



УДК 004.822:514

## АВТОМАТИЧЕСКОЕ ФОРМИРОВАНИЕ НЕОДНОРОДНОЙ СЕМАНТИЧЕСКОЙ СЕТИ НА ОСНОВЕ ВЫЯВЛЕНИЯ КЛЮЧЕВЫХ ПРЕДИКАТНЫХ СТРУКТУР ПРЕДЛОЖЕНИЙ ТЕКСТА

Харламов А.А. \*, Ермоленко Т.В. \*\*

\* *Институт высшей нервной деятельности и нейрофизиологии РАН,  
г. Москва, Россия*

kharlamov@analyst.ru

\*\* *Институт проблем искусственного интеллекта, г. Донецк, Украина*

etv@iai.dn.ua

В докладе представлен новый подход к автоматическому выявлению семантических отношений в тексте. Подход использует ту же методику формирования семантической сети, которая используется в ранее разработанной одним из авторов технологии автоматического смыслового анализа текстов TextAnalyst, только в качестве исходного материала к этому анализу вместо обычного текста используется текст с выявленными предикатными структурами предложений. Разметка осуществляется автоматически на основе правил согласования слов в синтаксических группах. В процессе анализа, также как и в случае однородной (ассоциативной) семантической сети, формируется семантическая сеть, ключевые понятия которой ранжированы по их смысловой значимости в тексте, но наряду с ассоциативными отношениями между понятиями сети используются некоторые общепринятые семантические отношения, в результате чего вместо однородной (ассоциативной) сети формируется неоднородная семантическая сеть. Эта неоднородная семантическая сеть является объединением точно выявленной (в виде предикатных структур) семантики отдельных предложений с обобщенной семантикой целого текста, представленной семантической сетью текста. **Ключевые слова:** автоматическое выявление предикатных структур, автоматический смысловой анализ текста, автоматическое формирование неоднородной семантической сети, синтаксические правила.

### ВВЕДЕНИЕ

Основное содержание текста может быть выражено разными способами: перечислением ключевых предикатных структур, содержащихся в предложениях текста, выявлением ключевых понятий текста в их взаимосвязях с последующим построением семантической сети. В первом случае предикатные структуры раскрывают содержание отдельных предложений текста. Причем сумма их смыслов не исчерпывает содержания текста [Ягунова, 2008]. Во втором случае семантическая сеть текста теряет подробности, содержащиеся в смысле отдельных предложений. Объединение двух подходов, заключающееся во включении подробной семантической структуры предложений в общую семантическую сеть текста позволит сохранить детальность описания семантики отдельных предложений с обобщенными семантическими представлениями целого текста. Такое объединенное представление совмещает детальность описания семантики текста с помощью

неоднородной семантической сети с автоматизмом формирования этой сети, что существенно отличает это от других подобных представлений, в которых неоднородная семантическая сеть строится исключительно вручную.

Для выявления предикатных структур необходимо сначала определить, какие понятия текста (слова и устойчивые словосочетания) являются ключевыми понятиями в тексте, а затем определить, какие предикатные структуры, их содержащие, являются наиболее важными в этом тексте.

Выявление ключевых понятий текста не является на настоящий момент очень сложной задачей. Существует множество подходов к их определению. Есть подходы, основанные на заранее сформированных семантических сетях предметных областей как в Excalibur [convera.com, 2011]. В других случаях в процессе поступления текстов формируется семантическая сеть конкретного текста как в TextAnalyst [Kharlamov, 2004; Харламов, 2001]. Однако выявление ключевых

понятий, даже в их взаимосвязях в тексте, оказывается недостаточным для описания смысла текста. Чаще всего, этого описания хватает лишь для классификации текста (отнесения текста к одной из предметных областей).

Поскольку предикатная структура характеризует отношения понятий в предложении, последовательность ключевых предикатных структур предложений текста исчерпывает содержание текста. Таким образом, выявление ключевых предикатных структур позволяет эксплицировать основное содержание текста. Причем ранжирование предикатных структур в тексте по их смысловому весу позволяет формировать более или менее компактные выжимки смысла текста.

