Пермский филиал федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики» Факультет экономики, менеджмента и бизнес-информатики

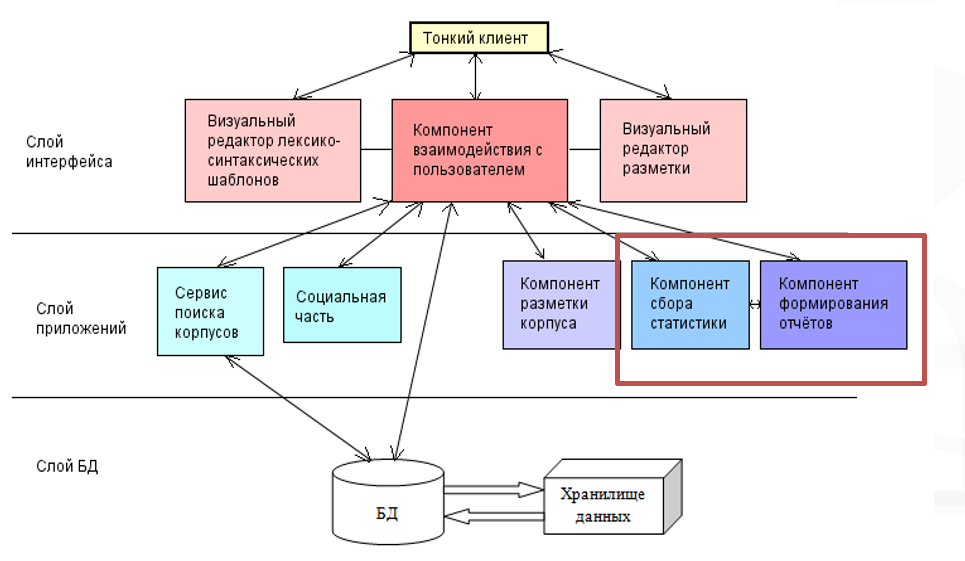
Горбунова Мария Андреевна

Доклад на тему

«СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ ДОКУМЕНТОВ КОРПУСА»

Пермь, 2017 год

В докладе «Разработка системы сбора статистики использования маркеров стиля в научных публикациях на английском языке» был описан процесс проектирования и разработки двух модулей. На Рис.1 можно увидеть какое место данные модули занимают в разрабатываемой системе Paper Cat.



**Рисунок 1. Структура системы Paper Cat**

Проведение анализа текста без использования компьютерных инструментальных средств – это достаточно затратный по времени и силам процесс, поэтому необходимо использовать программные инструментальные средства. На данный момент существует большое количество инструментов для обработки корпусов текстов.

Целью данной работы является демонстрация результатов работы компонентов системы, а именно предоставление отчетов о статистики собранной посредством модулей Paper Cat.

На данный момент реализовано огромное количество систем по обработке естественного языка. Для выяснения необходимого функционала для разрабатываемого программного модуля, необходимо провести обзор систем-аналогов, выявить существующие средства для сбора статистики и генерации отчетов, и каким образом они реализованы.

Для сравнительного анализа систем были сформулированы следующие критерии:

1. Возможность генерации отчетов.

2. Поддержка различных форматов отчетов.

3. Список параметров, заносимых в отчет.

4. Формирование пользовательских отчетов.

5. Возможность расширения функционала, а именно встраивания дополнительных программных модулей в систему.

6. Язык разработки.

Проанализируем наиболее популярные системы обработки естественного языка (Табл. 1).

**Таблица 1. Сравнение средств обработки языка**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Название системы** | **Генерация отчета** | **Форматы экспорта** | **Параметры отчета** | **Пользовательские отчеты** | **Расширение функционала** | **Язык разработки** |
| AntConc | + | + (.txt, .html, .xml, .ant) | Hit, KWIC,File | - | + | Perl |
| WordSmith Tools | + | + (.txt, .xml, .xls, .rtf, WordList files) | Word, Frequency, Texts, Lemmas, Set | - | - | Python |
| Sketch Engine | - | + (.txt, .xml) | Word, Co-occurrence count, Candidate count | - | + | C++, Python, JavaScript, Jquery |
| Gate Developer | - | + (.xml) | Type, Set, Start, End, ID, Features | - | + | Java |
| CQPWeb | + | + (.txt) | No, Search Results, No of occ., percent | - | + | Perl, R, PHP |

По итогам анализа систем обработки естественного языка, можно сделать ряд выводов. Ни одна из рассмотренных систем не позволяет пользователю вносить собственные параметры в отчет. Некоторые из систем не предоставляют возможности генерировать отчеты, также, как и не во всех реализована функция сбора статистики. В большинстве из рассмотренных систем отчет представляет собой файл в формате «\*.txt», с ограниченным набором параметров. Данный формат является универсальным для дальнейшей обработки, но не обеспечивает пользователю четкого графического отчета о произведенной обработке текста. Одна из систем, WordSmith Tools, позволяет отобразить отчет посредством MS Excel, что является преимуществом перед другими инструментами.

Таким образом, стоит отметить, что возможность генерировать пользовательские отчеты с графическим отображением полученных данных не реализована ни в одной из рассмотренных систем, тем не менее, подобная функция является полезной при представлении результатов обработки текста.

