	аварийное	аварийные	аварийных	аварию
3	<input checked="" type="checkbox"/> автобусы	15	--	автобусы	автобус	автобуса	автобусам	автобусами	автобусах
4	<input checked="" type="checkbox"/> Автозаводском	15	--	Автозаводском	Автозавод	автозавода	автозаводе	автозаводом	Автозаводск
5	<input checked="" type="checkbox"/> администрации	8	--	администрации	административную	административных	администраторов	администрацией	администрац
6	<input checked="" type="checkbox"/> аптеки	13	--	аптеки	аптек	аптека	аптеками	аптекарские	аптеках
7	<input checked="" type="checkbox"/> архитектуры	14	--	архитектуры	архитектор	архитектором	архитектур	архитектура	архитектуре
8	<input checked="" type="checkbox"/> бабушки	13	--	бабушки	бабка	бабкой	бабульки	Бабуля	бабушек
9	<input checked="" type="checkbox"/> бедные	8	--	бедные	беднее	бедность	бедный	бедным	бедных
10	<input checked="" type="checkbox"/> безопасности	14	--	безопасности	безопасен	Безопасна	безопаснее	безопасней	безопасно
11	<input checked="" type="checkbox"/> богатые	14	--	богатые	богат	богатая	богатеи	богатой	богатому
12	<input checked="" type="checkbox"/> больницы	18	--	больницы	больниц	Больница	больницам	больницами	больницах
13	<input checked="" type="checkbox"/> Большая	2	--	Большая	Большую				
14	<input checked="" type="checkbox"/> будни	7	--	будни	будней	будние	будничные	будничный	будням
15	<input checked="" type="checkbox"/> Верхней	9	--	Верхней	верхнее	Верхние	верхний	Верхним	Верхним
16	<input checked="" type="checkbox"/> вечером	14	--	вечером	вечера	вечерами	вечером	вечером	вечерком
17	<input checked="" type="checkbox"/> власти	5	--	власти	властей	власть	властям	властями	
18	<input checked="" type="checkbox"/> внуки	13	--	внуки	Внук	внука	внукам	внуками	внуках

Здесь мы можем редактировать наш набор фильтров, однако это бывает необходимо ОЧЕНЬ редко, поэтому сейчас мы не будем останавливаться на этом

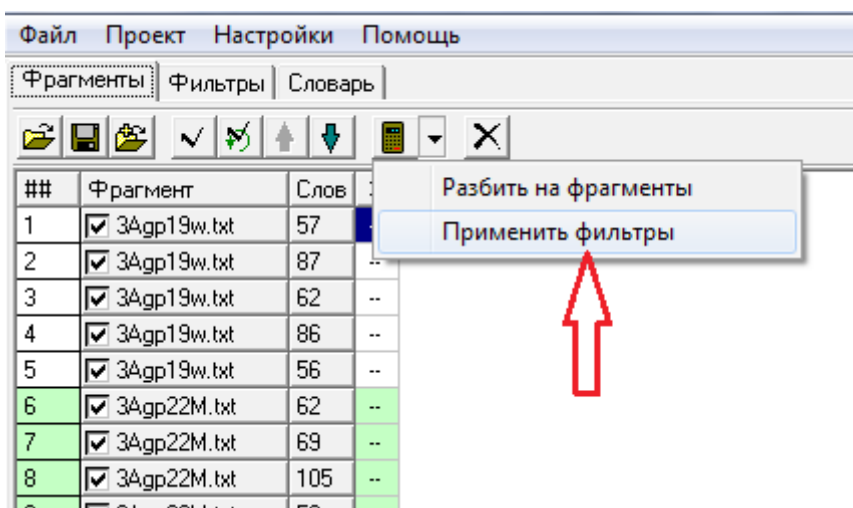
УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

функционале ЛЕКТА. Переходим вновь в закладку ФРАГМЕНТЫ. Здесь нам необходимо отметить каждый из фрагментов текстов галочкой. Для этого кликнем по соответствующей кнопке в меню. Этот шаг очень важен! Не забудьте про него!



##	Фрагмент	Отметить все	%
1	<input checked="" type="checkbox"/> 3Agp19w.txt	57	--
2	<input checked="" type="checkbox"/> 3Agp19w.txt	87	--
3	<input checked="" type="checkbox"/> 3Agp19w.txt	62	--
4	<input checked="" type="checkbox"/> 3Agp19w.txt	86	--
5	<input checked="" type="checkbox"/> 3Agp19w.txt	56	--
6	<input checked="" type="checkbox"/> 3Agp22M.txt	62	--
7	<input checked="" type="checkbox"/> 3Agp22M.txt	69	--
8	<input checked="" type="checkbox"/> 3Agp22M.txt	105	--

Теперь, когда все фрагменты выделены, применим наши фильтры — наборы лексем — к ним. Для этого выберем пункт меню «применить фильтры». Далее выбираете название и директорию для будущей матрицы, содержащей результат применения фильтров к текстам. С нетерпением ждём обработки материала...



##	Фрагмент	Слов	%
1	<input checked="" type="checkbox"/> 3Agp19w.txt	57	--
2	<input checked="" type="checkbox"/> 3Agp19w.txt	87	--
3	<input checked="" type="checkbox"/> 3Agp19w.txt	62	--
4	<input checked="" type="checkbox"/> 3Agp19w.txt	86	--
5	<input checked="" type="checkbox"/> 3Agp19w.txt	56	--
6	<input checked="" type="checkbox"/> 3Agp22M.txt	62	--
7	<input checked="" type="checkbox"/> 3Agp22M.txt	69	--
8	<input checked="" type="checkbox"/> 3Agp22M.txt	105	--

Обычно с массивом текста в 100-200 страниц операция длится около 30 секунд — минуты. С массивом англоязычного текста на 57 000 страниц (один из моих замеров, проведённых в рамках диссертационной работы) ЛЕКТА «возилась» около 12 часов. Для таких массивов нужен как опыт проведения контент-анализа в ЛЕКТА так и неплохой компьютер. Для моего несколько устаревшего ноутбука это было настоящее испытание.

Итак, мы дождалась завершения обработки массива. Находим наш файл. Сейчас он не имеет расширения. Кликаем по нему 2 раза и открываем с помощью EXCEL. Вот что мы видим: (я убрал почти все переменные, чтобы показать матрицу полностью в ширину)

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

HY27													
	A	B	C	D	E	F	G	H	HO	HP	HQ	HR	HS
1	##	Фрагмент	Слов	аварии	автобусы	Автозаво	админист	аптеки	%				
2	1	3Agp19w.txt	57	0	1	0	0	0	15	Проблема мобильности: внутригородская			
3	2	3Agp19w.txt	87	1	1	0	0	0	20	И.: По дороге вы попадаете в пробки? Р.:			
4	3	3Agp19w.txt	62	0	0	0	0	0	20	Замечала не раз, что движение становится			
5	4	3Agp19w.txt	86	0	2	0	0	0	16	Р.: На выходные многие выезжают за горо			
6	5	3Agp19w.txt	56	0	0	0	0	0	19	Р.: Мои родители ездят на личном автомо			
7	6	3Agp22M.txt	62	0	0	0	0	0	11	№8 Проблема мобильности: внутригород			
8	7	3Agp22M.txt	69	0	0	0	0	0	2	И.: Сталкиваетесь ли Вы с какими-нибудь			
9	8	3Agp22M.txt	105	0	0	0	0	0	11	##SL И.: А в пробки Вы попадаете? Р.: Поп			
10	9	3Agp22M.txt	58	0	0	0	0	0	22	И.: Приезжая на работу в будни, и до мест			
11	10	3Agp22M.txt	93	0	1	0	0	0	12	В некоторых ТЦ это может быть как крыта			
12	11	3Agp33M.txt	68	0	0	0	0	0	14	№10 Проблема мобильности: внутригоро			
13	12	3Agp33M.txt	102	0	0	0	0	0	13	И.: Можете привести некоторые примерь			
14	13	3Agp33M.txt	69	0	0	0	0	0	13	но зная ситуацию на дорогах, (благо есть с			
15	14	3Agp33M.txt	63	0	0	0	0	0	15	Р.: Думаю, что да! Конечно, преимущест			
16	15	3Agp33M.txt	79	0	0	0	0	0	16	Р.: С какой стороны посмотреть: если с то			
17	16	3Agp33M.txt	112	0	1	0	0	0	16	А сестра ездит на маршрутках и автобусах			
18	17	3Agp43w.txt	109	0	0	0	0	0	14	№7 Проблема мобильности: внутригород			
19	18	3Agp43w.txt	95	0	0	0	0	0	15	Р.: Обычно, после шести часов вечера и до			
20	19	3Agp43w.txt	122	0	1	0	0	0	18	И.: А этот припаркованный на обочинах тр			
21	20	3Agp43w.txt	97	0	0	0	0	0	10	У мужа свой автомобиль, но у него гибкий			
22	21	3Agp45w.txt	104	0	1	0	0	0	10	№9 Проблема мобильности: внутригород			
23	22	3Agp45w.txt	114	1	1	0	0	0	14	А далее это различного рода ДТП и авария			
24	23	3Agp45w.txt	78	0	0	0	0	0	16	##SL И.: А выходные Вы проводите в город			

В первом столбце отображается порядковый номер фрагмента текста. Во втором — код исходного текста фрагмента. В третьем — число слов во фрагменте. Далее идут названия папок, включивших себе актуальные лексемы, выбранные нами в ходе работы над словарём. Соответственно на пересечении строк и столбцов указано, какое количество раз встречается любая из лексем в папке словаря во фрагменте текста. Столбец % для нас не будет актуален. И последний столбец содержит сами фрагменты текстов.

