

## Программный продукт “TROPEs” и его возможности

Данный продукт был разработан во Франции при сотрудничестве компаний “Acetic” и “Cyberlex”. Авторы программы: Pierre Molette, Agnes Landre и Rodolphe Ghiglione. Первая версия программы была выпущена в 2009 году, работа над последующим обновлением была завершена в июне 2014 года. Форма использования продукта бесплатная. Данное программное обеспечение предназначено для семантической классификации, выделения ключевых слов, лингвистического и качественного анализа. Данный продукт является хорошим инструментом для исследований в области информационных технологий, исследования рынка, социологического анализа, научных исследований. Поддерживаемые языки: английский, французский, испанский, португальский, немецкий и итальянский.



На данном скриншоте представлен интерфейс программы. Сразу после запуска, в левом окне под панелью задач будет отображена краткая справка, в которой описан правильный алгоритм действий, выполнение которых необходимо для начала работы. Помимо этого данная справка предоставляет краткое описание содержания интерфейса. Анализируемый текст должен быть преобразован в формат “txt”; “ANSI”(Windows); “HTML” или “RTF”. Другими словами, текст должен содержаться на HTML странице, в

документе Microsoft Word или другом аналогичном текстовом редакторе. В тексте допускается наличие знаков препинания, цифр, круглых скобок и т.п. В процессе работы анализу подвергаются только буквы из алфавита и знаки препинания. Необходимо отметить, что между теми лексическими единицами, которые вы хотите интерпретировать, как одно слово, необходимо поставить нижнее тире (на пример: the\_head\_of\_government). Возможность непосредственного ввода текста в программу отсутствует. Анализ текста производится автоматически. Сразу после вставки текста в программу, на экране отображаются результаты анализа.

Любой текст, независимо от жанра цели высказывания, содержит в себе ключевые предложения, несущие в себе основную мысль дискурса. Такие предложения являются структурным стержнем текста. Помимо этого, любой текст содержит в себе информацию о тех или иных предметах, героях и событиях. Все перечисленные компоненты дискурса связаны между собой определёнными лексическими единицами. Эти лексические единицы можно распределить по уровню их значимости в дискурсе. В результате, мы можем сделать вывод о наличии в любом тексте определённого количества лексических единиц, без которых текст утратит свою семантическую конструкцию и основную мысль. Выявление в тексте таких структурно значимых элементов является первой и основной ступенью любого рода анализа.

Данное программное обеспечение позволяет мгновенно определить основную мысль текста и обозначить его главных героев. Используя данную программу, вы быстро найдете ответы на такие вопросы как «кто что кому говорит?»; «кто что делает?»; «где и когда было совершено, то или иное действие?»; «с какой целью было совершено то или иное действие?». С помощью технологии, речь о которой пойдет в следующем параграфе, данная программа способна выявлять стиль анализируемого текста. Из возможных вариантов, в программе представлено 4 различных стиля текста. Помимо этого, TROPES способна выявить в какой части текста повествование идёт от

первого лица и в какой от третьего. Посредством хронологического анализа, программа разбивает текст на эпизоды. Другими словами, используя TROPES, вам необязательно прибегать к прочтению того или иного текста для того что бы понять его основную мысль. Данная программа способна сделать это за вас.

### 2.1.1. Анализ стиля текста



Первое, что вы увидите в списке отображаемых результатов анализа – это стиль текста. В ходе данной работы мной был проведён анализ двух сказок «1001 ночь» и «Красавица и чудовище» на английском языке. В первом случае программа вынесла вердикт “Style rather argumentative”. После проведения анализа второго произведения, стиль дискурса был охарактеризован как “rather Enunciative”. Для того что бы увидеть обоснования результатов, достаточно кликнуть левой кнопкой мыши на надпись “text style” под вкладкой “results”, после чего на экране будут отражены лексические единицы, наличие которых в тексте объясняет сделанный вывод. Так, например, при определении стиля первого произведения, программа выделила из текста такие словоформы как “but”, “like”, “not”, “so”, “as”, “than” и т.д.