Проанализировав средства для обработки языка было выявлено, что ни одно из них, включая Gate Developer, не предоставляет пользователю возможность сгенерировать отчет. Данный факт является показателем того что системы недостаточно доработаны, так как отсутствие генерации отчетов — это существенный недостаток. Для того чтобы построить программный модуль, который выполнял бы такую функцию, необходимо просмотреть решения, которые существуют на рынке в данный момент и выбрать подходящую систему. Определим, что является генератором отчетов.

Генератор отчетов – это библиотека или программа, позволяющая представить данные в структурированном виде, удобном для чтения и печати [8]. Чаще всего для пользователя достаточно затруднительно, после завершения работы с какими-либо данными, формировать отчет вручную. Инструменты для генерации отчетов значительно облегчают задачу.

Стоит отметить, что генераторы отчетов нужны не только пользователям, но и разработчикам, так как система, оснащенная функцией выведения отчетности, имеет некоторое преимущество перед аналогами.

Для дальнейшей разработки программного модуля необходимо проанализировать имеющиеся на рынке средства генерации отчетов и выбрать тот, который максимально удовлетворяет требованиям. Были выбраны 5 генераторов отчетов: Crystal Reports, StimulSoft Reports, Jasper Report, Next Report, BIRT. Для оценивания каждого из инструментов были разработаны следующие критерии:

1. Язык внедрения.

2. Возможность создания диаграмм.

3. Форматы документов отчетов, выгружаемых пользователю.

4. Возможность выбора визуального оформления отчета (шрифты, цветовое оформление, отступы и т.д.).

5. Стоимость программы.

6. Открытый код.

7. Возможность работы с базами данных.

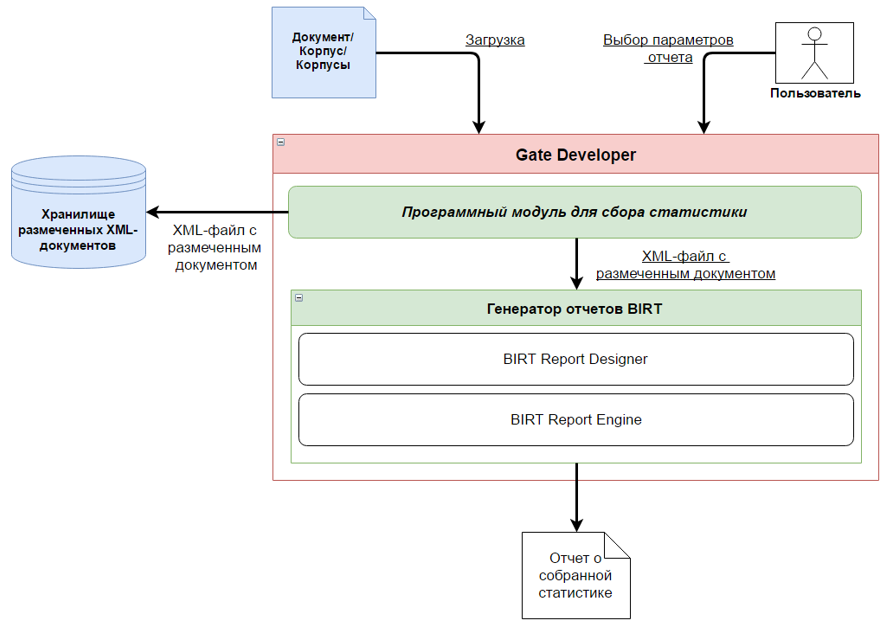
Результаты сравнения генераторов отчетов представлены ниже (Табл. 2).

**Таблица 2. Сравнение генераторов отчетов**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Язык внедрения** | **Возможность создания диаграмм** | **Форматы отчетов, выгружаемых пользователю** | **Визуальное оформление** | **Стоимость** | **Открытый код** | **Возможность работы с базами данных** |
| **Crystal Reports** | Visual Basic, C++,C#, Java и т.д. | + | RPT, RPTR, HTML, MS Word, MS Excel, ODBC, PDF, RTF, CSV, TTX, TXT, XML | + | ~ 75 000 руб. | - | + |
| **StimlSoft Reports** | JavaScript, HTML5, PHP, Java, и Flex | + | PDF, XPS, MS Power Point, HTML, MHT, TXT, RTF, MS Word, MS Excel, CSV, DBF, XML | + | ~ 29 000 руб. | + (при определенных условиях покупки) | + |
| **Jasper Report** | Java | + | PDF, RTF, HTML, XLS, CSV и XML | + | Бесплатно | + | + |
| **Next Report** | Java | + | HTML, EXCEL, EXCELX, PDF, DOCX, RTF,  CSV, TSV, TEXT и XML. | + | Бесплатно | - | + |
| **BIRT** | Java | + | HTML, PDF, Excel, Power Point,  Word | + | Бесплатно | + | + |

Таким образом, среди всех рассмотренных средств генерации отчетов, наиболее подходящим для решения поставленной задачи является BIRT. Данный инструмент поддерживает внедрение с помощью языка программирования Java, позволяет визуализировать данные в виде диаграмм, а также выбрать параметры отображения отчета. Также выбранное средство предоставляется бесплатно, что является важным критерием при академических исследованиях.