Как вы понимаете, уже в EXCEL методом простой сортировки и базовых функций с этой матрицей можно плодотворно работать. Однако, наша работа будет глубже и продуктивнее. В следующей статье я расскажу Вам о том, как идентифицировать отдельные тематические блоки в массиве текстов — найти группы сильно коррелирующих лексем и фрагментов текстов с помощью факторного анализа. Эта работа будет проведена в программе STATISTICA, хотя в новой версии ЛЕКТА модуль факторного анализа позволяет не прибегать к стороннему софту.

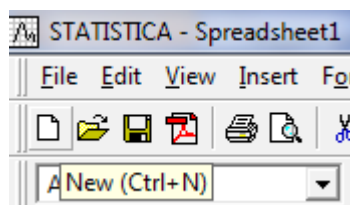
Глава четвёртая. Факторный анализ в Statistica

Теперь я расскажу вам как обрабатывать базовую матрицу контент-анализа в программе STATISTICA. Обработка будет производиться с помощью факторного анализа. Подчеркну, что в новой версии ЛЕКТА можно проводить факторный анализ без использования сторонних программ типа Statistica или SPSS. Однако, у меня не руках нет новой версии ЛЕКТА, поэтому объяснять принцип идентификации ключевых тематических блоков текстового массива буду на примере программы Statistica. Не пугайтесь — ничего сложного здесь нет, особенно если не лезть в дебри. Итак, поехали: факторный анализ в Statistica!

Объяснять факторизацию буду с использованием Statistica 8 в англоязычной версии. В подробности вдаваться не буду — просто повторяйте те же действия, что и в инструкции, и всё получится! Напомню — суть этой работы на данном этапе сводится к определению главных тем текстового массива. Выражены они будут набором переменных (наших единиц счёта — групп лексем) и набором фрагментов текстов по каждой из тем.

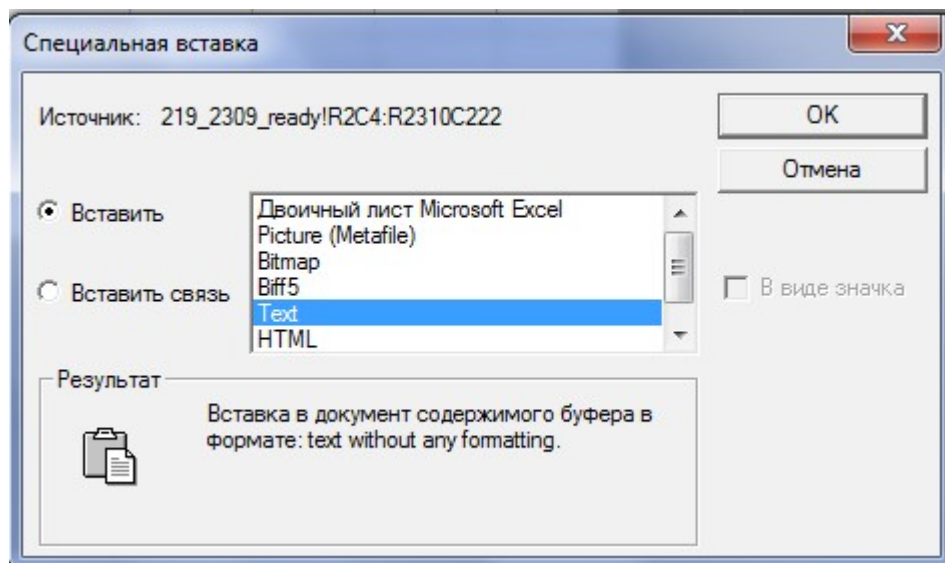
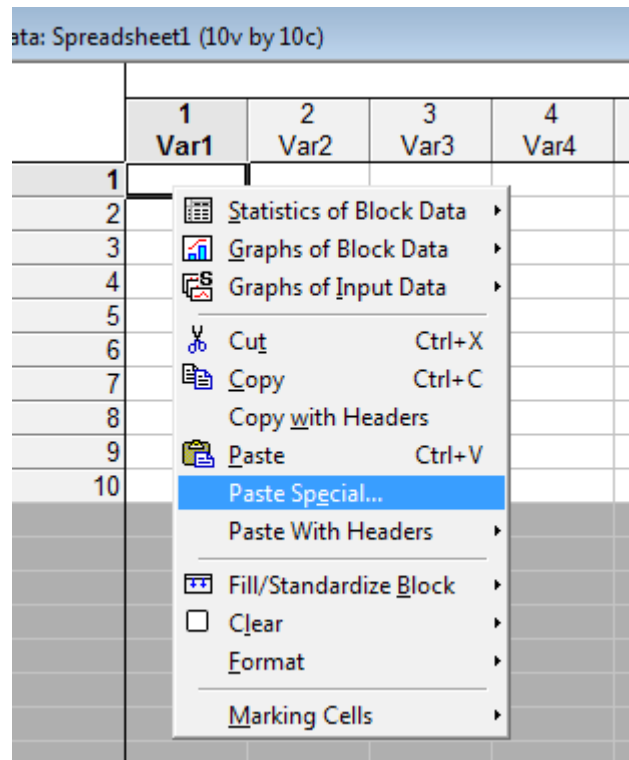
Открываем нашу матрицу контент-анализа в EXCEL. Далее нам потребуется скопировать только ту часть таблицы, в которой содержатся показатели частотности. То есть мы не будем копировать первые три столбца, последний, содержащий фрагменты текстов, предпоследний столбец и первую строку — копируем только столбцы, в которых отражена частотность встречаемости переменных.

Тем временем открываем статистику и создаём новый лист, кликнув по иконке CREATE



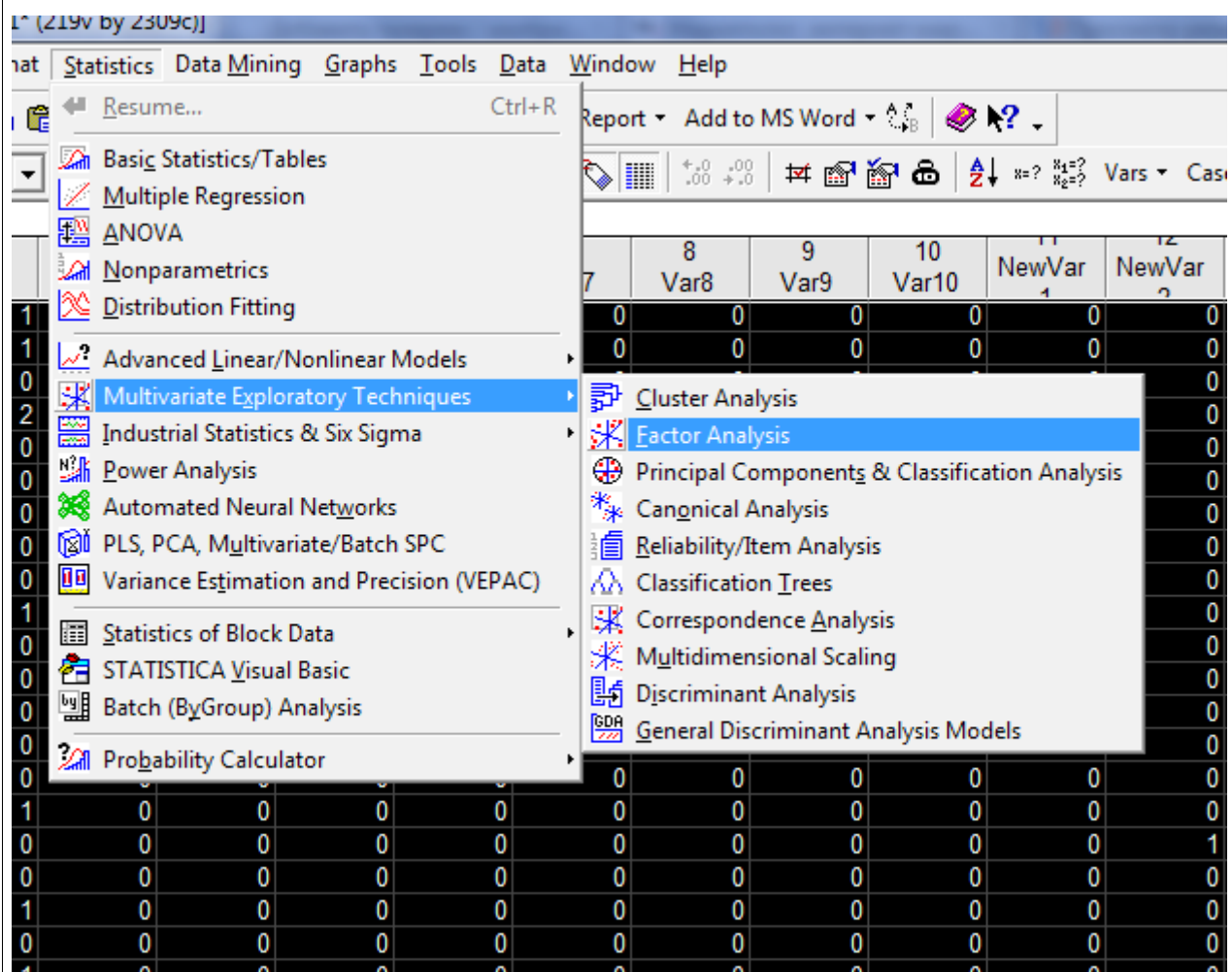
Появляется рабочий лист — пустая таблица. Ставим курсор в первую ячейку первой строки, нажимаем правую кнопку мыши, выбираем опцию PASTE SPECIAL (специальная вставка). Выбираем тип вставки TEXT. Нажимаем ОК. Довольно странно, что делать нужно так, но это помогает избежать некоторых ошибок при вставке.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.



Видим основную часть нашей матрицы, вставленной в рабочий лист STATISTICA. Теперь открываем пункт меню STATISTICS, подменю MULTIVARIATE EXPLORATORY TECHNIQUES и далее пункт FACTOR ANALYSIS.