Выявление ключевых предикатных структур в тексте возможно с использованием того же подхода, который был использован для выявления ключевых понятий в тексте в технологии TextAnalyst и зарекомендовал себя достаточно хорошо [Gartner Symposium, 2001]. Отличие в использовании технологии в случае выявления ключевых понятий, и в случае выявления ключевых предикатных структур заключается в первичной обработке текста. Во втором случае специфическая первичная обработка позволяет выявлять ключевые предикатные структуры вместо ключевых слов. Первичная обработка для выявления предикатных структур в русскоязычных текстах может использовать, например, согласование в синтаксических группах и формальные признаки для выявления типов отношений [Сокирко, 2001], например как в словаре валентностей глаголов.

После выявления предикатных структур и построения их иерархий в предложениях становится возможным построить первичную семантическую (уже неоднородную, размеченную типами связей) сеть также как ассоциативную сеть в технологии TextAnalyst. В этом случае звездочки ключевых понятий с их ближайшими ассоциантами как их (ключевых понятий) семантическими признаками заменяются на предикатные структуры. Таким образом, ассоциативные отношения заменяются на неоднородные отношения актантов предикатива. Перенормировка весов ключевых понятий (как вершин) уже этой сети затем позволит получить семантическую сеть текста, и, тем самым, ранжировать предикатные структуры по их смысловому весу в тексте.

## **1. Отображение в многомерное пространство как способ формирования поуровневых словарей событий разной частоты встречаемости**

Обработка внутренне структурированной информации различных модальностей (в том числе текстовой), имеющей многоуровневую структуру, с помощью нейронных сетей на основе нейроподобных элементов с временной суммацией

сигналов [Харламов, 2001] позволяет автоматически сформировать словари событий разной частоты встречаемости (словари разных уровней иерархии). Такая обработка сводится к отображению  $F$ , подвергнутой предварительно первичной обработке информационной последовательности  $A$ , в многомерное пространство  $R^n$ , в результате которого информационная последовательность преобразуется в связанную последовательность точек многомерного пространства – траекторию  $\hat{A}$ :

$$\hat{A} = F(A). \quad (1)$$

Это отображение обладает свойством ассоциативности обращения к траектории: как только в информационной последовательности появляется повторяющийся фрагмент, траектория возвращается к ранее пройденному участку. Запоминание числа прохождений траекторией точек многомерного пространства, с последующим применением порогового преобразования, позволяет выявлять фрагменты траектории заданной частоты появления, которые составляют словари событий входной информации заданной частоты встречаемости  $\{\hat{B}_i\}$ . Для лингвистической информации это, например, словари флективных морфем, корневых основ, синтаксических групп. Выявленные таким образом лингвистические единицы в дальнейшем можно использовать для обработки текстовой информации. Словарь флективных морфем можно использовать для морфологического анализа, словарь корневых основ – для выявления ключевых понятий в тексте и формирования однородной (ассоциативной) семантической сети [Харламов, 2008], словарь синтаксических групп – для формирования неоднородной семантической сети.

Ранее одним из авторов была реализована технология обработки текстовой информации TextAnalyst [Sullivan, 2001], позволяющая автоматически выявлять ключевые понятия в тексте на основе только информации о структуре самого текста (независимо от предметной области). Для этого формировался частотный портрет текста, содержащий информацию о частоте встречаемости понятий текста, представленных как корневые основы соответствующих слов, или их устойчивых сочетаний, встречающихся в тексте, а также об их совместной (попарной) встречаемости в смысловых фрагментах текста (предложениях). Частотный портрет, таким образом, содержал информацию о частоте встречаемости понятий и их попарной (в терминах их ассоциативной связи) встречаемости в тексте. Использование хопфилдоподобного алгоритма [Hopfield, 1982] позволяло перейти от частоты встречаемости к смысловому весу (вес связей при этом оставался неизменным).

Этот способ включал несколько этапов. На этапе первичной обработки из текста удалялась нетекстовая информация, текст сегментировался на

слова и предложения, из текста удалялись стоп-слова, рабочие и общеупотребимые слова, а оставшиеся слова подвергались морфологической обработке. Для простоты анализа морфологическая обработка производилась с использованием традиционного морфологического словаря – словаря первого уровня –  $\{\hat{B}_i\}_1$ . Далее формировался словарь второго уровня –  $\{\hat{B}_i\}_2$  – словарь корневых основ (и устойчивых словосочетаний). На следующем этапе строился частотный портрет текста, то есть выявлялись частоты  $p_i$  встречаемости корневых основ понятий (полученных в результате морфологического анализа) и их устойчивых сочетаний, и частоты  $p_{ij}$  их попарной встречаемости в предложениях текста (то есть формировался словарь третьего уровня  $\{\hat{B}_i\}_3$ ). И, наконец, на третьем этапе, частоты встречаемости перенормировывались в смысловые веса с использованием итеративной процедуры, похожей на алгоритм искусственной нейронной сети, предложенной Хопфилдом:

$$w_i(t+1) = \left( \sum_{i \neq j} w_i(t) w_{ij} \right) \sigma(\bar{E}), \quad (2)$$