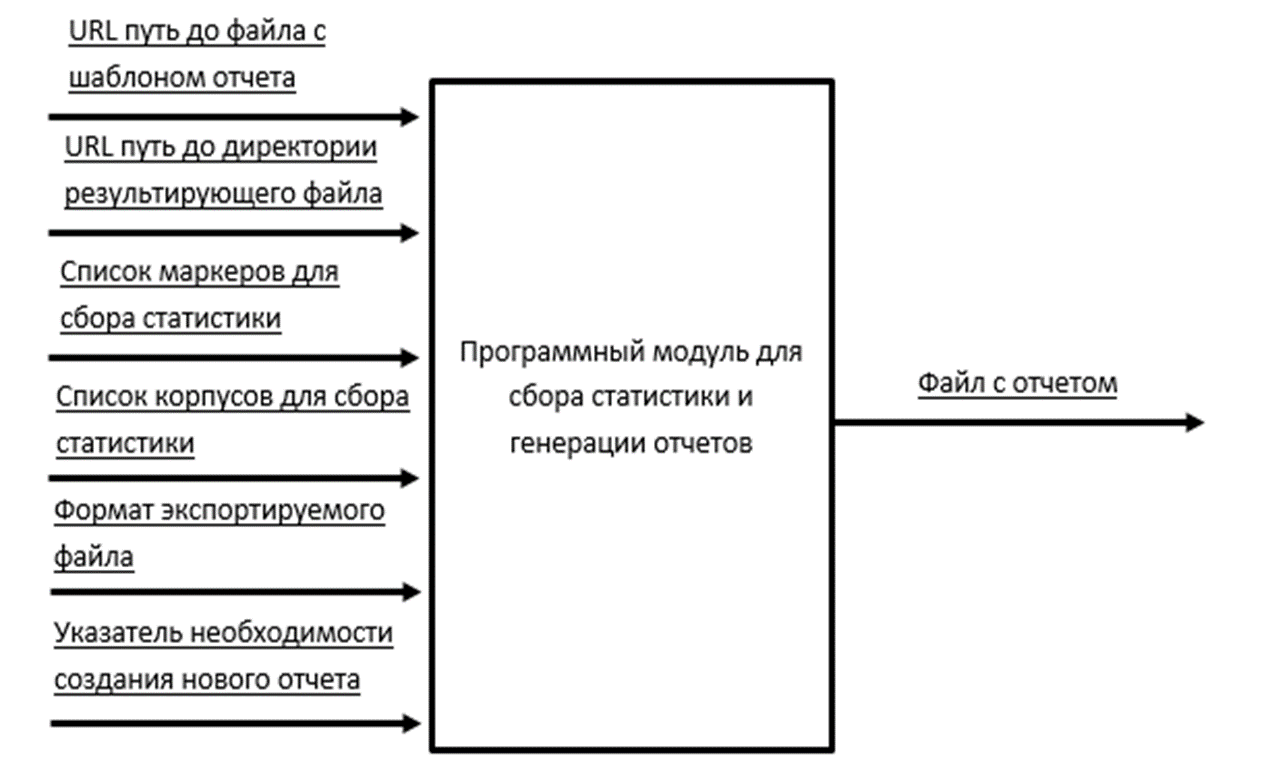
Перед тем как приступить к непосредственной разработке программного модуля, необходимо спроектировать архитектуру программного модуля (Рис. 2).



**Рисунок 2. Архитектура программного модуля**

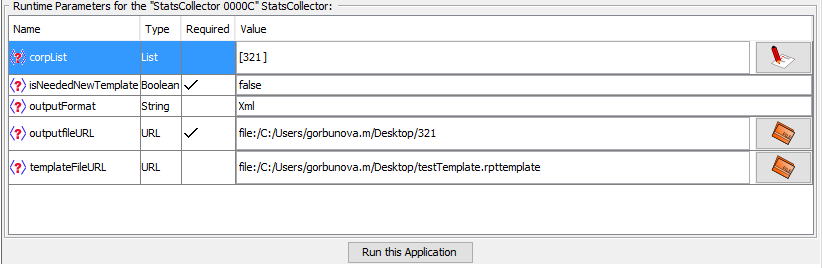
Как можно увидеть, пользователю необходимо загрузить документ и выбрать параметры отчета, после чего Gate Developer предоставляет размеченный документ в формате XML. Данный файл помещается в хранилище документов (на данный момент папка на локальном компьютере пользователя) и затем переходит в генератор отчета, где он будет использоваться в качестве источника данных. В соответствии с загруженным шаблоном отчета, BIRT генерирует отчет и выгружает его в директорию, указанную пользователем.

Так как плагин должен генерировать отчет на основе шаблона и экспортировать готовый отчет, в качестве входных параметров необходимо получать путь до файла с шаблоном отчета и путь до директории, в которой необходимо сохранить отчет. Схема передачи параметров изображена на рис. 1.