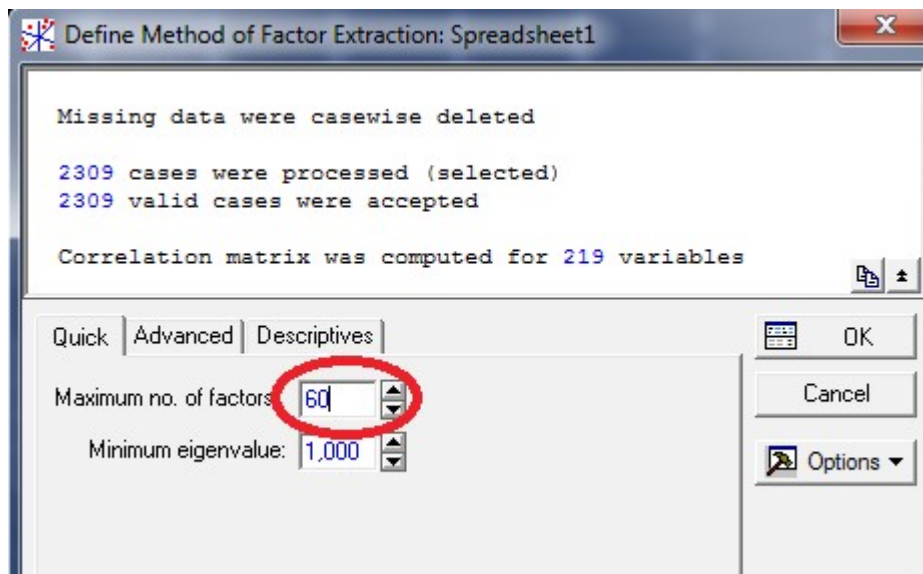
УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.



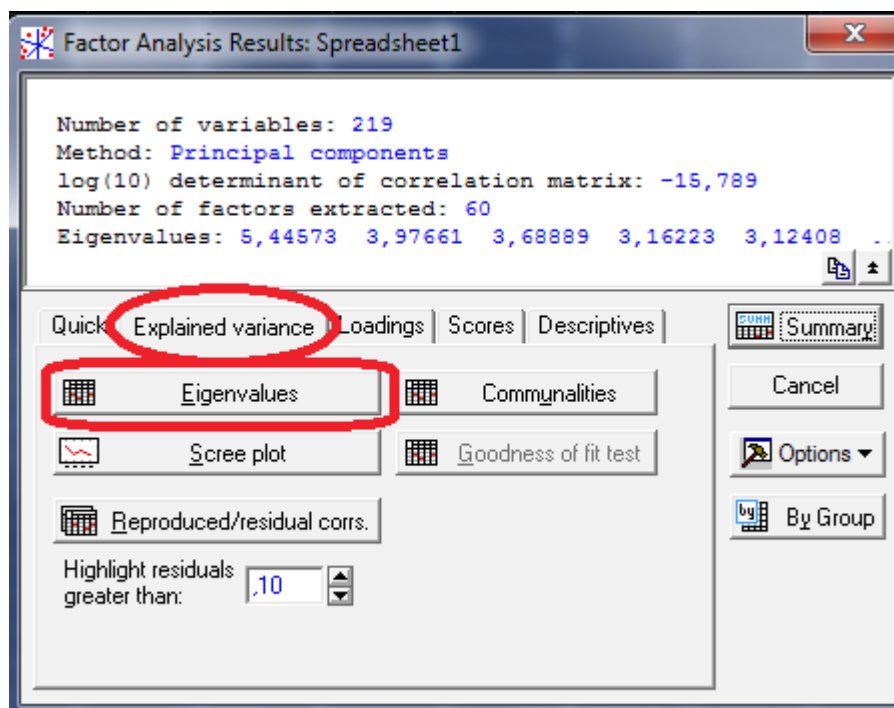
Кликаем по кнопке VARIABLES и выбираем опцию SELECT ALL — то есть выбираем все переменные из нашего набора — в данном примере все 219 штук. Обычно их бывает меньше — 100-150. Другими словами, мы задействуем для анализа все созданные нами наборы лексем — все папки. Нажимаем ОК и потом снова ОК.

Далее нам предстоит выбрать оптимальное количество факторов (микротем) в нашей факторной модели, своеобразной структуре информационного пространства изучаемого предмета. По умолчанию нам предлагается выбрать 2 фактора — нас это явно не устраивает, так как такую модель будет очень трудно анализировать и она будет слишком генерализована, каждый из факторов будет содержать слишком много переменных — например 100 и 119. Для удобства анализа нам надо добиться иного распределения переменных — приблизительно 4-8 на каждый из факторов. Учитывая то, что у нас всего 219 переменных, факторов должно быть 27-54 или чуть больше или чуть меньше. Теперь определим оптимальное количество факторов. Для этого впишем в поле MAXIMUM NO OF FACTORS например число 60 (с запасом). Нажмём ОК.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.



Перейдём в закладку EXPLAINED VARIANCE и кликнем по кнопке EIGENVALUES.



Открывается таблица. Из неё копируем первый столбец (EIGENVALUES) и вставляем его в новый лист EXCEL, желательно, в новую книгу — наш основной рабочий документ. Вставляем в первый столбец — курсор на A1

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

Eigenvalues (Spreadsheet1)				
Extraction: Principal components				
Value	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	5,445726	2,486633	5,4457	2,48663
2	3,976611	1,815804	9,4223	4,30244
3	3,688895	1,684427	13,1112	5,98686
4	3,162233	1,443942	16,2735	7,43081
5	3,124084	1,426522	19,3975	8,85733
6	2,777752	1,268380	22,1753	10,12571
7	2,713239	1,238922	24,8885	11,36463
8	2,348840	1,072530	27,2374	12,43716
9	2,297016	1,048866	29,5344	13,48603
10	2,233476	1,019852	31,7679	14,50588
11	2,153506	0,983336	33,9214	15,48921
12	2,109697	0,963332	36,0311	16,45255
13	2,040619	0,931790	38,0717	17,38434
14	1,993436	0,910245	40,0651	18,29458
15	1,947153	0,889111	42,0123	19,18369
16	1,902546	0,868742	43,9148	20,05243
17	1,881808	0,859273	45,7966	20,91171
18	1,857345	0,848103	47,6540	21,75981
19	1,807921	0,825535	49,4619	22,58534
20	1,753133	0,800517	51,2150	23,38586
21	1,746787	0,797620	52,9618	24,18348
22	1,702455	0,777377	54,6643	24,96086
23	1,676062	0,765325	56,3403	25,72618
24	1,654447	0,755455	57,9948	26,48164
25	1,637771	0,747841	59,6326	27,22948
26	1,606615	0,733614	61,2392	27,96309
27	1,566203	0,715161	62,8054	28,67825
28	1,555338	0,710200	64,3607	29,38845

Теперь в ячейке B2 напишем формулу =A1-A2 и «протащим» формулу на всю длину столбца со значениями. Скопируем полученные значения (дельты) и вставим их в строку (транспонируем). Такая операция производится через специальную вставку. Выберем вставить «значения» и поставим галочку на «транспонировать».

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

D1

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	5,445726	1,469115											
2	3,976611	0,287716											
3	3,688895	0,526662											
4	3,162233	0,038149											
5	3,124084	0,346332											
6	2,777752	0,064513											
7	2,713239	0,364400											
8	2,348840	0,051824											
9	2,297016	0,063540											
10	2,233476	0,079970											
11	2,153506	0,043809											
12	2,109697	0,069078											
13	2,040619	0,047183											
14	1,993436	0,046283											
15	1,947153	0,044608											
16	1,902546	0,020737											
17	1,881808	0,024463											
18	1,857345	0,049424											
19	1,807921	0,054788											
20	1,753133	0,006346											
21	1,746787	0,044332											
22	1,702455	0,026393											
23	1,676062	0,021615											
24	1,654447	0,016676											

