

ОБРАЗЕЦ ОПИСАНИЯ ПРОГРАММЫ.

Рассмотрим программу анализа и лингвистической обработки текстов **TextAnalyst2.0**. Это смысловой анализатор, который за считанные минуты позволяет ознакомиться с текстами любой тематики. Программа разработана в качестве инструмента для анализа содержания текстов, смыслового поиска информации, формирования электронных архивов и предоставляет пользователю следующие основные возможности:

- анализ содержания текста с автоматическим формированием семантической сети с гиперссылками;
- получение смыслового портрета текста в терминах основных понятий и их смысловых связей;
- анализ содержания текста с автоматическим формированием тематического древа с гиперссылками;
- выявление семантической структуры текста в виде иерархии тем и подтем;
- смысловой поиск с учетом скрытых смысловых связей слов запроса со словами текста;
- автоматическое реферирование текста;
- автоматическая индексация текста с преобразованием в гипертекст;
- ранжирование всех видов информации о семантике текста по «степени значимости» с возможностью варьирования детальности ее исследования.

При исследовании искусствоведческих текстов многие из этих возможностей могут быть полезными. Для начала работы запускаем программу и выбираем текстовый файл для анализа. В нашем случае это будет файл "Поэзия земли.txt", содержащий статью о русском художнике Владимире Копылове. Теперь главное окно TextAnalyst должно выглядеть примерно так:

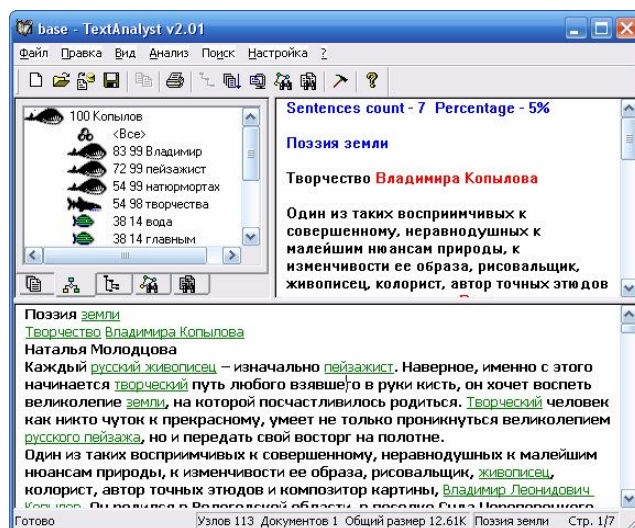


Рисунок 1 – Интерфейс программы TextAnalyst

Вся работа по анализу текста уже сделана, остается лишь ознакомиться с ее результатами. Прежде всего, изучив предложенный материал, TextAnalyst формирует сеть основных (наиболее значимых) понятий, содержащихся в представленных ему текстах (верхнее левое окно на Рис. 1). В нашем случае в сеть понятий входят: "Копылов", "Владимир", "пейзажист", "натюрмортах", "творчества", "вода", "главным", "живописца", "рисунок", "художественный" и др.

Такая сеть служит представлением смысла текста и основой для всех видов дальнейшего анализа. **Сеть понятий** – это множество терминов из текстов – слов и словосочетаний, связанных между собой по смыслу. В сеть включены не все термины текста, а лишь наиболее значимые, несущие основную смысловую нагрузку. Аналогичным образом представлены и смысловые связи между понятиями текстов – отражаются лишь наиболее явно выраженные из них. Поэтому, с одной стороны сеть достаточно полно описывает смысл текстов, а с другой – позволяет отбросить несущественную информацию и представить содержание в сжатом виде, так называемом “смысловом портрете”.

Таким образом, можно сразу увидеть всю информацию по каждому понятию, буквально бросив единственный взгляд на набор его связей в сети. В результате, передвигаясь по смысловым связям от понятия к понятию, можно находить и прицельно исследовать лишь интересующие места текстов, не затрудняя себя просмотром всей попавшейся на пути информации.