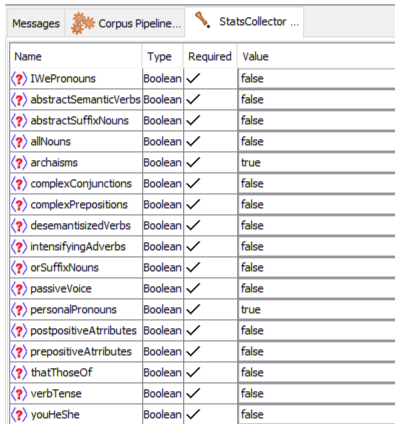


**Рисунок 3. Параметры входа/выхода**

Полный листинг добавления параметров можно увидеть в приложении Б. В результате выполнения плагина, в окне редактора параметров обрабатывающего ресурса отображаются добавленные параметры (рис. 2) и список необходимых маркеров (рис. 3).



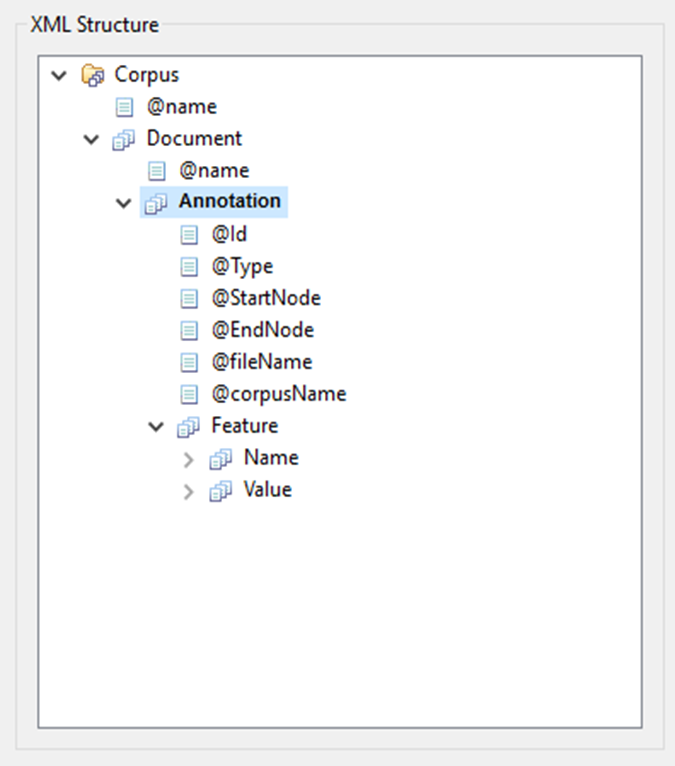
**Рисунок 4. Параметры ввода**



**Рисунок 5. Список маркеров**

Ранее было сказано, что в системе Gate Developer уже существует плагин для сбора статистических данных. На данный момент StatsCollector, программный компонент по сбору статистики, экспортирует результат работы в виде файла в формате \*.txt. Данный формат не удобен для дальнейшей обработки, к тому же результирующий файл не обладает необходимой для этого структурой. Таким образом, в плагине необходимо реализовать алгоритм, который бы выгружал аннотированные тексты в структурированную форму (XML) и передавал их в генератор отчетов, в качестве источника данных. В зависимости от значения параметра corpList, плагин формирует .xml документ с данными корпуса или же документ с данными из нескольких корпусов, с помощью функции getcorp(). В качестве входного параметра getcorp() принимает название корпуса, и формирует строку с данными. Для работы с XML и его обработкой, была использована библиотека DOM4J. В случае, если плагин применяется к нескольким корпусам, метод применяется к каждому из них и формирует общую строку для набора корпусов.

Размеченные корпуса в формате XML после обработки обладают структурой, изображенной на рис. 4.



**Рисунок 6. Структура документа**

Для того чтобы сгенерировать отчет, пользователю необходимо передать в плагин URL адрес расположения шаблона отчета. В текущем контексте шаблон рассматривается как спроектированная структура отчета. Для создания и редактирования шаблонов отчетов существует независимый от среды разработки Eclipse интерфейс, к которому можно получить доступ посредством введения значения true во входной параметр isNeededDrawTemplate. После введения данного параметра необходимо нажать кнопку “Run Application”, после чего пользователь видит окно.