Специальная вставка

Вставить

- все
- формулы
- значения
- форматы
- примечания
- условия на значения

Операция

- нет
- сложить
- вычесть

пропускать пустые ячейки

с исходной темой

без рамки

ширины столбцов

формулы и форматы чисел

значения и форматы чисел

все условные форматы объединения

умножить

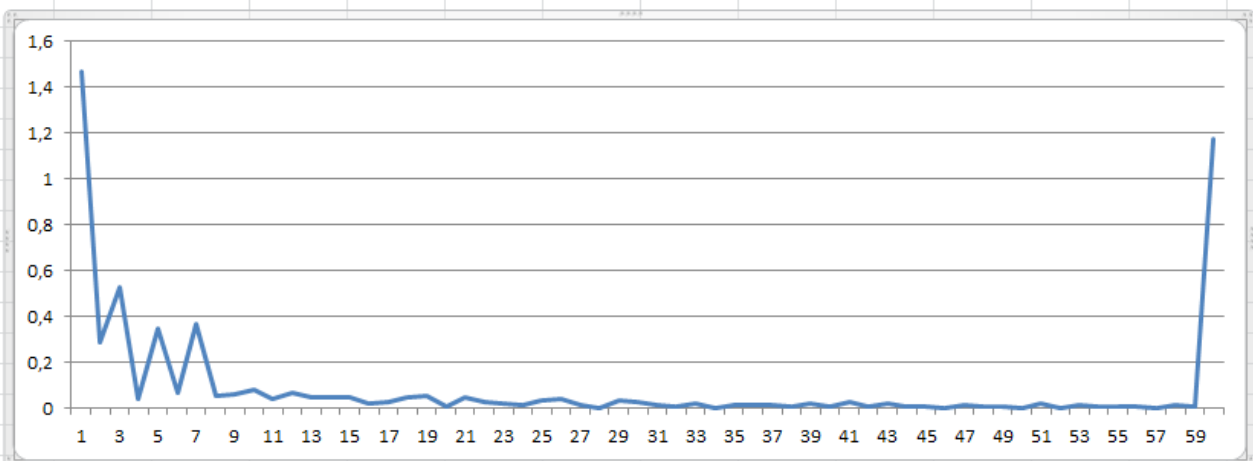
разделить

транспонировать

Вставить связь

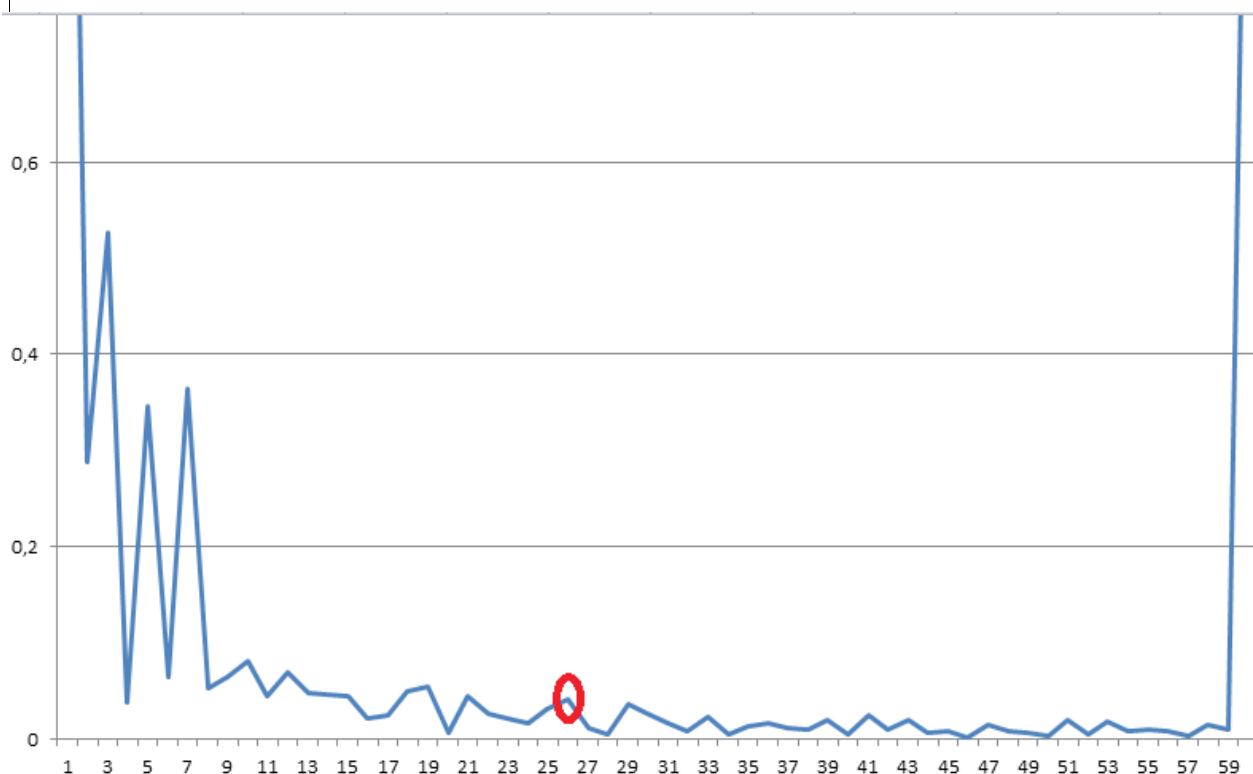
OK Отмена

Теперь из этой линейки значений построим график. Обводим все значения и выбираем диалоговое окно построения диаграмм — тип диаграммы — график.

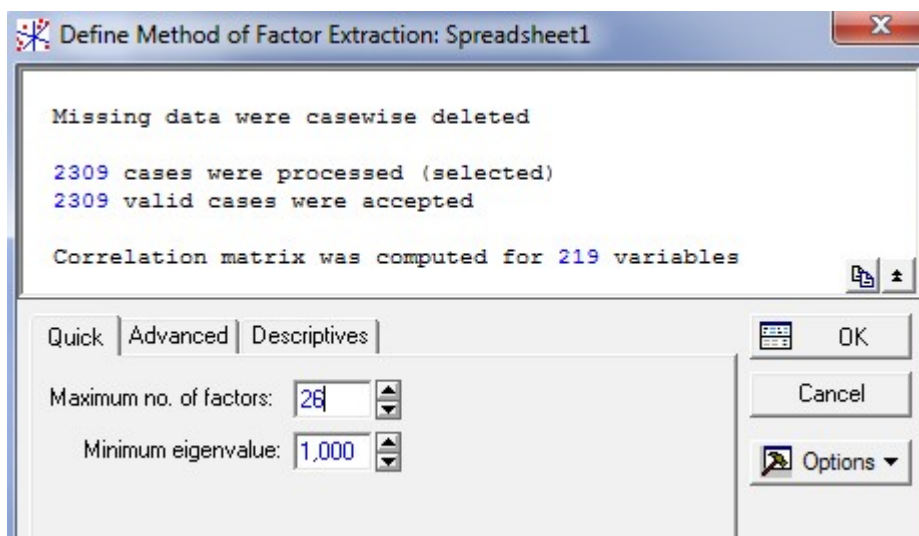


Нам надо выбрать значение на графике, после которого происходит резкий спад — снижение объясняющей способности модели. Сейчас мы видим, что таким значением может быть 7 (7 факторов), однако, как ранее уже было сказано, нам недостаточно такого числа факторов, поэтому нам надо растянуть график вниз и искать другие спады. Растягиваем... Получаем что-то вроде этого:

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

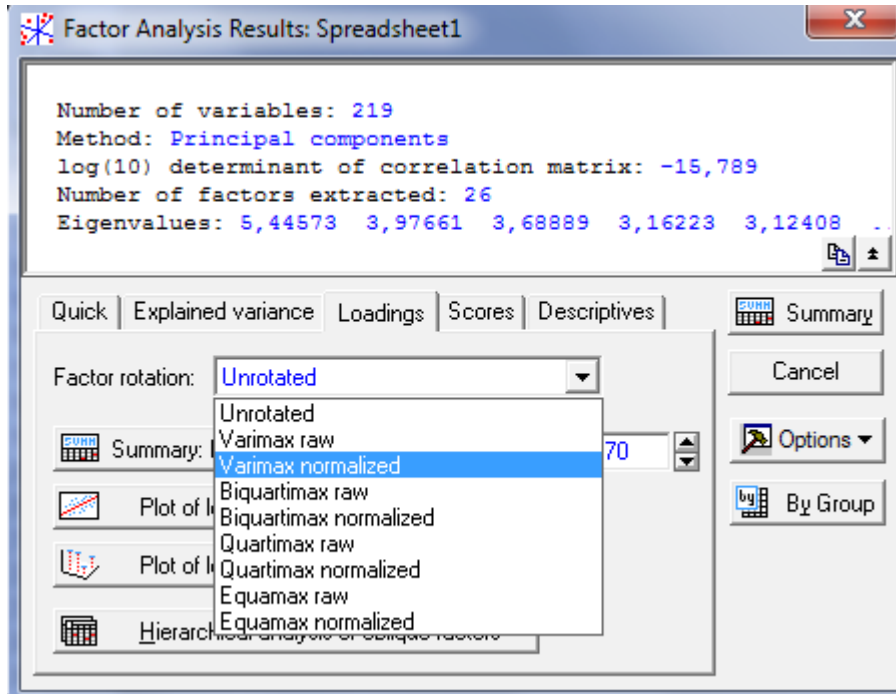


Запредельной точности нам не требуется. Можем выбрать, например, 26 факторов. Итак — 26 факторов будут адекватно объяснять тематический разброс информационного пространства изучаемого предмета. Возвращаемся в Statistica, нажимаем CANCEL и вписываем уже не ориентировочное количество факторов (не 60), а определённое, оптимальное — 26. (сразу подчеркну — на этом этапе я предпочитаю проводить работу несколько раз, дабы получить несколько вариантов матриц факторного анализа с разным количеством факторов — просто подстраховаться на случай получения малопригодной для работы матрицы). Вписав 26, нажимаем ОК.



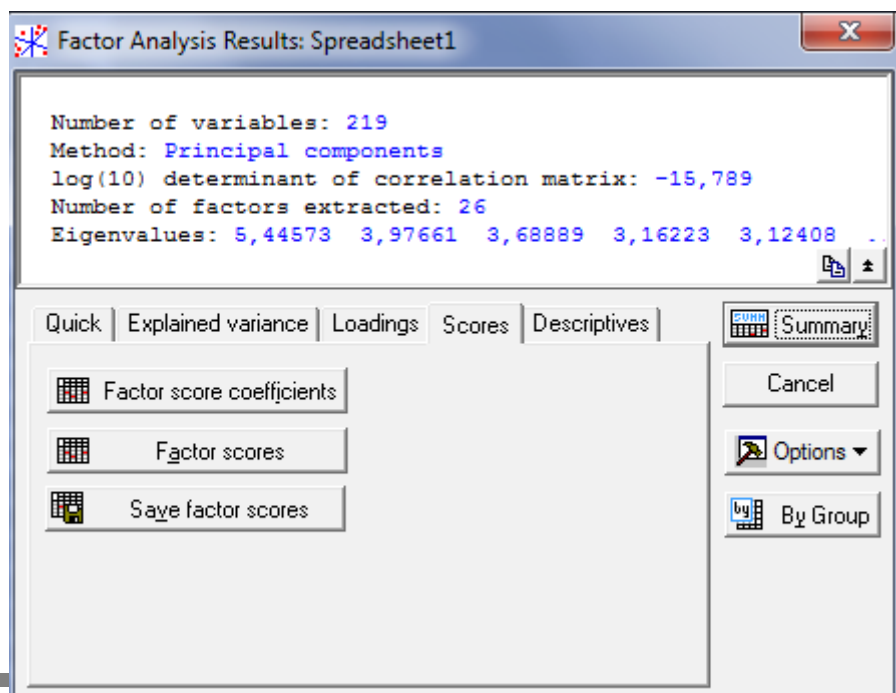
УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

В закладке LOADINGS выбираем тип факторного анализа — VARIMAX NORMALIZED (меню FACTOR ROTATION).