Теперь обратим внимание на числа, стоящие в дереве возле понятий. Ближайшее к понятию число представляет его **смысловый вес** (например, "Копылов" – 100, "Владимир" – 99, "творчества" – 98). Его значение (от 1 до 100) показывает, сколь важную роль играет понятие для смысла всего текста

– как много информации в тексте касается данного понятия. Максимальное значение, равное 100, говорит о том, что понятие является ключевым и представляет важнейшую тему текста. Маленькое, близкое к единице значение показывает, что соответствующая тема лишь вскользь упомянута в тексте и в нем очень мало информации, относящейся к данному понятию. Второе число, стоящее перед смысловым весом, ближе к раскрытому узлу, представляет **вес связи** от понятия в вершине раскрытого списка к данному. Вес связей также всегда принимает значение от 1 до 100.

Кроме того, программа предоставляет услугу автоматического **реферирования** (в меню "Анализ" выбираем пункт "Реферирование"). Формируемый реферат содержит список наиболее информативных предложений текстов (тезисов) (Рисунок 2, правое верхнее окно). Конечно, это еще не полноценный реферат, так как тезисы в основном не связаны между собой стилистически, а просто выбраны из текста и расположены в порядке их встречаемости. Однако и такой подстрочник реферата оказывается достаточно информативным, чтобы составить общее представление о тексте и уяснить его основные мысли. Более того, все предложения реферата снабжены отсылками к соответствующим местам исходных текстов, что позволяет просмотреть контекст интересующего тезиса. Подробность реферата можно легко настраивать, изменяя количество формирующих его предложений. При этом каждое предложение реферата характеризуется относительной степенью значимости во всем тексте.

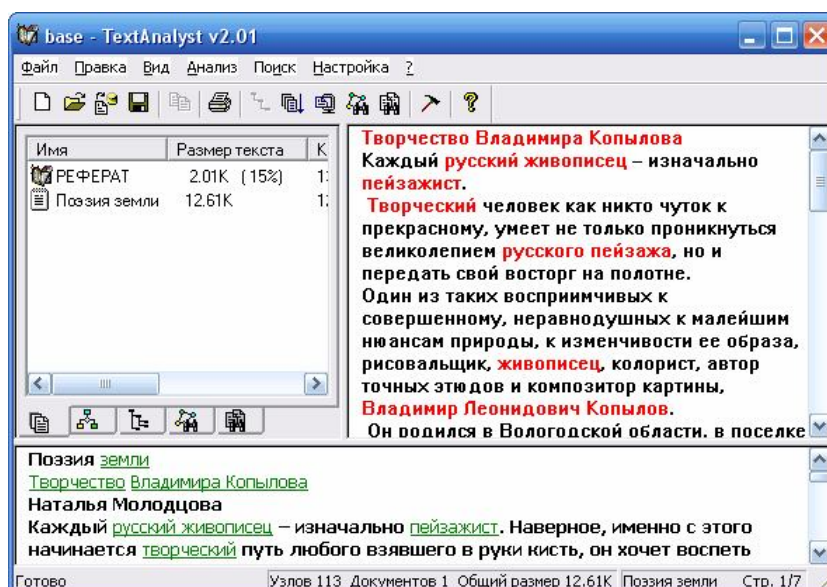


Рисунок 2 – Реферирование

Таким образом, данная программа может быть достаточно полезной в исследовании автора, поскольку позволяет быстро получить смысловой портрет текста, что необходимо в процессе работы с большим количеством

разнообразных текстов, когда нелегко удержать в памяти содержание каждого из них. TextAnalyst удобно использовать и при поиске текстов для исследования: не обязательно тратить время на чтение полного текста, с помощью реферирования можно быстро сориентироваться, о чем идет речь (о каких художниках, картинах, направлениях в живописи и т.д.), и сделать вывод о том, подходит ли данный текст для исследования. Однако серьезным недостатком программы применительно к нашему исследованию, связанному с англоязычными текстами, является отсутствие возможности обрабатывать такие тексты.