В программу заложены следующие стили:

- 1) Argumentative: discussing, comparing or criticizing
- 2) Narrative: telling a story, at a time, in a certain place
- 3) Enunciative: setting some influence, or revealing a point of view
- 4) Descriptive: describing something or somebody

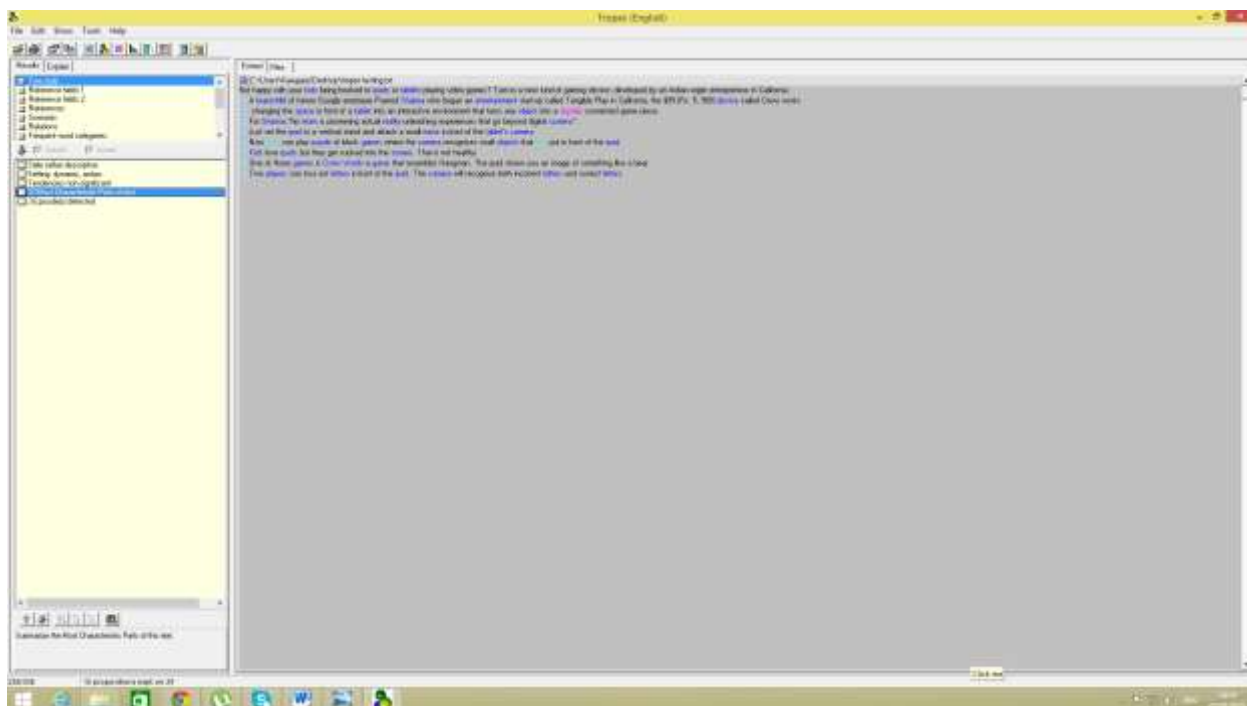
В этом же разделе, программа разбивает текст на две части: на ту, в которой повествование идёт от первого лица, и ту, в которой повествование идёт от третьего лица. В левой части экрана вы сможете увидеть подразделы “involving the narrator” и “involving with I”. После нажатия на первую вкладку, в главном окне будет отражена та часть текста, которая была выражена третьим лицом или автором. Соответственно, нажав на вторую вкладку, вы увидите в главном окне части текста, выраженные прямой речью. Однако именно здесь проявляется первый недостаток данной программы. В ходе работы, мне стало очевидно, что данная функция не всегда выполняется корректно. При анализе обоих произведений, имели место быть многочисленные случаи, в которых части текста, выраженные прямой речью, отображались под вкладкой повествования от лица автора. Так же, возникала проблема обратного характера, когда высказывания, выраженные от лица автора, отображались под вкладкой прямой речи.

### **2.1.2. Анализ наиболее характерных частей текста**

Данная функция предназначена для выявления в тексте предложений, имеющих непосредственное отношение к основной сюжетной линии. Как правило, это предложения, несущие сведения о главных героях и событиях, играющих важную роль по отношению к основной мысли всего дискурса. Программа выявляет эти части текста в основной сюжетной линии подобно тому, как строители обозначают несущие элементы (такие как стены) все конструкции, которые нельзя сносить ни при каких обстоятельствах.