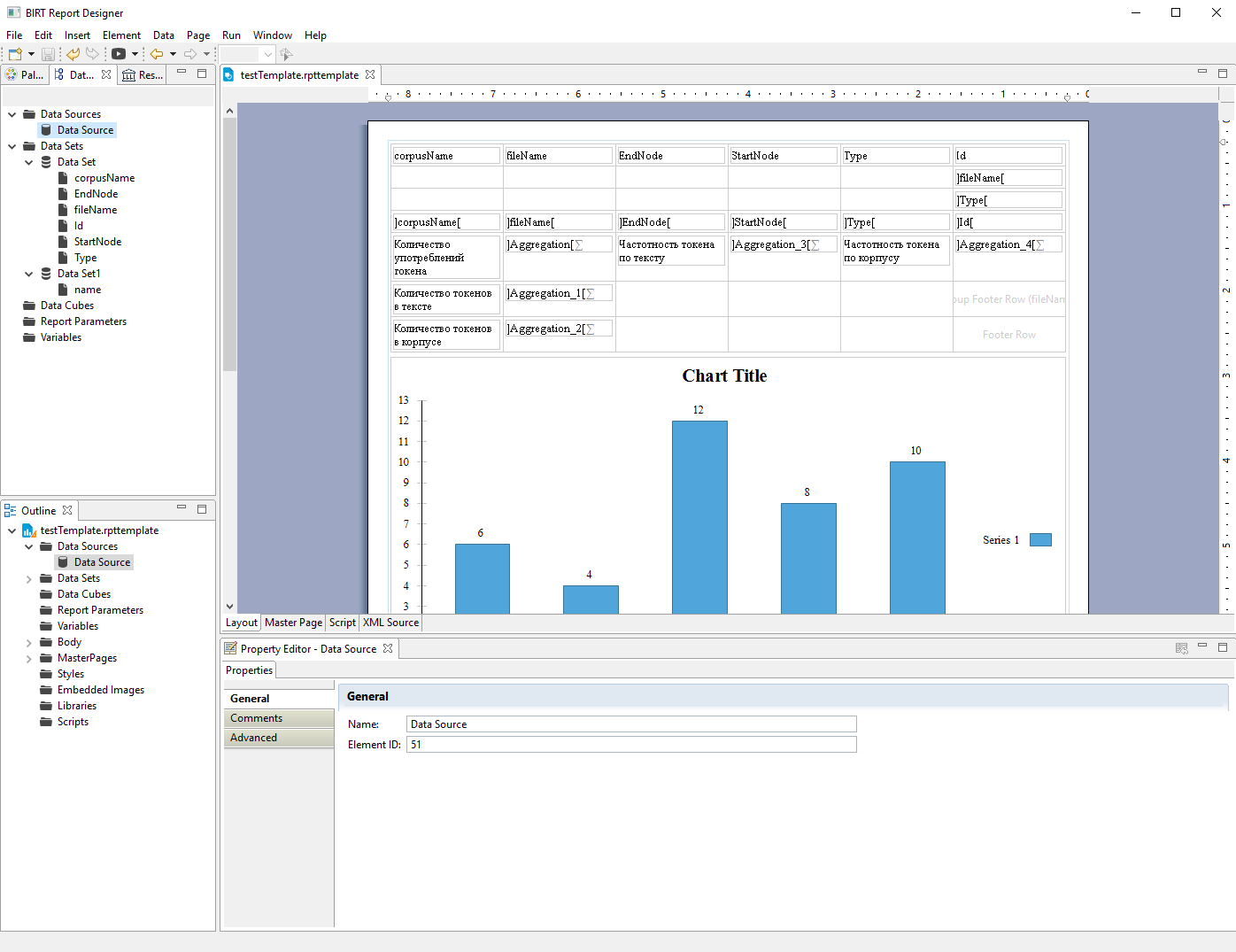
С помощью интерфейса пользователь может добавлять элементы из палитры и таким образом заполнять отчет. Так как на основе этого шаблона в дальнейшем будет генерироваться плагин необходимо, чтобы в нем была указана структура данных. Рассмотрим добавление данных на примере вышеуказанного xml-документа.

Сначала необходимо создать шаблон (.rpttemplate) или дизайн(.rptdesign),в данном случае это будет шаблон. После этого перейти на вкладку Data Explorer и с помощью контекстного меню вызвать окно создания источника данных.

Далее необходимо последовательно выбрать «XML Data Source», а затем указать путь до самого источника данных. Схему данных программа определяет автоматически.

Далее необходимо создать DataSet, так же как для предыдущего действия, для этого необходимо воспользоваться контекстным меню. После того как заданы название и другие настройки набора данных, необходимо создать структуру колонок, выбрав их из структуры xml-файла.

После добавления данных пользователь может выполнять действия с ними, а также использовать различные элементы для создания шаблона. В конце работы необходимо сохранить шаблон в формате \*.rpttemplate (или \*.rptdesign). Указанные процедуры являются необходимыми, так как структура данных, включенная в шаблон, позволяет загрузить результаты вычисления статистики в отчет и отобразить их, так как было задано пользователем.



**Рисунок 7. Пример шаблона отчета**

На рис.5 представлен пример шаблона. На нем мы можем увидеть функцию Aggregation, которая позволяет выполнять стандартные математические функции, а также функции математической статистики, что является применимым к данной работе.

По решению экспертов, для анализа были выбраны следующие корпуса текстов:

* Competent
  + - Competent BI PE
    - Competent Law
    - Competent Politology
    - Competent History
* Learners
  + - Learners BI PE
    - Learners Law
    - Learners Politology
    - Learners History

Условно корпуса разделены на два блока – блок академических работ и блок работ студентов. На момент сбора статистики были представлены тексты работ по 4 направлениям – Бизнес-информатика, Юриспруденция, История и Политология.

По запросу экспертов были собраны 3 вида показателей для каждого корпуса:

* Мода
* Медиана
* Среднее нормализованное значение (на 1000 слов)

Для отображения полной картины, показатели были собраны по всем маркерам в каждом корпусе, а именно:

* Abstract semantic verbs
* Archaisms
* Complex conjunctions
* Desemantisized verbs
* Future
* IPronoun
* Intensifying adverbs
* Logic connectors
* Noun
* OrSuffix
* PassiveVoice
* Past
* PersonalPronoun
* Prepositive attributes
* Postpositive attributes
* Present
* ThatThose
* WePronoun
* YouHeShe

Ниже приведены показатели для каждого направления – Бизнес-информатика (Табл. 3), Юриспруденция (Табл.4), История (Табл.5) и Политология (Табл.6).