Кликаем на кнопку Summary Factor loadings и получаем первую матрицу факторного анализа — матрицу факторных нагрузок. Копируем её и вставляем на новый лист в EXCEL (курсор на B2)

Переходим в закладку SCORES и кликаем по кнопке FACTOR SCORES. Получаем вторую матрицу и вставляем её на новый лист в EXCEL — курсор на D2.



УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

Теперь из нашей исходной матрицы нам необходимо взять список названий переменных. Вначале скопируем их, а далее с помощью транспонирования поместим их вертикально в лист матрицы факторных нагрузок в первый столбец — курсор при вставке на А2. Во второй матрице нам необходимо заполнить первые 3 пустых столбца и вставить 2 после матрицы. В первые 3 вставляем значения из первых трёх столбцов исходной матрицы контент-анализа. В последние 2 также вставим значения и текст последний двух столбцов. Из матрицы факторных нагрузок удалим последние 2 строки (суммы) — они нам не потребуются в работе. К обоим матрицам добавим шапки F1, F2... F26. Ниже приложу файл с матрицами, так как столько скриншотов делать просто лень, и смотреть их Вам будет неудобно 😊

[матрицы факторного анализа](#)

Не беспокойтесь о том, что выглядит это непросто. Просто следуйте алгоритму и результат приятно обрадует! Далее я расскажу о небольшой обработке этих матриц в Excel. Это уже финишная прямая, к слову! Дальше — проще!

Глава пятая. Обработка матриц факторного анализа в Excel.

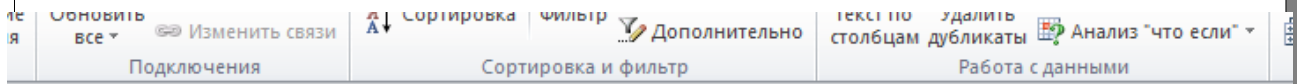
Сейчас мы поговорим об итоговой обработке полученных матриц факторного анализа. На этом этапе мы определим содержание факторов — тематических групп, выраженных набором переменных (лексем) и наборами фрагментов текстов.

Мы продолжим работу с теми же матрицами. Скачать их можно [по этой ссылке](#).

Для начала скопируем шапку (F1,F2...F26) и вставим её справа. Далее один столбец назовём тах и вставим шапку ещё раз. Чтобы было понятнее, ниже привожу скриншоты:

	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	
	F23	F24	F25	F26	F1	F2	F3	F4	F5	F6	F7
545	-0,026246	0,080735	-0,058412	0,001855							
988	-0,022868	0,027952	0,011481	-0,255552							
671	-0,010660	-0,070892	-0,017146	0,002341							
314	0,017954	0,018398	-0,038913	0,056884							
564	0,024950	0,039656	0,143003	0,003726							
578	-0,011629	0,006717	-0,002803	0,002901							
821	0,024083	-0,009022	0,011109	0,048885							
064	-0,730089	0,041743	-0,038984	0,019517							
186	-0,059077	-0,005015	-0,036966	-0,073156							
411	-0,700106	-0,056450	-0,055780	0,016581							
416	0,010229	0,070863	-0,046970	0,044261							
053	-0,079988	-0,062893	0,146206	-0,058108							
706	0,011634	-0,030511	0,008925	0,006233							
763	-0,299674	-0,179873	0,021927	0,135398							
866	0,064108	0,037514	-0,046857	0,143511							
574	0,019794	0,032478	-0,002458	0,072560							
572	0,008777	0,017313	0,033791	0,019664							
384	0,027130	-0,006464	-0,041087	-0,459505							
687	-0,000628	0,067594	0,336837	0,016178							
638	0,042390	0,021556	0,479851	0,017808							
687	0,066599	0,018857	0,052063	0,027110							
403	0,043412	-0,117334	0,114062	-0,116711							
241	0,029540	0,095067	-0,019244	0,017915							

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.



	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG
	F22	F23	F24	F25	F26	max	F1	F2	F3	F4	F5

Итого у нас в матрице factor loadings три одинаковых шапки Теперь немножко посчитаем. В столбце фактора F1 под второй шапкой пишем формулу — ABS(самое первое значение в матрице — в примере это B2). Протягиваем формулу по длине всей второй шапки и вниз до упора. В общем получаем абсолютные значения всей матрицы, создавай её дубль под второй шапкой.



AB2		fx		=ABS(B2)							
Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
4	F25	F26	F1	F2	F3	F4	F5	F6	F7	F8	F9
	.080735	-0,058412	0,001855								
	.027952	0,011481	-0,255552								
	.070892	-0,017146	0,002341								
	.018398	-0,038913	0,056884								
	.039656	0,143003	0,003726								
	.006717	-0,002803	0,002901								
	.009022	0,011109	0,048885								
	.041743	-0,038984	0,019517								
	.005015	-0,036966	-0,073156								
	.056450	-0,055780	0,016581								

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

		Подключения			Сортировка и фильтр			Работа с данными			
fx		=ABS(AA2)									
AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	
F21	F22	F23	F24	F25	F26	max	F1	F2	F3	F4	
0,058215	0,036545	0,026246	0,080735	0,058412	0,001855						
0,029514	0,011988	0,022868	0,027952	0,011481	0,255552						
0,130395	0,033671	0,01066	0,070892	0,017146	0,002341						
0,588622	0,026314	0,017954	0,018398	0,038913	0,056884						
0,048163	0,002564	0,02495	0,039656	0,143003	0,003726						
0,050748	0,061578	0,011629	0,006717	0,002803	0,002901						
0,01389	0,019821	0,024083	0,009022	0,011109	0,048885						
0,030334	0,008064	0,730089	0,041743	0,038984	0,019517						
0,066222	0,067186	0,059077	0,005015	0,036966	0,073156						
0,027270	0,000411	0,700106	0,05545	0,05570	0,015581						

Теперь в столбце max напишем формулу =МАКС(диапазон абсолютных значений с первого по 26 факторы). У нас в примере формула будет выглядеть как =МАКС(AB2:BA2). Далее протягиваем формулу по всей высоте матрицы.

		Подключения			Сортировка и фильтр			Работа с данными				Структура	
fx		=МАКС(AB2:BA2)											
AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG
19	F20	F21	F22	F23	F24	F25	F26	max	F1	F2	F3	F4	F5
1,015249	0,040614	0,058215	0,036545	0,026246	0,080735	0,058412	0,001855	0,249996					
0,03333	0,002108	0,029514	0,011988	0,022868	0,027952	0,011481	0,255552	0,61437					
1,057914	0,133713	0,130395	0,033671	0,01066	0,070892	0,017146	0,002341	0,445882					
1,039047	0,051905	0,588622	0,026314	0,017954	0,018398	0,038913	0,056884	0,588622					
1,053702	0,045573	0,048163	0,002564	0,02495	0,039656	0,143003	0,003726	0,615401					
1,013674	0,017287	0,050748	0,061578	0,011629	0,006717	0,002803	0,002901	0,372608					
1,001588	0,098021	0,01389	0,019821	0,024083	0,009022	0,011109	0,048885	0,098021					
0,01097	0,008911	0,030334	0,008064	0,730089	0,041743	0,038984	0,019517	0,730089					
1,031729	0,069443	0,066222	0,067186	0,059077	0,005015	0,036966	0,073156	0,18672					
1,018236	0,031635	0,037379	0,000411	0,700106	0,05645	0,05578	0,016581	0,700106					
1,026077	0,088152	0,023244	0,035416	0,010229	0,070863	0,04697	0,044261	0,343071					
1,014731	0,02254	0,054192	0,244053	0,079988	0,062893	0,146206	0,058108	0,244053					
1,009679	0,10372	0,012717	0,016706	0,011634	0,030511	0,008925	0,006233	0,539386					
1,051176	0,077867	0,177338	0,094763	0,299674	0,179873	0,021927	0,135398	0,378875					

И последняя формула. Пишем под третьей шапкой. Делим первое абсолютное значение на максимальное в строке. У нас в примере это будет выглядеть как =AB2/\$BB2. Далее соответственно протягиваем формулу по всей длине третьей шапки и высоте матрицы.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

Подключения			Сортировка и фильтр						
=AB2/\$B\$2									
AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI
25	F26	max	F1	F2	F3	F4	F5	F6	F7
,058412	0,001855	0,249996	0,106389	0,352766	0,305348	0,077955	0,040064	0,052779	0,04610
,011481	0,255552	0,61437	0,015448	1	0,126897	0,04998	0,081195	0,026377	0,03248
,017146	0,002341	0,445882	0,027964	0,055557	1	0,157952	0,023278	0,042051	0,15885
,038913	0,056884	0,588622	0,030257	0,08147	0,002929	0,077654	0,064276	0,054316	1,76E-0
,143003	0,003726	0,615401	0,034951	0,014068	0,078542	0,022547	0,222884	0,013331	0,14236
,002803	0,002901	0,372608	0,304097	0,110971	0,068859	0,298451	0,099826	0,032993	0,07570
,011109	0,048885	0,098021	0,198474	0,444219	0,340323	0,289642	0,24879	0,218054	0,06856
,038984	0,019517	0,730089	0,047179	0,041265	0,220982	0,047538	0,033623	0,035949	0,04253
,036966	0,073156	0,18672	0,212106	0,321694	0,315279	0,07317	0,083118	0,006306	0,11013
0,05578	0,016581	0,700106	0,020523	0,067409	0,141119	0,013006	0,009591	0,026764	0,0396
0,04697	0,044261	0,343071	0,109101	0,056703	0,01871	0,161574	0,16565	0,015341	0,05611
,146206	0,058108	0,244053	0,507347	0,043951	0,052729	0,082784	0,065988	0,006924	0,0956
,008925	0,006233	0,539386	0,017801	0,010185	0,007862	0,052661	0,016377	0,007376	0,10650
,021927	0,135398	0,378875	0,031177	0,208561	0,153216	0,16158	0,102446	0,317979	0,38622
,046857	0,143511	0,279596	0,119292	0,633001	0,008807	0,150077	0,175379	0,110321	0,11009

Далее сортируем значения в столбце F1 под третьей шапкой, так, чтобы вверху оказались единицы.