По словам разработчиков, данная функция осуществляется за счёт дискурсивно-когнитивного анализа. Каждая семантическая единица текста

(то или иное суждение, высказывание или фраза) связана с другими посредством семантических сетей подобно паутине. В пределах этой паутины, каждая такая единица обладает своей значимостью или «влиятием» в той или иной степени. В ходе этой части анализа, семантические единицы дискурса распределяются по мере своей значимости. Части текста, обладающие наибольшей «сферой влияния», попадают в список наиболее характерных частей текста.

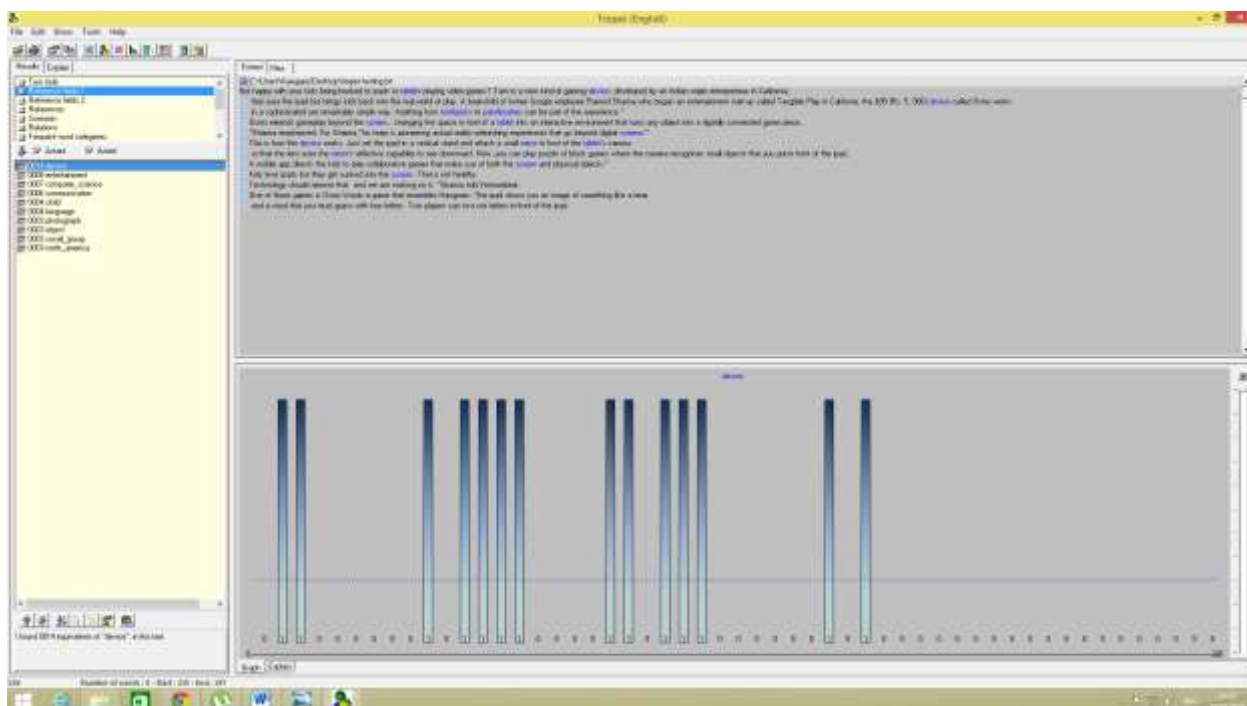


Данная функция предназначена для того, что бы сориентировать пользователя в случае необходимости сокращения текста. В то же время, вы можете прибегнуть к данной возможности программы с целью выявления основной мысли текста, сэкономив при этом время на его прочтение.

### 2.1.3. Анализ предметных областей

В данной графе отображаются лексические единицы, которые могут иметь отношение к одной предметной области. Например, слово «самолёт» имеет непосредственное отношение к такой предметной области как «техника» или «авиация». Причем, в данной программе такие отношения между лексическими единицами приобретают иерархическую структуру. В окне результатов вы увидите три раздела, посвященных этой части анализа.