**Таблица 3. Показатели корпусов направления Бизнес-информатика**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Маркер** | **Competent** | | | | **Learners** | | | |
| **Мода** | **Медиана** | **Среднее значение (на 1000 слов)** | **Мода** | | **Медиана** | **Среднее значение (на 1000 слов)** |
| **Abstract semantic verbs** | 0,01 | 0,031 | 0,04136 | 0,134 | | 0,068 | 0,07568 |
| **Archaisms** | 0 | 0 | 0,0000482 | 0 | | 0 | 0,00026 |
| **Complex conjunctions** | 0 | 0,004 | 0,00453 | 0 | | 0,006 | 0,00568 |
| **Desemantisized verbs** | 0,006 | 0,019 | 0,0281 | 0,01 | | 0,043 | 0,05042 |
| **Future** | 0 | 0,001 | 0,00136 | - | | 0,016 | 0,01937 |
| **IPronoun** | 0 | 0 | 0,00035 | 0 | | 0 | 0,00137 |
| **Intensifying adverbs** | 0 | 0,001 | 0,0017 | 0 | | 0,001 | 0,00137 |
| **Logic connectors** | 0,003 | 0,003 | 0,00449 | 0,002 | | 0,004 | 0,00537 |
| **Noun** | 0,834 | 0,37 | 0,483349 | - | | 0,645 | 0,765 |
| **OrSuffix** | 0 | 0,002 | 0,00422 | 0,01 | | 0,007 | 0,00637 |
| **PassiveVoice** | 0,002 | 0,013 | 0,02434 | - | | 0,061 | 0,05011 |
| **Past** | 0 | 0,012 | 0,02002 | 0,033 | | 0,017 | 0,01811 |
| **PersonalPronoun** | 0,005 | 0,017 | 0,01858 | - | | 0,016 | 0,02253 |
| **Prepositive attributes** | 0,433 | 0,177 | 0,22371 | - | | 0,284 | 0,33742 |
| **Postpositive attributes** | 0,151 | 0,073 | 0,09554 | 0,27 | | 0,104 | 0,14274 |
| **Present** | 0,029 | 0,065 | 0,07999 | - | | 0,101 | 0,12426 |
| **ThatThose** | 0 | 0 | 0,00022 | 0 | | 0 | 0,00005 |
| **WePronoun** | 0 | 0,01 | 0,01086 | 0,001 | | 0,003 | 0,00579 |
| **YouHeShe** | 0 | 0 | 0,00011 | 0 | | 0 | 0,00074 |

**Таблица 4. Показатели корпусов направления Юриспруденция**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Маркер** | **Competent** | | | | **Learners** | | |
| **Мода** | **Медиана** | **Среднее значение (на 1000 слов)** | **Мода** | | **Медиана** | **Среднее значение (на 1000 слов)** |
| **Abstract semantic verbs** | 0,092 | 0,0735 | 0,059857 | 0,079 | | 0,074 | 0,06788 |
| **Archaisms** | 0 | 0 | 0,000298 | 0 | | 0 | 0,000059 |
| **Complex conjunctions** | 0,002 | 0,008 | 0,009333 | 0,004 | | 0,006 | 0,00659 |
| **Desemantisized verbs** | 0,072 | 0,0545 | 0,046833 | 0,055 | | 0,052 | 0,04276 |
| **Future** | 0 | 0,003 | 0,003536 | 0,002 | | 0,005 | 0,008 |
| **IPronoun** | 0 | 0,005 | 0,006929 | 0 | | 0,001 | 0,00165 |
| **Intensifying adverbs** | 0 | 0,002 | 0,002 | 0 | | 0,001 | 0,001 |
| **Logic connectors** | 0,01 | 0,007 | 0,007369 | 0,006 | | 0,007 | 0,00641 |
| **Noun** | 0,895 | 0,7555 | 0,578214 | - | | 0,9 | 0,76147 |
| **OrSuffix** | 0 | 0,004 | 0,004131 | 0,021 | | 0,008 | 0,1024 |
| **PassiveVoice** | 0,003 | 0,02 | 0,019571 | 0,033 | | 0,033 | 0,02835 |
| **Past** | 0,002 | 0,0295 | 0,027202 | 0,01 | | 0,018 | 0,01647 |
| **PersonalPronoun** | 0,06 | 0,052 | 0,043012 | 0,024 | | 0,038 | 0,03394 |
| **Prepositive attributes** | 0,042 | 0,3145 | 0,243179 | 0,391 | | 0,361 | 0,30912 |
| **Postpositive attributes** | 0,21 | 0,1745 | 0,131167 | 0,211 | | 0,209 | 0,182 |
| **Present** | 0,173 | 0,127 | 0,10356 | 0,136 | | 0,134 | 0,11359 |
| **ThatThose** | 0 | 0 | 0,000524 | 0 | | 0 | 0 |
| **WePronoun** | 0,001 | 0,009 | 0,00956 | 0,009 | | 0,016 | 0,1735 |
| **YouHeShe** | 0 | 0 | 0,000238 | 0 | | 0 | 0,001 |