здесь  $w_i(0) = \ln p_i$ ;  $w_{ij} = \ln p_{ij} / \ln p_j$  и  $\sigma(\bar{E}) = 1/(1 + e^{-k\bar{E}})$  функция, нормирующая на среднее значение энергии всех вершин сети  $\bar{E}$ . В результате итеративной процедуры перенормировки наибольшие веса получали понятия, связанные с наибольшим числом других понятий с большим весом, то есть те понятия, которые стягивают на себя смысловую структуру текста. Полученные таким образом смысловые веса ключевых понятий показывают значимость этих понятий в тексте. В дальнейшем эта информация используется для выявления предложений текста, содержащих наиболее важную информацию в тексте.

В результате получается так называемая ассоциативная (однородная) семантическая сеть  $N$  как совокупность несимметричных пар понятий  $\langle c_i c_j \rangle$ , где  $c_i$  и  $c_j$  – понятия, связанные между собой отношением ассоциативности (совместной встречаемости в некотором фрагменте текста):

$$N \cong \{ \langle c_i c_j \rangle \}, \quad (3)$$

В данном случае отношение ассоциативности несимметрично:  $\langle c_i c_j \rangle \neq \langle c_j c_i \rangle$ .

Семантическая сеть, описанная таким образом, может быть переопределена как множество так называемых звездочек  $z_i = \langle c_i \langle c_j \rangle \rangle$ :

$$N \cong \{ z_i \} = \{ \langle c_i \langle c_j \rangle \rangle \}, \quad (4)$$

Под звездочкой  $z_i = \langle c_i \langle c_j \rangle \rangle$  понимается конструкция, включающая главное понятие  $c_i$ , связанное с множеством понятий-ассоциантов  $\langle c_j \rangle$ , которые являются семантическими признаками главного понятия, отстоящими от главного понятия в ассоциативной сети на одну связь. Ассоциативные связи направлены от главного понятия к понятиям-ассоциантам.

## 2. Синтаксический анализ как средство первичной обработки для последующего семантического анализа

В процессе синтаксического анализа из предложений текста выделяются содержащиеся в них предикатные структуры, а также отношения сочинения, и атрибутивные отношения. Общую схему действий можно представить в виде последовательности шагов.

- Членение предложения по знакам пунктуации и сочинительным союзам на начальные сегменты (фрагментация). Определение вершин и типов начальных сегментов.

- На декартовом произведении омонимов внутри начальных сегментов построение множества однозначных морфологических интерпретаций каждого сегмента.

Построение синтаксических групп для каждой интерпретации сегмента с помощью синтаксических правил, выявляющих синтаксические связи между словами. Оценка синтаксического покрытия каждой интерпретации.

- Установление иерархии между сегментами с помощью синтаксических правил: вложения контактно расположенных сегментов (причастных, деепричастных оборотов, обособленного определения); определение однородности между контактно расположенными сегментами; определение отношения импликации между сегментами по подчинительным союзам, в них входящим.

- Поиск шаблона, соответствующего одной из минимальных структурных схем предложений (МСС), описывающих предикативный минимум предложения [Белошарова и др., 1997].

Множество простых предложений русского языка задается перечнем МСС, описывающих предикативный минимум предложения. МСС — это модель, отвлеченный образец, отражающий способ выражения предикативности.

Использование МСС в качестве формального образца позволяет получить предикативную основу (структурную схему) простого предложения, и в дальнейшем — его предикатную структуру [Харламов и др., 2011]. Это первый и обязательный шаг для проведения первичного семантического анализа в формировании информационного

портрета текста, поскольку смысловая связь между понятиями предложения (объектом/субъектом) в общем случае может быть описана предикатом, актантами которого выступают данные понятия. Установление таких синтактико-семантических связей позволяет сформировать схему ситуации, описываемой во фразе.