Подключения			Сортировка и фильтр				
=AB2/\$B\$2							
AZ	BA	BB	BC	BD	BE	BF	BI
5	F26	max	F1	F2	F3	F4	F5
,052063	0,02711	0,160909	1	0,343192	0,178129	0,065109	0,11
,041585	0,064658	0,220127	1	0,055882	0,59116	0,077948	0,20
,035003	0,023839	0,361246	1	0,119202	0,019962	0,12988	0,04
,110423	0,040824	0,332292	1	0,295525	0,028421	0,128412	0,07
,161657	0,079169	0,20782	1	0,242048	0,724674	0,270155	0,19
,000513	0,011615	0,229643	1	0,369343	0,295648	0,303819	0,13
,066574	0,072963	0,332002	1	0,030525	0,103925	0,129893	0,03
,082016	0,018703	0,742839	1	0,024861	0,008049	0,018301	0,0
,009835	0,010001	0,504425	1	0,075666	0,121679	0,175729	0,01
,044769	0,058868	0,441047	1	0,114855	0,324803	0,420346	0,06
,071962	0,006706	0,670807	1	0,051337	0,006655	0,002918	0,03
,047411	0,011377	0,12151	1	0,01104	0,18217	0,001057	0,11
,065254	0,055959	0,509944	1	0,020358	0,067863	0,055425	0,00
,091414	0,047265	0,216531	0,931887	0,183156	0,279242	0,263152	0,05
,063685	0,031161	0,151686	0,584709	0,030992	0,412529	1	0,18
,107453	0,045854	0,286705	0,570713	0,144049	0,451832	0,262211	0,1
,032791	0,026419	0,327431	0,528712	0,02504	0,008886	0,05601	0,09

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

Теперь после строк, содержащих единицы, вставим пустую строку.

BA	BB	BC	BD	BE	BF	F
0,02711	max	F1	F2	F3	F4	F
0,064658	0,160909	1	0,343192	0,178129	0,065109	
0,023839	0,220127	1	0,055882	0,59116	0,077948	
0,040824	0,361246	1	0,119202	0,019962	0,12988	
0,079169	0,332292	1	0,295525	0,028421	0,128412	
0,011615	0,20782	1	0,242048	0,724674	0,270155	
0,072963	0,229643	1	0,369343	0,295648	0,303819	
0,018703	0,332002	1	0,030525	0,103925	0,129893	
0,010001	0,742839	1	0,024861	0,008049	0,018301	
0,058868	0,504425	1	0,075666	0,121679	0,175729	
0,006706	0,441047	1	0,114855	0,324803	0,420346	
0,011377	0,670807	1	0,051337	0,006655	0,002918	
0,055959	0,12151	1	0,01104	0,18217	0,001057	
	0,509944	1	0,020358	0,067863	0,055425	
0,047265	0,216531	0,931887	0,183156	0,279242	0,263152	
0,031161	0,151686	0,584709	0,030992	0,412529	1	
0,045854	0,286705	0,570713	0,144049	0,451832	0,262211	

Переходим к столбцу F2. Также применяем сортировку по убыванию, получаем несколько единиц вверху и также как и ранее добавляем пустую строку после них. Так «лесенкой» двинемся вниз до столбца 26 включительно.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

В	BC	BD	BE	BF	BG	BH	BI	BJ
38836	0,146628	1	0,074009	0,111705	0,125834	0,098858	0,041358	0,251588
19474	0,145877	1	0,123681	0,059517	0,338471	0,014528	0,100317	0,078710
31097	0,140768	1	0,035074	0,071251	0,053941	0,079088	0,062314	0,037188
71427	0,07502	1	0,101889	0,097179	0,161017	0,077052	0,041047	0,362258
5831	0,069386	1	0,118561	0,035445	0,08059	0,061641	0,163957	0,213328
31688	0,061015	1	0,051529	0,033032	0,019573	0,075725	0,056713	0,068188
23666	0,059183	1	0,52874	0,12704	0,245161	0,036532	0,046604	0,095898
36952	0,058059	1	0,127602	0,084382	0,043239	0,048184	0,105731	0,215688
8108	0,042545	1	0,127062	0,048251	0,105309	0,006096	0,105189	0,009358
51437	0,015448	1	0,126897	0,04998	0,081195	0,026377	0,032483	0,073328
55588	0,003713	1	0,097417	0,037959	0,126158	0,036174	0,015208	0,123488
26879	0,08604	0,601286	1	0,360246	0,081575	0,119937	0,159975	0,127768
10917	0,1764	0,115207	1	0,186933	0,01954	0,145764	0,548406	0,171768
56694	0,04409	0,102836	1	0,055254	0,046564	0,080173	0,13117	0,155538
33859	0,005855	0,084566	1	0,040597	0,009577	0,060952	0,035255	0,129448
57105	0,055671	0,06693	1	0,014311	0,079937	0,003149	0,112153	0,245088
45882	0,027964	0,055557	1	0,157952	0,023278	0,042051	0,158853	0,055808
71069	0,068389	0,049455	1	0,056841	0,010484	0,030335	0,072008	0,072808
59447	0,129966	0,045221	1	0,128895	0,019815	0,017242	0,129976	0,036688
39321	0,002076	0,031683	1	0,021392	0,035949	0,014018	0,025976	0,099128
51686	0,584709	0,030992	0,412529	1	0,181578	0,18899	0,399154	0,153938
35153	0,412065	0,292032	0,209718	1	0,175418	0,135997	0,014444	0,085858
8241	0,004722	0,059975	0,097679	1	0,070064	0,090037	0,12965	0,001368
33692	0,279767	0,098047	0,074925	1	0,062068	0,002284	0,073731	0,051958

Теперь, рекомендую закрепить самую первую строку матрицы. Так всегда будет видно с каким фактором ведётся работа. Также покрасим красным значения факторов как в примере, двигаясь лесенкой вниз.

Результат смотрим здесь: [готовые матрицы](#)

Пока всё. Посмотрите на то, что вышло в итоге, и, думаю, вы догадаетесь к чему всё идёт, и зачем мы проводили все эти манипуляции с данными.

Часть шестая. Интерпретация матриц факторного анализа и написание отчёта.

Это завершающая часть книги о контент-анализе в программе ЛЕКТА. Сейчас я расскажу вам, как работать с матрицами факторного анализа, получаемыми в ходе контент-анализа. Напомню ещё раз, что я описываю предыдущую версию программы, функционал которой несколько менее полон — в ней нет модуля факторного анализа, что немного осложняет работу. В новой версии факторизация происходит намного быстрее и удобнее.

Итак.

Структура будущей аналитической записки

Возможно, вы уже догадались, как работать с полученными матрицами. Напомню, что матрицы представлены факторами, а факторы — это тематические блоки, каждый из которых и предстоит описать. Учитывая то, что факторы представлены наиболее часто встречающимися лексемами, мы можем предположить, что и сами факторы описывают наиболее популярные (актуальные) в изучаемом информационном пространстве тематические блоки. Давайте просто посмотрим на нашу матрицу factor loadings. ([скачать её можно здесь](#))

Посмотрим на первый фактор (колонка F1 или столбец В в примере). Отсортируем значения, отмеченные красным, так, чтобы сверху оказались наибольшие по модулю значения. В данном случае сверху оказываться значение 0,74 . Как мы видим, все значения отрицательны. Условно будем называть фактор отрицательным. Однако, это не значит, что фактор носит какой-либо негативный оттенок. Если бы все значения были в плюсе, фактор бы можно было классифицировать как положительный, что опять же не говорит о его позитивности. Если бы были и отрицательные и положительные значения, фактор можно было бы назвать поляризованным. Поляризованный фактор представляет не один тематический блок, а два. Положительный или отрицательный факторы отражают только один тематический блок. (строго говоря, это не так — какие-то микротемы могут «привязаться» к фактору, расширяя его тематический диапазон).

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

		Получение внешних данных		Подключения			
B2		fx		-0,742838663500113			
	A	B	C	D	E	F	G
1		F1	F2	F3	F4	F5	F6
2	Минина	-0,742839	0,018468	0,005979	0,013595	0,020688	0,020688
3	Пожарского	-0,670807	0,034437	-0,004464	-0,001957	0,022342	0,022342
4	Чкалову	-0,509944	0,010382	0,034606	0,028263	-0,000674	-0,000674
5	памятник	-0,504425	0,038168	0,061378	-0,088642	0,006854	0,006854
6	Площадь	-0,441047	-0,050656	-0,143253	-0,185392	-0,030707	-0,030707
7	Горький	-0,361246	0,043061	0,007211	-0,046919	0,015557	0,015557
8	горького	-0,332292	-0,098201	-0,009444	-0,042670	0,023724	0,023724
9	лестница	-0,332002	0,010134	0,034503	0,043125	-0,010663	-0,010663
10	история	-0,229643	0,084817	0,067894	-0,069770	0,031887	0,031887
11	ГАЗ	-0,220127	-0,012301	-0,130130	0,017158	-0,045039	-0,045039
12	завод	-0,207820	0,050303	-0,150602	0,056144	0,039568	0,039568
13	войны	-0,160909	0,055223	0,028662	-0,010477	0,019233	0,019233
14	связи	-0,121510	-0,001341	-0,022135	-0,000128	0,014056	0,014056
15							
16	молодежи	0,031301	0,107965	-0,033760	0,076637	0,002221	0,002221
17	такси	0,053831	-0,187365	0,033016	0,032667	0,008645	0,008645
18	пешком	-0,045286	-0,260032	-0,011972	0,009677	0,031761	0,031761
19	транспорта	0,073143	-0,498836	-0,036918	0,055722	0,062770	0,062770

Как мы видим, фактор представлен следующим набором переменных:

Минина
Пожарског

o

Чкалову
памятник
Площадь
Горький
горького
лестница
история
ГАЗ
завод
войны
связи

Лексемы, имеющие наибольший коэффициент играют большую роль в факторе и лучше его описывают. Лексемы, имеющие малый коэффициент, часто можно просто исключить, так как они описывают периферийные темы, а не тематические доминанты. Просто посмотрев на набор лексем, мы можем сделать вывод, что первый фактор

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

описывает достопримечательности Нижнего Новгорода, главные исторические улицы, памятники. Не всем знакомы достопримечательности Нижнего Новгорода, поэтому давайте посмотрим на любой другой фактор.