**Таблица 5. Показатели корпусов направления История**

| **Маркер** | **Competent** | | | **Learners** | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Мода** | **Медиана** | **Среднее значение (на 1000 слов)** | **Мода** | **Медиана** | **Среднее значение (на 1000 слов)** |
| **Abstract semantic verbs** | 0,076 | 0,067 | 0,05712 | 0,118 | 0,118 | 0,0849 |
| **Archaisms** | 0 | 0 | 0,00013 | 0 | 0 | 0,00014 |
| **Complex conjunctions** | 0,014 | 0,014 | 0,013 | 0,004 | 0,008 | 0,00905 |
| **Desemantisized verbs** | 0,056 | 0,05 | 0,04467 | - | 0,078 | 0,06157 |
| **Future** | 0 | 0,1 | 0,00157 | 0,004 | 0,004 | 0,00729 |
| **IPronoun** | 0,004 | 0,004 | 0,00573 | 0 | 0,005 | 0,00638 |
| **Intensifying adverbs** | 0,001 | 0,001 | 0,00163 | 0,001 | 0,002 | 0,0021 |
| **Logic connectors** | 0,004 | 0,004 | 0,00455 | 0,007 | 0,009 | 0,00833 |
| **Noun** | 0,941 | 0,8635 | 0,68272 | - | 1,334 | 0,95662 |
| **OrSuffix** | 0,001 | 0,004 | 0,00433 | 0,015 | 0,009 | 0,01014 |
| **PassiveVoice** | 0,032 | 0,028 | 0,02467 | 0,06 | 0,042 | 0,03614 |
| **Past** | 0,117 | 0,115 | 0,0981 | 0,048 | 0,056 | 0,05633 |
| **PersonalPronoun** | 0,059 | 0,0685 | 0,06252 | 0,022 | 0,058 | 0,04867 |
| **Prepositive attributes** | 0,387 | 0,349 | 0,28315 | - | 0,509 | 0,35971 |
| **Postpositive attributes** | 0,172 | 0,1585 | 0,1343 | 0,279 | 0,279 | 0,20457 |
| **Present** | 0,078 | 0,061 | 0,05507 | 0,188 | 0,165 | 0,1261 |
| **ThatThose** | 0 | 0 | 0,00063 | 0 | 0 | 0,00014 |
| **WePronoun** | 0,006 | 0,0055 | 0,00555 | 0,011 | 0,011 | 0,01057 |
| **YouHeShe** | 0 | 0 | 0,00032 | 0 | 0 | 0,00129 |

**Таблица 6. Показатели корпусов направления Политология**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Маркер** | **Competent** | | | | **Learners** | | | |
| **Мода** | **Медиана** | **Среднее значение (на 1000 слов)** | **Мода** | | **Медиана** | **Среднее значение (на 1000 слов)** |
| **Abstract semantic verbs** | 0,071 | 0,69 | 0,06166 | - | | 0,1125 | 0,08785 |
| **Archaisms** | 0 | 0 | 0,0000781 | 0 | | 0 | 0,0001 |
| **Complex conjunctions** | 0,01 | 0,0145 | 0,01519 | 0 | | 0,008 | 0,0076 |
| **Desemantisized verbs** | 0,046 | 0,047 | 0,04494 | 0,087 | | 0,079 | 0,0631 |
| **Future** | 0,001 | 0,002 | 0,00281 | 0 | | 0,0165 | 0,01505 |
| **IPronoun** | 0,001 | 0,002 | 0,00294 | 0 | | 0,003 | 0,00315 |
| **Intensifying adverbs** | 0,002 | 0,002 | 0,00241 | 0 | | 0,001 | 0,00175 |
| **Logic connectors** | 0,009 | 0,006 | 0,0067 | 0,015 | | 0,0095 | 0,009 |
| **Noun** | 0,908 | 0,908 | 0,76631 | - | | 1,231 | 0,9343 |
| **OrSuffix** | 0,006 | 0,0055 | 0,0055 | 0,015 | | 0,011 | 0,0114 |
| **PassiveVoice** | 0,02 | 0,022 | 0,02156 | 0,048 | | 0,048 | 0,03965 |
| **Past** | 0,051 | 0,047 | 0,04452 | 0,043 | | 0,024 | 0,02755 |
| **PersonalPronoun** | 0,027 | 0,0325 | 0,03313 | 0 | | 0,047 | 0,03935 |
| **Prepositive attributes** | 0,407 | 0,382 | 0,33511 | 0,592 | | 0,494 | 0,374 |
| **Postpositive attributes** | 0,191 | 0,1835 | 0,15916 | 0,349 | | 0,2695 | 0,20845 |
| **Present** | 0,116 | 0,109 | 0,09561 | - | | 0,1785 | 0,14135 |
| **ThatThose** | 0 | 0 | 0,00034 | 0 | | 0 | 0,0001 |
| **WePronoun** | 0,01 | 0,01 | 0,01063 | 0 | | 0,0195 | 0,01555 |
| **YouHeShe** | 0 | 0 | 0,00028 | 0 | | 0 | 0,0012 |

На основе данных были построены графики. Для начала сравним показатели академических работ (Рис. 8).

**Рисунок 8. Сравнение показателей академических работ**

Как можем видеть, на графике представлены показатели по всем маркерам по каждому направлению. Проделаем тоже самое для письменных работ студентов НИУ ВШЭ Пермь (Рис. 9).

