

Хабаров А.Р., Лебедев В.В., Карельская К.А. Формирование семантической сети в интеллектуальном анализаторе текстовой информации TextAnalyst. // Проблемы информатики в образовании, управлении, экономике и технике: Сб. статей XVI Междунар. научно-техн. конф. – Пенза: ПДЗ, 2016. – С. 248-252.

УДК 004

ФОРМИРОВАНИЕ СЕМАНТИЧЕСКОЙ СЕТИ В ИНТЕЛЛЕКТУАЛЬНОМ АНАЛИЗАТОРЕ ТЕКСТОВОЙ ИНФОРМАЦИИ TextAnalyst

А.Р. Хабаров, В.В. Лебедев, К.А. Карельская

FORMATION OF SEMANTIC NETWORK IN THE INTELLIGENT ANALYZER OF TEXT INFORMATION TextAnalyst

A.R. Khabarov, V.V. Lebedev, K.A. Karelskaya

Аннотация. Рассматривается использование интеллектуальных систем для анализа содержания и смыслового поиска текстовой информации в процессе обучения.

Ключевые слова: искусственный интеллект, семантическая сеть, интеллектуальный анализатор.

Abstract. The article discusses the use of intelligent systems for the analysis of contents and semantic searching of text information in the learning process.

Keywords: artificial intelligence, semantic network, intelligent analyzer.

Для дальнейшего повышения качества выполнения студентами рефератов, лабораторных и курсовых работ в обучающий процесс активно внедряются современные интеллектуальные системы для анализа содержания и смыслового поиска текстовой информации.

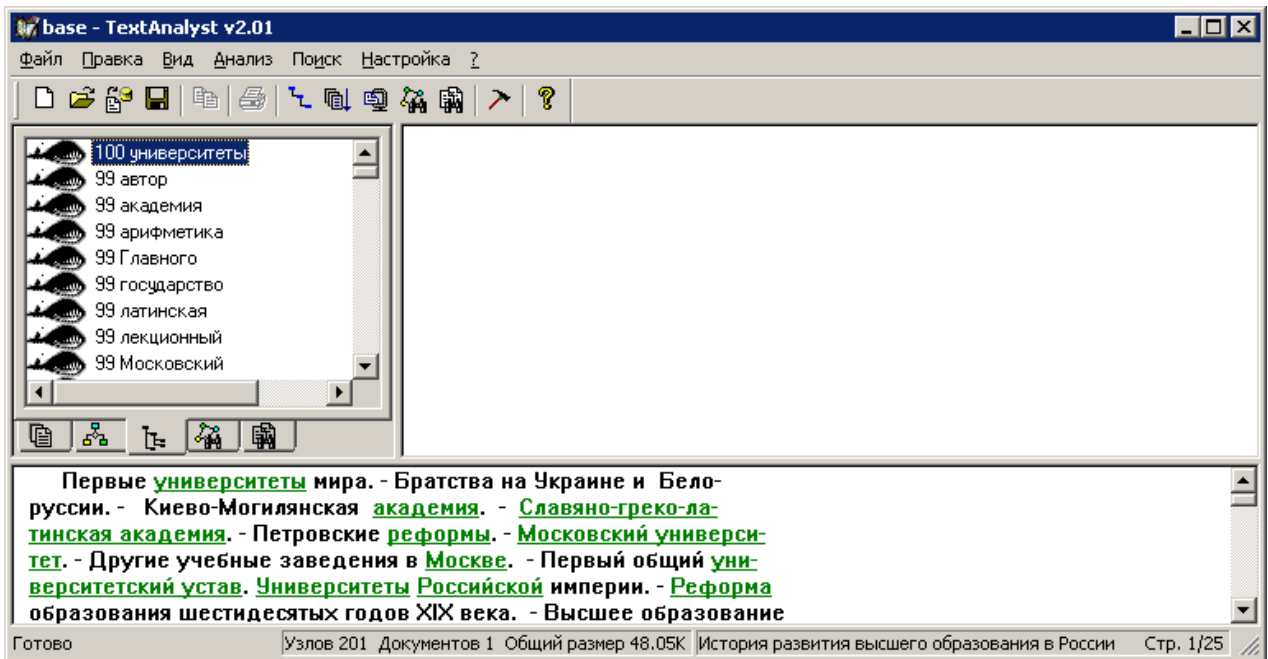
Для удобства работы весь интерфейс TextAnalyst построен на базе трех взаимосвязанных окон:

«Окно 1» – окно значимых элементов текста, располагается в левом верхнем углу экрана;

«Окно 2» – окно ссылок к предложениям текстов, располагается в правом верхнем углу экрана;

«Окно 3» – окно анализируемых текстов, располагается в нижней части экрана.