Далеко ходить не будем. Давайте взглянем на второй фактор.

Режимы просмотра книги		Показать			
C29		fx 0,107964972899182			
	A	B	C	D	
1		F1	F2	F3	F4
3	войны	-0,160909	0,055223	0,028662	-0,01
4	связи	-0,121510	-0,001341	-0,022135	-0,00
5					
6	автобусы	0,009491	-0,614370	0,077962	0,03
7	маршрутки	0,035492	-0,581688	0,029974	0,01
8	транспорта	0,073143	-0,498836	-0,036918	0,05
9	проезд	-0,001729	-0,465588	0,045356	0,01
10	добираться	-0,031628	-0,455831	-0,054044	0,01
11	общественном	0,060685	-0,431097	-0,015120	0,03
12	езжу	0,061192	-0,419474	0,051881	0,02
13	пересадками	-0,017363	-0,408108	0,051855	0,01
14	остановки	-0,022466	-0,386952	-0,049376	0,03
15	работу	0,019155	-0,323666	-0,171135	0,04
16	утром	0,020362	-0,271427	0,027656	0,02
17	пешком	-0,045286	-0,260032	-0,011972	0,00
18	такси	0,053831	-0,187365	0,033016	0,03
19	молодежи	0,031301	0,107965	-0,033760	0,01

Снова сортируем значения. Вверху оказываются наиболее значимые переменные.

автобусы
 маршрутки
 транспорта
 проезд
 добираться
 обществен
 ном
 езжу
 пересадкам
 и
 остановки
 работу
 утром

пешком
такси
молодежи

Легко понять, что фактор описывает функционирование транспортной системы Нижнего Новгорода. Фактор отрицательный. Однако, мы видим, что в факторе присутствует одна переменная со знаком +. Это лексема молодёжь. Исключаем её — сама по себе она нас мало интересует. Лексемы представляют интерес при наличии выраженной в факторе корреляции. По одиночке рекомендую их игнорировать.

Как мы поступаем далее? Всё довольно просто. Каждому из факторов мы даём название. Например, второму можно дать название «функционирование транспортной инфраструктуры» или что-то похожее. Дав название каждому из факторов, подумаем, как можно их сгруппировать в более крупные блоки. Можно назвать их темами. В итоге получим 3-7 тем.

Тема 1

фактор a
фактор b
фактор c

Тема 2

фактор d
фактор e
фактор f

Тема 3

фактор g
фактор h
фактор i

Каждой из тем также даём название. Поясню на всякий случай — факторы группируются исходя из тематичности, а не их порядковых номеров. То есть мы просто находим более общие темы и интегрируем факторы в крупные блоки.

Сама эта структура может стать основой аналитической записки. Тема = глава. Фактор = параграф. Факторы, разумеется, в случае их схожести также можно объединять между собой.

Аналитическая записка по результатам контент-анализа

Что же мы делаем дальше... Набор тематических групп, представляющих

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

наибольшую популярность в информационном пространстве, выраженном массивом текстов, нас, безусловно, интересует. Однако он даёт сравнительно мало информации. Эта информация сугубо количественная. Мы же планируем перейти к качественному этапу интерпретации текстов.

Допустим, мы продолжаем работу со вторым фактором, описывающим транспортную инфраструктуру Нижнего Новгорода. Открываем текстовый редактор, например, MS Word. Вставляем туда группу лексем второго фактора.

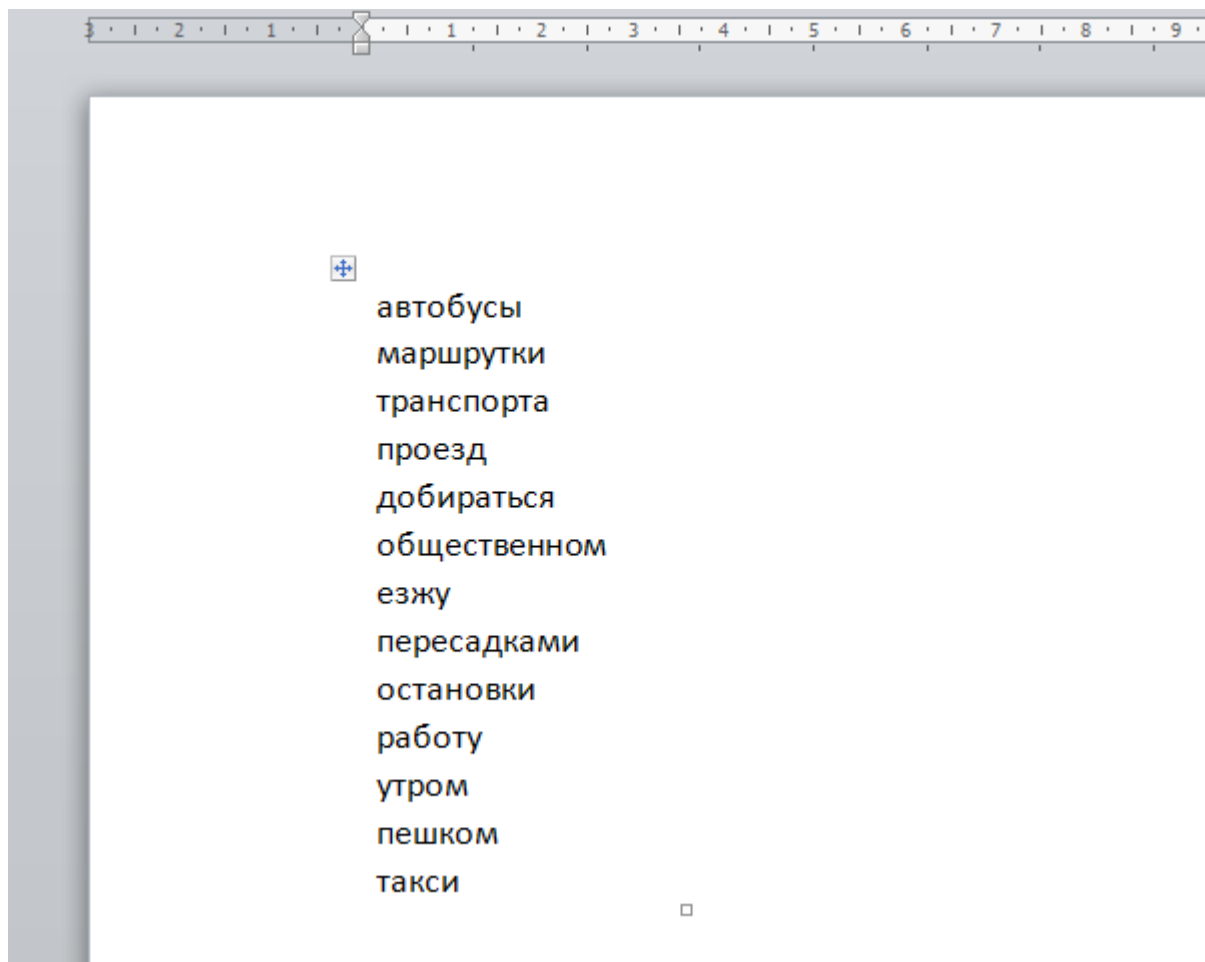
автобусы
маршрутки
транспорта
проезд
добираться
обществен

ном

езжу
пересадкам

и

остановки
работу
утром
пешком
такси



В голове держим мысль о том, что фактор отрицательный. Открываем вторую матрицу — Factor Scores — находим столбец F2, сортируем значения так, чтобы вверху оказались отрицательные значения, то есть по возрастанию. Опускаемся вниз до значений по модулю в столбце не менее 1.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

	A	B	C	D	E	F	G	H	I	J	
	##	Фрагмент	Слов	F1	F2	F3	F4	F5	F6	F7	F8
27	462	3KTZ43M.t	107	-0,2148	-1,0371	-1,37348	-0,2673	0,3293	0,1180	-0,56970	
28	456	3KTZ41M.t	71	0,0922	-1,0369	-3,59674	0,2210	0,3322	0,2879	0,52167	
29	448	3KTZ35W.t	104	0,0767	-1,0223	-0,92265	0,0405	-0,0021	-2,4043	-0,31510	
30	2269	zol23w.tx	66	0,1791	-1,0206	0,71468	0,3406	0,1241	-0,0211	0,19966	
31	776	3Pntl.W.6	95	-0,1082	-1,0197	1,25046	-0,0922	0,0168	-0,5028	0,35614	
32	1584	mar21w.t)	65	0,3603	-1,0110	-0,07194	0,2990	0,1927	0,1241	-0,30110	
33	1605	mar24b1.t	75	0,0431	-1,0095	-1,19163	0,0897	0,4582	0,0574	0,64794	
34	2285	zol35w.tx	76	0,5482	-0,9922	0,89219	-0,2081	0,5338	0,9714	-0,33592	
35	64	3BLTKV62l	108	0,1513	-0,9875	0,2922	-0,0003	-1,7127	0,2604	0,23666	
36	18	3Agr43w.	95	0,3317	-0,9863	-0,04917	0,3051	0,4960	0,2156	0,06243	
37	1319	kac56m.tx	67	0,5475	-0,9811	0,11655	-1,7019	-0,3173	0,0769	-0,30156	
38	433	3KLN64W.	87	0,6836	-0,9770	0,34983	-0,1178	0,4060	-2,4739	-0,68133	
39	34	3BLTKV60'	110	0,2082	-0,9711	0,43268	0,1363	-0,0296	0,0976	0,43522	
40	1555	lez36b.txt	65	0,3990	-0,9696	0,07996	0,3017	0,1661	0,3089	-0,01852	
41	2273	zol23w.tx	85	0,4255	-0,9683	-0,26370	0,3087	0,2394	-0,0549	0,48305	