**Рисунок 9. Сравнение показателей письменных работ студентов**

Наиболее ценным результатом является то что с помощью статистики возможно просмотреть численное выражение того насколько студенческие работы отличны от академических. Для примера рассмотрим такое сравнение на корпусах работ направления Бизнес-информатика (Рис. 10).

**Рисунок 10. Сравнение показателей академических работ и работ студентов направления Бизнес-информатика**

На основе этих данных система в дальнейшем сможет сформировать памятку для студента в которой будет отображено каких элементов не хватает в работе, а каких слишком много.

**Библиографический список**

1. Трофимов Сергей Варианты использования (Use Case) [Электронный ресурс]. URL: http://www.caseclub.ru/articles/use\_case.html (дата обращения: 06.03.2017).

2. Cunningham H. [и др.]. Developing language processing components with GATE (a user guide) // University of Sheffield. 2001. № Gate 2 (2006). C. 1–457.

3. Laurence Anthony’s Homepage [Электронный ресурс]. URL: http://www.laurenceanthony.net/ (дата обращения: 03.02.2017).

4. GATE.ac.uk - index.html [Электронный ресурс]. URL: https://gate.ac.uk/ (дата обращения: 14.03.2017).

5. The Stanford Natural Language Processing Group [Электронный ресурс]. URL: https://nlp.stanford.edu/software/tagger.shtml (дата обращения: 16.03.2017).

6. Лингвистический Энциклопедический Словарь. Кореферентность [Электронный ресурс]. URL: http://tapemark.narod.ru/les/243a.html (дата обращения: 16.03.2017).

7. Чанкер для русского языка [Электронный ресурс]. URL: http://web-corpora.net/wsgi/chunker.wsgi/npchunker/ (дата обращения: 16.03.2017).

8. генератор отчетов - это... Что такое генератор отчетов? [Электронный ресурс]. URL: http://dic.academic.ru/dic.nsf/eng\_rus/56647/генератор (дата обращения: 14.03.2017).

9. Crystal Reports [Электронный ресурс]. URL: https://msdn.microsoft.com/ru-ru/library/bb126227(v=vs.90).aspx (дата обращения: 15.03.2017).

10. Генератор отчетов для Java. Экспорт отчетов в PDF, MS Office, Open Office и т.д. :: Стимулсофт [Электронный ресурс]. URL: https://www.stimulsoft.com/ru/products/reports-java (дата обращения: 15.03.2017).

11. Бизнес аналитика. Генератор отчетов для JavaScript, .NET, ASP.NET, WPF, Flex, PHP, Silverlight, Java, ASP.NET MVC, WinRT, HTML5. Дизайнер отчетов. Создание отчетов. Сервер отчетов. Бизнес-аналитика. :: Стимулсофт [Электронный ресурс]. URL: https://www.stimulsoft.com/ru (дата обращения: 15.03.2017).

12. JasperReports® Library | Jaspersoft Community [Электронный ресурс]. URL: http://community.jaspersoft.com/project/jasperreports-library (дата обращения: 15.03.2017).

13. Business reporting software - NextReports Designer, Engine and Server [Электронный ресурс]. URL: http://www.next-reports.com/ (дата обращения: 15.03.2017).

14. BIRT Home [Электронный ресурс]. URL: http://www.eclipse.org/birt/ (дата обращения: 15.03.2017).

15. About [Электронный ресурс]. URL: http://www.eclipse.org/birt/about/project-organization/ (дата обращения: 15.03.2017).

16. Hardie A. The IMS Open Corpus Workbench ( CWB ) CQPweb System Administrator ’ s Manual 2014. C. 1–45.

17. Scott M. WordSmith Tools step by step 2010.

18. Talk T., Diachronic T., Corpus E. A Guide to using AntConc C. 1–9.

19. Сравнение утилит анализа корпусов — NLPub [Электронный ресурс]. URL: https://nlpub.ru/Сравнение\_утилит\_анализа\_корпусов (дата обращения: 21.05.2017).

20. Каталог лингвистических программ и ресурсов в Cети, ч.1 / Linguistics Software Catalogue, ch.1 [Электронный ресурс]. URL: http://rvb.ru/soft/catalogue/c01.html (дата обращения: 21.05.2017).

21. Sketch Engine | language corpus management and query system [Электронный ресурс]. URL: https://www.sketchengine.co.uk/ (дата обращения: 04.02.2017).

22. Обзор платформы Eclipse - как её использовать | Статьи о программном обеспечении [Электронный ресурс]. URL: https://hightech.in.ua/content/art-eclipse-platform (дата обращения: 23.03.2017).

23. Welcome to NetBeans [Электронный ресурс]. URL: https://netbeans.org/ (дата обращения: 23.03.2017).

24. Обзор NetBeans IDE | Обзор IDE [Электронный ресурс]. URL: https://aboutide.wordpress.com/2011/03/23/обзор-netbeans-ide/ (дата обращения: 23.03.2017).

25. IntelliJ IDEA [Электронный ресурс]. URL: http://jetbrains.ru/products/idea/ (дата обращения: 23.03.2017).

26. SAP Crystal Reports 2011 User вЂTM s Guide ■ SAP Crystal Reports 2011 2012.