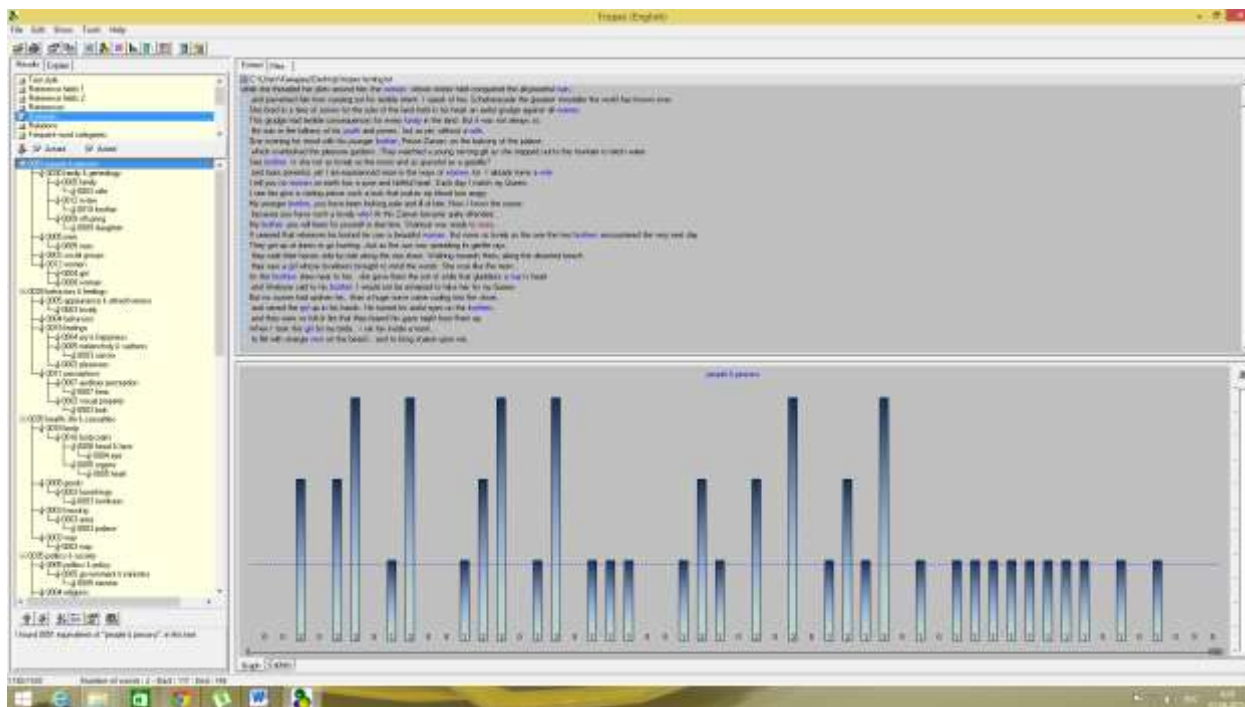
Первый раздел носит название “Reference fields 1”. Второй раздел называется так же, но обозначен вторым номером. Третий раздел называется “References”. Для наглядности можно привести следующий пример. Словосочетание “Lord Chancellor” определяется словом “minister”, которое в свою очередь попадает в категорию “politics”.



В первом разделе размещаются лексические единицы, занимающие наивысшую позицию в иерархической структуре отношений предметных областей. Так, например, при анализе произведения «1001 ночь», было выделено 19 наименований потенциальных предметных областей. Первое место в этом списке заняла предметная область под названием “family”. После клика по данной позиции в меню, в главном окне были отображены все предложения, содержащие в себе слова, имеющие отношения к данной предметной области. Сюда вошли такие слова как “brother”, “sister”, “wife” и т.д. Все эти слова были выделены другим цветом. Помимо этого, в нижнем правом окне было приведено количество словоупотребления в тексте каждой из представленных лексических единиц.

#### 2.1.4. Анализ семантических сетей

В меню результатов анализа данный раздел обозначен как “Scenario”. В этом разделе, предметные области и лексические единицы, относящиеся к ним, визуализированы в форме древовидных семантических сетей. Данные семантические сети иллюстрируют отношения между самими предметными областями лексических единиц.