Напомню — последний столбец этой матрицы — фрагменты текстов. Соответственно при сортировке и его содержимое также отсортировалось. Чтобы было понятнее, скажу по-проще: нас интересует сильно нагруженные темой транспорта фрагменты текстов. В них будут часто встречаться вместе лексемы, представляющие фактор — то есть те лексемы, которые содержатся в первой матрице. Фрагменты текстов, имеющие коэффициент менее 1 будут содержать мало этих слов, соответственно разговор в них будет идти уже на смежные или иные темы. Нам же при работе с анализом фактора о транспорте, имеет смысл анализировать текстовые фрагменты только о транспорте — ничего лишнего нам не нужно. Итак, двигаемся в конец второй матрицы и копируем тексты, имеющие коэффициент как минимум 1. Чтобы было понятнее, я скрою почти все столбцы, оставив последний и несколько столбцов факторов.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

A	B	C	D	E	F	AB	AC	AD	AE	AF	AG	AH
	Фрагмент	Слов	F1	F2	F3	F25	F26	%				
2154	VOR 21W.	108	-0,0127	-1,1756	-0,42595	-0,18622	2,0800	21	Девушка 21 год, студентка заочной форм			
44	3BLTKV60	91	0,0693	-1,1598	-0,34724	1,41108	0,6754	17	Там немного поменьше цены. Там чего-н			
1292	kac22w.tx	68	0,0322	-1,1560	-1,31434	-0,30324	1,3434	17	- с того момента как себя помню, когда б			
829	3RBKN27V	100	0,0793	-1,1535	-0,04299	-0,01753	-0,0730	5	Пошли какие-то бараки, еще чего-то, я е			
95	3BLTKV63	76	0,2890	-1,1498	0,17995	-0,46688	2,2602	10	И вот в пятницу вечером мы уезжаем, в г			
833	3RBKN27V	61	-0,3176	-1,1455	0,55736	0,45088	-0,3213	16	Те, которые живут в заречной части, у ни			
12	3Agp33M.	102	-0,4996	-1,1347	0,02569	0,33496	0,7584	13	И.: Можете привести некоторые пример			
1340	kol20m.tx	142	-6,7412	-1,1279	-0,52474	0,89765	1,6009	26	Даже аквапарка того же самого, куда све			
2284	zol35w.tx	62	0,2037	-1,1076	-0,18940	0,77529	0,1684	16	Бывает правда, что и пьяные могут полез			
1563	lez52w.tx	103	-0,2117	-1,1026	0,36331	1,62232	-5,3209	17	Ходить куда то — желания нет, да и возра			
1705	pan47w.tx	57	-0,0565	-1,0897	-0,45524	-0,38896	-0,8358	26	А по большому счету, район этот неплох			
1837	sol27w.tx	101	-0,0108	-1,0891	0,22093	-0,54485	0,9621	16	##SL А ваши братья живут с вами? Один			
1315	kac56m.tx	75	0,0936	-1,0639	-1,59950	-0,15194	0,6762	28	Расскажите немного о себе Меня зовут В			
395	3IGNT75M	86	0,2117	-1,0531	0,63787	0,80520	-1,7575	13	- Да, конечно. Параж находится недале			
1469	koz21w.tx	55	0,2390	-1,0525	-0,22336	-0,54896	-1,1097	27	Чтобы спокойно было и тихо. Можешь сн			
147	3BLTKV72	127	0,3584	-1,0481	-0,52570	1,91906	-0,2290	12	Не только город, страна не готова к экстр			
462	3KTZ43M.	107	-0,2148	-1,0371	-1,37348	-0,98906	0,6520	19	##SL - Где географически вы работаете? -			
456	3KTZ41M.	71	0,0922	-1,0369	-3,59674	0,59050	1,6926	36	Алексей, 41 - В каком районе вы прожива			
448	3KTZ35W.	104	0,0767	-1,0223	-0,92265	-0,84833	1,4369	31	Да и мне тоже удобно, все места, котор			
2269	zol23w.tx	66	0,1791	-1,0206	0,71468	-0,39315	-0,4263	7	Да и к тому же у нас был специальный пр			
776	3Pntl.W.6	95	-0,1082	-1,0197	1,25046	2,79496	-2,7207	15	Женщина, 67 лет, пенсионерка. Интервь			
1584	mar21w.tx	65	0,3603	-1,0110	-0,07194	-1,19657	-0,3627	10	В выходные я частенько подрабатываю, т			
1605	mar24b1.t	75	-0,0431	-1,0095	-1,19163	0,76641	0,9406	33	Хотя дороги -это проблема всего города,			
2285	zol35w.tx	76	0,5482	-0,9922	0,85219	2,77669	0,2194	21	Но есть один минус по этой причине нел			

В общем, копируем все текстовые фрагменты до самого верха. Вставляем их в текстовый редактор ниже фрагментов текстов.

Теперь нам надо отметить лексемы фактора в тексте и желательно разными цветами. Соответственно через функцию поиск (Ctrl F) раскрашиваем лексемы, а точнее их основные части без окончаний, так как наши лексемы, как Вы помните — фильтры, включающие в себя наборы словоформ.

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

+

Р.: Ну, лично для меня ездить на социальном транспорте гораздо выгоднее, проезд в маршрутках довольно дорогой, а иногда приходится ездить с пересадками и по несколько раз в день в разные места. Что касается качества, то битком сейчас не только автобусы, но и маршрутки. Это уже привычное явление. В чём преимущество маршруток, так это в скорости. Автобусы ездят медленнее и соответственно дольше. И.: Может быть просто сейчас недостаточно общественного транспорта? И количество автобусов и маршруток меньше, чем нужно, если они не справляются с таким количеством пассажиров? Как вы считаете, если общественного транспорта станет больше, каким образом это повлияет на ваше передвижение и на ситуацию дорог Нижнего Новгорода в целом?

Жизнь в городе: проблема удобства и комфорта. Респондент – мужчина, 37 лет, не женат, высшее образование. ВОПРОС: Расскажите, пожалуйста, как Вы добираетесь на работу. ОТВЕТ: На общественном транспорте. ВОПРОС: Нравится? ОТВЕТ: Ну, в общем, да. Я проездной не покупаю, езжу на первом, что подойдет, поскольку если проездной покупаешь, то себя чем-то связываешь, особенно в нашем городе, где не существует единого проездного на электротранспорт и автобус. То есть тут есть разные версии, почему этого нет, вплоть до обвинений власти в неких намеренных злодеяниях. Но вот, тем не менее, факт – такого проездного нет. К тому же есть маршрутки, на которые в принципе проездного нет, поэтому... ВОПРОС: То есть, Вы можете доехать от дома до работы на любом виде транспорта? Во дворе не гуляю. - ##SL - Как ты добираться на учебу? - Чаще всего добираюсь на автобусе, но когда опаздываю еду на маршрутке. От моего дома до учебы не идет прямой автобус, поэтому приходится ездить с пересадками. Очень удобно ездить с проездным, он подстраховывает, потому что даже когда нет денег, ты можешь уехать домой, но с ним ограничены маршруты, поэтому приходится несколько раз делать пересадки. Хотя автобусы в мою сторону не так часто ходят. Да и народу в автобусах бывает иногда так много, что даже не влезешь туда. Когда куда-нибудь едешь, всегда есть пробки: на Пролетарской и на мосту. На мосту бывает и не один раз в пробке постоишь.

Ну, дальше всё просто. Дальше идёт качественная интерпретация — описываем, как конкретно преподносится каждый из тематических блоков, отвечая на вопросы КАК, ПОЧЕМУ, ЗАЧЕМ и т.д. Оцениваем тексты по заданному набору параметров — оцениваем мотивации, эмоции и что угодно. Можем сгруппировать тексты с точки зрения оценки предмета обсуждения фактора — например, закрасив положительные высказывания жёлтым, нейтральные серым, негативные красным. Далее их можно сгруппировать, перетащив в единые блоки и работать с ними по очереди, также определив, какие оценки доминируют в пространстве фактора, подсчитав соотношение оценок. Интерпретация качественного характера уже была описана на блоге. Если не хотите сильно заморачиваться — просто обобщайте материал каждого из факторов, передавайте в записке основные мысли, эмоциональные особенности их преподнесения, оценки обсуждаемого предмета авторами текстов, пишите выводы в конце параграфа.

На этом про ЛЕКТА всё. Кстати, если уж речь идёт о качественной интерпретации, то хочу отметить — следующих цикл статей на блоге посвящён программе ATLAS.TI — мощном инструменте качественного контент-анализа. Соответственно обе программы можно использовать вместе на разных этапах для получения более качественного результата.

Спасибо и удачи с освоением метода контент-анализа! Всегда буду рад помочь вам в этом отношении — пишите на мой блог. Адрес внизу.

Всего доброго!

УСЛУГИ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА ИНТЕРВЬЮ, ФОКУС ГРУПП, СМИ, ФОРУМОВ, БЛОГОВ. ОБУЧЕНИЕ МЕТОДИКАМ ПРОВЕДЕНИЯ КОНТЕНТ-АНАЛИЗА.

P.S. Скоро на блоге появится и видео курс по ЛЕКТА, качественным методам, программе ATLAS.ti + книга по ней и многое другое! Следите за обновлениями на блоге о контент-анализе!