В процессе работы все три окна всегда расположены на экране. В этих окнах TextAnalyst покажет всю информацию, полученную им при изучении текста. Для формирования семантической сети в интеллектуальном анализаторе текстовой информации TextAnalyst выполняют следующие действия: выберите закладку «Сеть понятий» или выполните команду «Вид, Семантическая сеть». Сеть основных понятий проанализированного текста в пакете TextAnalyst представлена на рисунке.



Сеть основных понятий проанализированного текста

Теперь в окне 1 представлена (в виде графа обычного дерева) сеть основных понятий проанализированных текстов. Прежде всего, изучив предложенный материал, TextAnalyst формирует сеть основных (наиболее значимых) понятий, содержащихся в представленных ему текстах. 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, в виде так называемого «смыслового портрета». При этом каждое понятие, повторявшееся в различных местах текстов множество раз, оказывается представлено в единственном узле сети. В этом узле также собирается разбросанная информация, касающаяся понятия, формируется список предложений, в которых оно употреблялось. А различные формы слов приводятся к общей грамматической форме для отображения в один элемент сети. Аналогичным образом собирается информация по смысловым связям каждого понятия в виде списка всех связанных с ним в тексте понятий, дополненного предложениями, в которых отражаются данные связи.

Таким образом, можно сразу увидеть всю информацию по каждому понятию – теме текста, 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.

Каждый элемент сети (понятие) характеризуется числовой оценкой – так называемым смысловым весом. Связи между парами понятий, в свою очередь, также характеризуются весами. Эти оценки позволяют сравнить относительный вклад различных понятий и их связей в семантику текста, выявить более или менее подробно проработанную в тексте тематику, задать способ сортировки информации.

И наконец, позволяют взглянуть на весь текстовый материал по пластам – смысловым срезам различной глубины.

Необходимо обращать внимание на числа, стоящие в дереве возле понятий. Ближайшее к понятию число представляет его смысловой вес, значение (от 1 до 100) которого показывает, сколь важную роль играет понятие для смысла всего текста, т.е. 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 означает, 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. Большое значение веса связи (близкое к 100) от одного понятия к другому указывает на то, что подавляющая часть информации в тексте, касающаяся первого, касается в то же время и второго понятия (например, вторая тема почти всегда излагается в контексте первой).

Следует отметить, что кроме перечисленного, программа TextAnalyst предоставляет следующие 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;
 - смысловой поиск с учетом скрытых смысловых связей слов запроса со словами текста;
 - автоматическое реферирование текста – формирование его смыслового портрета в терминах наиболее информативных фраз;
 - кластеризацию информации – анализ распределения материала текстов по тематическим классам;
 - автоматическую индексацию текста с преобразованием в гипертекст;
- ранжирование всех видов информации о семантике текста по «степени значимости» с возможностью варьирования детальности ее исследования.

Таким образом, благодаря развитой функциональности в области семантического анализа текстов пакет персонального анализа текстов TextAnalyst успешно используется для выполнения курсовых работ по дисциплине «Интеллектуальные системы».

Библиографический список

1. Советов Б.Я., Цехановский В.В., Чертовской В.Д. Интеллектуальные системы и технологии. М.: Академия, 2013.
2. Матвейкин В.Г., Дмитриевский Б.С., Ляпин Н.Р. Информационные системы интеллектуального анализа. М.: Машиностроение, 2008.

Хабаров Алексей Ростиславович

Тверской государственный
технический университет,
г. Тверь, Россия

E-mail: al_xabarov@mail.ru

Лебедев Владимир Владимирович

Тверской государственный
технический университет,
г. Тверь, Россия

E-mail: Lebedev_vl.69@mail.ru

Карельская Катерина Александровна

Тверской государственный
технический университет,
г. Тверь, Россия

Khabarov A.R.

Tver State Technical University,
Tver, Russia

Lebedev V.V.

Tver State Technical University,
Tver, Russia

Karelskaya K.A.

Tver State Technical University,
Tver, Russia