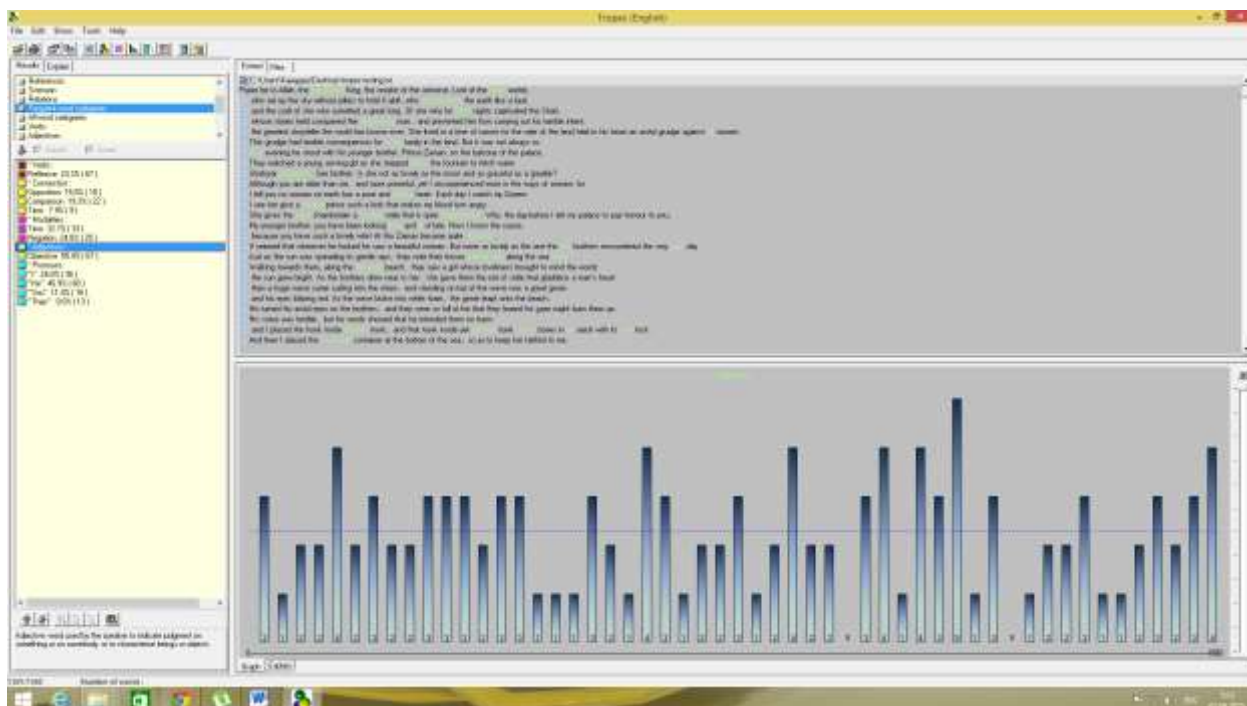


В качестве примера можно привести одно из деревьев зависимости, составленное в ходе анализа. Корнем первого дерева стала категория под названием “people & persons”. Отсюда исходят такие ответвления как “family & genealogy”, “social groups” и т.д. Это лишь одно из многочисленных деревьев зависимости, выявленных в ходе анализа текста. Помимо всего прочего, программа предоставляет возможность вносить свои поправки в структуру данных семантических сетей. Данная функция предназначена для решения проблемы двусмысленности, которая возникает в ходе анализа текста.

### 2.1.5. Анализ частей речи

В процессе анализа текста, программа автоматически разделяет все лексические единицы на соответствующие части речи. Эта часть анализа текста разбита на наибольшее количество разделов, чем все остальные. В

первом разделе представлены данные, отображающие количество употреблений в тексте каждой отдельной части речи и их процентное соотношение. На первом месте стоят глаголы. За глаголами следуют союзы. За союзами следуют различного рода модальности (modalities), например лексические единицы или фразы, обозначающие время (then, sooner, each day и т.д.). Сюда же относятся отрицательные частицы (not, never, none и т.д.). Далее представлены прилагательные. В конце списка стоят местоимения.



Следующий раздел называется “All word categories”. В данном разделе подробно разобраны все части речи. Например, все союзы разделены по своим разновидностям, таким как “disjunction”, “opposition”, “addition” и т.д.