Обусловленный валентностью предиката семантико-синтаксический уровень анализа конструкций, не соответствующий узкому собственно формально-синтаксическому подходу, дает возможность даже из набора неправильных форм (посредством приведения их к начальным формам) с помощью заполнения валентных гнезд определить схему предложения.

Семантико-синтаксический анализ предложения предусматривает создание электронного словаря валентности глаголов. При этом для каждого из глаголов (около 20 тысяч в русском языке) необходимо указать, какими падежами и с какими предлогами он может управлять, а также — в каких семантических ролях (семантических падежах) выступают актанты глагола.

Предикатная структура предложения по сути дела является звездочкой. Эти звездочки объединяются в сеть. Объединение иерархично: к главному понятию первой звездочки (первого предложения) присоединяются звездочки других предложений, в которых главным понятием является субъект предикатной структуры, они объединяются с главным понятием первой звездочки ассоциативными связями. К их ассоциантам как главным понятиям уже присоединяемых звездочек присоединяются с помощью соответствующих связей (предикатных, атрибутивных, ассоциативных) второстепенные понятия этих звездочек, и так до включения в сеть всех понятий. Затем сеть перенормируется как и в разделе 2. Используя (4), предикатную структуру, описанную с помощью (1), можно представить в виде звездочки

$$\langle Pred \langle Obj_i \rangle \rangle, i = 0 \dots n, \quad (5)$$

где *Pred* является главным понятием, *Obj<sub>i</sub>* — понятиями-ассоциантами (*Obj<sub>0</sub>* — активный субъект *Subj*). Поскольку глагольная конструкция считается несущей семантической нагрузки в ассоциативной семантической сети, место главного понятия в звездочке занимает субъект.

Затем, эти звездочки объединяются в сеть. Объединение иерархично: к ассоциантам субъекта предикатной структуры (главного понятия) первой звездочки (первого предложения) присоединяются звездочки других предложений, в которых главные понятия (субъекты их предикатных структуры) совпадают с ассоциантами первой звездочки. К ассоциантам главного понятия первой звездочки также присоединяются с помощью соответствующих связей (атрибутивных,

ассоциативных) второстепенные понятия первой звездочки (соответствующие актанты ее предикатной структуры). Этот процесс продолжается до включения в сеть всех понятий. Затем сеть перенормируется как в разделе 1.

### 3. Семантический анализ текста

Предикатные структуры предложений анализируемого текста являются входными единицами для последующей обработки информации, почти повторяющей обработку, описанную в разделе 1, для выявления ключевых понятий текста. Синтаксические группы канонического представления предложения являются элементами звездочек в терминах такой обработки [Харламов, 2008]. Главное слово синтаксической группы соответствует главному понятию звездочки  $c_i$ , оно связано с второстепенными словами, которые являются его ближайшими ассоциантами —  $\langle c_j \rangle$ . В отличие от подобного представления в разделе 1, звездочки являются не парами, а тройками. В этом случае под размеченной (неоднородной) семантической сетью  $N$  понимается совокупность несимметричных троек понятий  $\langle c_i, r_o, c_j \rangle$ , где  $c_i$  и  $c_j$  — понятия, связанные между собой отношением  $r_o$ , принадлежащим множеству  $R$  отношений, принятых в используемой модели мира:

$$N \cong \{ \langle c_i, r_o, c_j \rangle \}. \quad (6)$$

Несимметричность семантической сети предполагает неравенство двух троек:  $\langle c_i, r_o, c_j \rangle \neq \langle c_j, r_o, c_i \rangle$ . Семантическая сеть, описанная таким образом, может быть переописана как множество звездочек:  $z_i = \langle c_i \langle r_o, c_j \rangle \rangle$ :

$$N \cong \{ z_i \} = \{ \langle c_i \langle r_o, c_j \rangle \rangle \}. \quad (7)$$

Под звездочкой  $z_i = \langle c_i \langle r_o, c_j \rangle \rangle$  в этом случае понимается конструкция, включающая главное понятие  $c_i$  связанное с множеством понятий-ассоциантов  $c_j$ , которые являются семантическими признаками главного понятия, отстоящими от главного понятия на одну связь. Связи направлены от главного понятия к понятиям-ассоциантам, и размечены отношениями  $r_o$ .