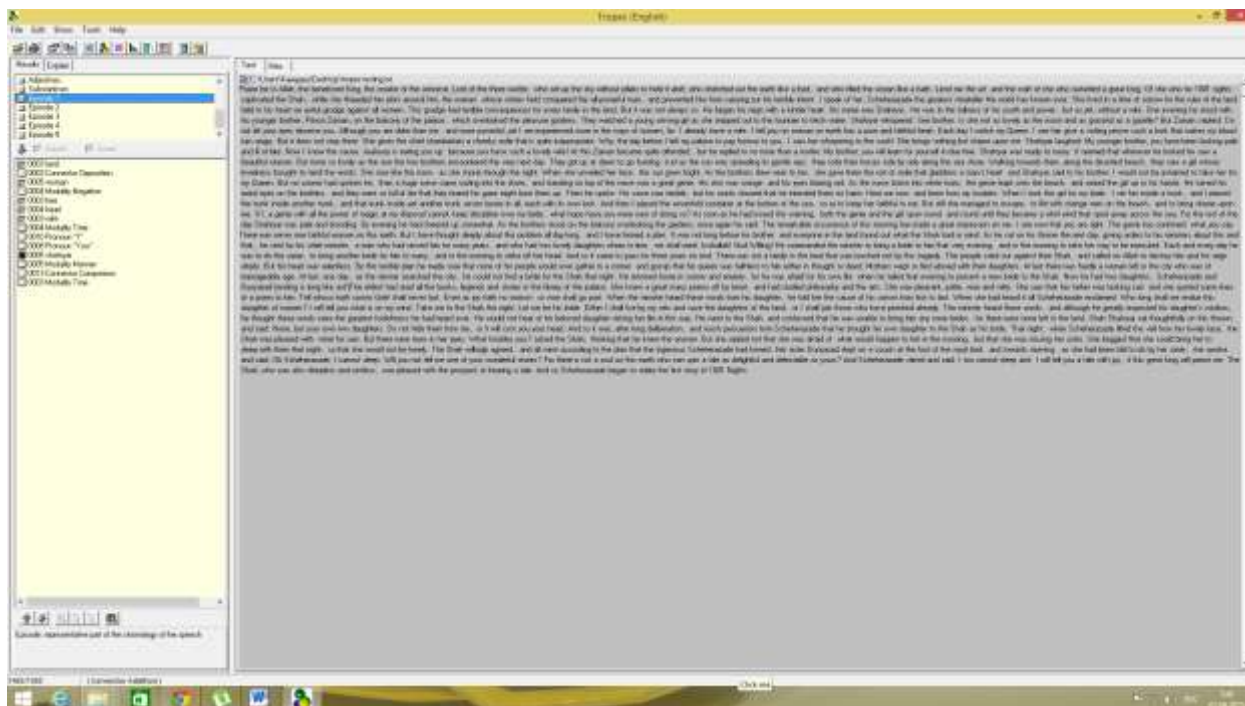
В следующих трёх разделах перечисляются все соответствующие им части речи, к каждой из которых прилагается статистика частоты употребления. Сюда относятся глаголы, существительные и прилагательные.

### 2.1.6. Анализ отдельных глав текста

Как говорилось выше, в процессе анализа, программа разбивает текст на главы или эпизоды. Этот процесс происходит за счёт выявления в тексте обособленных групп лексических единиц, связанных между собой по смыслу и употреблённых в тексте с определённой плотностью в той или иной его



части.



Так, например, в первой главе декларации по правам человека наибольшей плотностью обладает связка слов «государство», «член», «свобода». Во второй главе данной декларации наибольшей плотностью обладает связка таких слов как «работа», «образование», «защита».

Количество разделов, посвящённых данной части анализа, зависит от количества выявленных в тексте глав. В каждом разделе представлены именно те слова, которые выделяют эту конкретную главу из анализируемого текста.

## **2.2. Недостатки программного продукта “TROPES”**

В данный продукт заложена функция, которая предполагает возможность отделения частей анализируемого текста, повествуемых третьим лицом, от прямой речи. В ходе работы с программой в этой области были выявлены нарушения. При анализе обоих произведений, случаи употребления прямой речи неоднократно были отнесены к той части текста, повествование которой исходило от третьего лица. Однако, следует отметить что обратной проблемы не возникало.

Помимо этого, заложенных в данную программу вариантов стилей текста явно недостаточно, для того что бы дать анализируемому дискурсу полноценную характеристику. Говоря о возможных доработках касаясь этой проблемы, можно сделать вывод о необходимости наличия в данном программном продукте возможности определения жанра.

Помимо вышесказанного, других серьёзных технических недоработок в ходе работе выявлено не было. Однако стоит отметить, что дизайн интерфейса находится на уровне Windows 98 и непременно должен быть доработан.