Семантическая сеть строится из таких звездочек. Второстепенное слово первой звездочки  $z_1$  становится главным во второй звездочке  $z_2$ . И так далее. В некоторый момент какое-нибудь второстепенное слово одной из предыдущих звездочек  $z_i$  становится главным в следующей звездочке (уже встречавшейся ранее)  $z_1$ . Все

вместе они формируют сеть – циклический граф. Причем частоты встречаемости характеризуют первичные весовые характеристики понятий звездочки, а связь получает метку отношения, которое ею реализуется. Таким образом, сеть становится неоднородной (размеченной), в отличие от однородной (ассоциативной) сети в разделе 1. Сеть строится с учетом анафор и сочинений в рамках абзаца таким образом, что главное слово первого предиката первого анализируемого предложения данного абзаца становится главным словом первой звездочки. Главные слова вершинных предикатов всех остальных предложений одного абзаца становятся второстепенными понятиями в этой первой звездочке и присоединяются к главному слову сочинительными (ассоциативными) связями  $r_s$ . Подчиненные предикаты присоединяются к второстепенным словам первой звездочки своими главными словами, каждый – на свое место. Связи в предикатных звездочках размечаются типами связей  $r_p$ , принятых в заданной модели предметной области  $R \in \{r_s, \{r_p\}, r_a\}$ . На свои места присоединяются атрибутивными связями  $r_a$  определяющие понятия.

Далее, также как в разделе 1, первичная сеть перенормируется. Частоты встречаемости  $p_i$  и попарной встречаемости  $p_{ij}$  перевычисляются в смысловые веса  $w_i$ . Получаем окончательную неоднородную (размеченную) семантическую сеть.

Необходимо напомнить, что связи, полученные на этапе синтаксического анализа предложений, дополняются сочинительными связями, полученными в результате раскрытия анафорических ссылок и объединения предложений абзаца как смысловой единицы.

Итак, на этом этапе обработки сначала, на основе предложений, содержащих ключевые понятия, выделенные на этапе лексической обработки, из предикатных пар формируется частотный портрет текста. Его единицами являются пары понятий, выделенные на этапе анализа предложений, а связями – попарные включения этих понятий, с учетом их связей, анафор и сочинений в рамках абзаца. После формирования частотного портрета частоты встречаемости единиц перенормируются в смысловые веса. Взвешенные по смыслу предикатные структуры ранжируются и составляют иерархию предикатных структур, описывающих предметную область.

#### 4. Прагматический анализ текста

Описанным в разделе 3 способом строятся семантические сети целых предметных областей. В этом случае сети формируются на основе корпусов текстов, описывающих предметные области.

Последующий прагматический анализ входного текста заключается в выделении цепочки наиболее весомых ключевых предикатных структур, характерных для предметной области, к которой относится текст. Объем этой цепочки зависит от порога, примененного к смысловому весу предикатных структур (выделение главных предикатных структур), а порядок этих предикатных структур – от порядка их следования в предложениях текста. Такая цепочка полностью характеризует смысловое содержание текста в терминах выбранной предметной области.

Отнесение текста к одной из предметных областей осуществляется сравнением ассоциативной сети текста и ассоциативных сетей предметных областей (классификация) [Харламов, 2008]. Степень пересечения ассоциативных сетей текста и предметной области характеризует смысловое пересечение текста и предметной области.

Последовательность главных предикатных структур, выбранных из текста, с использованием порогового преобразования, описывает основное

#### ЗАКЛЮЧЕНИЕ

Автоматический смысловой анализ текста с использованием технологии TextAnalyst позволяет выявить ключевые понятия в их взаимосвязях в тексте, а также ранжировать их по степени их смысловой значимости в данном тексте. Замена первичной обработки текста, заключающейся в морфологическом анализе слов, на первичную обработку, связанную с выявлением предикатных структур в предложениях текста, позволяет перейти от автоматического формирования ассоциативной сети текста к автоматическому формированию неоднородной семантической сети. Проекция исходного текста на сформированную семантическую сеть, с учетом смысловых весов ключевых понятий, позволяет выявить прагматическую структуру текста как последовательность наиболее существенных предикатных структур в данном тексте.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- [Ягунова Е.В., 2008] Ягунова Е.В. Вариативность стратегий восприятия звучащего текста (экспериментальное исследование на материале русскоязычных текстов разных функциональных стилей) / – Пермь: Пермский университет, 2008 - 395с.
- [convera.com, 2011] URL: <http://www.convera.com>.
- [Kharlamov, 2004] Kharlamov A.A. Networks constructed of neuroid elements capable of temporal summation of signals / Alexander A. Kharlamov, Vladimir V. Raevsky // “Neural Information Processing: Research and Development”, Jagath C. Rajapakse and Lipo Wang, Editors, “Springer-Verlag”, May, 2004 - 478 p., ISBN 3-540-21123-3.
- [Харламов, 2001] Харламов А.А. Нейросетевая технология представления и обработки информации (естественное представление знаний). / А.А. Харламов // - М.: «Радиотехника», 2006. - 89 с.
- [Gartner Symposium, 2001] Data Mining, Gartner Symposium ITXPO 2001.

[Сокирко А.В., 2001] Сокирко А.В. Семантические словари в автоматической обработке текста (по материалам системы ДИАЛИНГ) / А.В. Сокирко // Диссертация на соискание ученой степени кандидата технических наук – М.: 2001.

[Харламов, 2008] Харламов А.А. Перестройка модели мира, формируемой на материале анализа текстовой информации с использованием искусственных нейронных сетей, в условиях динамики внешней среды. / А.А. Харламов, В.В. Раевский // Речевые технологии, N 3, 2008. Стр. 27-35.

[Sullivan, 2001] Sullivan D. Document Warehousing and Textmining./ Dan Sullivan // - N.Y.: "Wiley publishing house", 2001.

[Hopfield, 1982] Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. / J.J. Hopfield // Proc. Natl. Acad. Sci. 79, 1982. Pp. 2554 – 2558.

[Белошапкова и др., 1997] Белошапкова В.А. Современный русский язык: Учебник для филологических специальностей высших учебных заведений / В.А. Белошапкова [и др.] // Под ред. Белошапковой – 3-е изде, испр. и доп. - М.: Азбуковник, 1997 - 928 с.

[Харламов и др., 2011] Харламов А.А. Метод выделения главных членов предложения в виде предикатных структур, использующий минимальные структурные схемы. / А.А. Харламов, Т.В. Ермоленко, Г.В. Дорохина, Д.С. Гнигтько // Речевые технологии, N 2, 2011 (В печати).

#### **AUTOMATICAL MAKING OF HETEROGENEOUS SEMANTIC NETWORK ON THE BASE OF EXTRACTION OF KEY PREDICATE STRUCTURES OF TEXT SENTENCES**

Kharlamov A.A. \*, Yermolenko T.V. \*\*,

*\*Institute of Higher Nervous Activity of the  
Russian Academy of Science, Moscow, Russia*  
**kharlamov@analyst.ru**

*\*\*Institute of Artificial Intelligence Problems,  
Donetsk, Ukraine*  
**etv@iai.dn.ua**

In the work a new approach to automatical extraction of semantic relationships of texts keywords is presented. The TextAnalyst technology of making semantic network is used but instead of key words the approach uses key predicate structures. This approach combines two other ones: the extraction of predicate structure of sentence as a sense of sentence; and extraction of key words of text with their relationships as a semantic network of the whole text. Their combination is a detailed sense of text which can be extracted automatically.

The main text content can be shown as a list of its key words for example. Extraction of key words from text is not a problem now [Kharlamov, 2004; Харламов, 2001]. But there are no relationships their as a very important information. Extraction of the information about relationships of key words in text can be fulfilled by using its predicate structures. Automatic extraction of predicate structures from sentences of the text is the main goal of the work.

Automatic extraction of key word from text can be

made by TextAnalyst technology of automatic semantic text analysis. The main key words with their relationships are extracted from the text by using artificial neuron network on the basis of neuroliked elements with time summation of input signals [Kharlamov, 2004]. The semantic network elements (nodes and edges) have their weights. The network is homogenous (associative) in this case.

For doing the analysis more exact we can use the information about relationships between the key words. In such a way we are making the heterogeneous semantic network. For this we can use predicate structures of the text sentences which we can extract by using special syntactical rules [Харламов, 2011].

Extracted from the text sentences predicate structures include relationships of their actants. Those relationships allow paint the edges of previously prepared semantic network.

The sequence of the predicate structures which include the main key words of the text characterize the text pragmatic.

The TextAnalyst technology allows extract automatically the main text key words in their associative relationships. By changing the text preprocessing we can automatically extract from the text not only associative but other types of relationships and do the extracted semantic network the heterogeneous one. By projecting the text onto the semantic network we can find in the text a sequence of the main predicate structures which include the text pragmatic.