

# Понимание текста как интерпретация цепочек предикатных структур важных предложений текста, полученных на этапе прагматического анализа (объединение лингвистического и статистического подходов)

*Харламов А.А., доктор технических наук, с.н.с.,*

*Ермоленко Т.В., кандидат технических наук*

В статье описан подход к представлению текста в минимальном по объёму его описании на метаязыке, на основе которого возможно восстановление текста в близком по смыслу к исходному тексту виде. Таким представлением является цепочка расширенных предикатных структур предложений текста, выделенных путём ранжирования с последующим удалением несущественных, с точки зрения текста, предложений. Подход объединяет лингвистический и статистический методы анализа смысла текста.

• полный лингвистический анализ предложения • статистический семантический анализ текста • прагматический анализ • предикатные структуры • сравнение текстов • классификация текстов.

This paper reports on an approach to presentation of a text in its minimized form in metalanguage that allows restoring a text similar to the origin. Here such text representation is a string of extended predicative structures of the text sentences, isolated by ranging and further removal of sentences insignificant according to the semantic net of the text. The extended predicative structures are a result of a comprehensive linguistic analysis of text sentences. Analysis of the semantics of the whole text is made by statistical methods.

## Введение

Смысловый анализ естественно-языкового текста в автоматизированном режиме обычно реализуется методами либо лингвистического, либо статистического анализа. В первом случае выявляется смысловая структура отдельного предложения (или нескольких следующих друг за другом предложений, связанных анафорическими, или эллиптическими связями) в виде его расширенной предикатной структуры, включающей в свой состав наряду с предикатом субъект, главный и второстепенные объекты [Харламов с соавт., 2012а; Харламов с соавт., 2012b]. Смысловый же анализ целого текста реализуется с использованием статистических подходов. В этом случае строится семантическая сеть всего текста, содержащая ключевые понятия текста (слова и устойчивые словосочетания) в их взаимосвязях [Харламов, 2006].



Под пониманием смысла текста (далее для краткости — текста) обычно понимается интерпретация содержания текста на некоем метаязыке, который, при необходимости, позволяет восстановить текст, пусть даже в другой форме, но с сохранением его смысла [Мартынов, 1977].

До настоящего времени все методы понимания текста сводились к построению семантической сети текста. Большая часть подходов к построению семантической сети текста основывалась на ее формировании экспертом (вручную) [Голенков с соавт, 2011]. Однако семантическая сеть текста, которая удобна для интерпретации смысла текста экспертом, не является представлением текста, удобным для последующего восстановления первоначального, или близкого по форме текста.

Целью исследования, результаты которого представлены далее, является разработка алгоритмов представления текста в минимальном по объему описании текста на метаязыке, на основе которого возможно восстановление текста в близком по смыслу к исходному тексту виде. Таким представлением является цепочка расширенных предикатных структур предложений текста, выделенных путем ранжирования с последующим удалением несущественных с точки зрения текста предложений этого текста.

Разделение семантического анализа на два этапа — собственно семантического анализа, а также прагматического анализа — позволяет построить некоторое представление текста, которое, с одной стороны, сохраняет существенные смысловые черты исходного текста, а с другой, позволяет восстановить текст из этого представления в близкой к исходной текстовой форме. Под семантическим анализом целого текста в этом случае понимается выявление ключевых понятий текста в их взаимосвязях в тексте с их весовыми характеристиками. Под прагматическим анализом понимается извлечение из текста некоторых высказываний, которые описывают основное содержание текста. Единицами прагматического уровня анализа текстов являются высказывания, которые есть последовательности предложений, относящиеся к одной общей теме. При этом этап семантического анализа целого текста вследствие взвешивания ключевых понятий позволяет ранжировать ключевые понятия текста и, следовательно, предложения текста, их содержащие, и, таким образом, извлекать из текста наиболее существенные с точки зрения структуры текста, предложения. На этапе прагматического анализа эти предложения объединяются в последовательности, которые характеризуют прагматику текста: это последовательности предложений, составляющие или описание чего-либо, или алгоритм реализации чего-либо. Цепочки расширенных предикатных структур, соответствующих предложениям выявленных последовательностей предложений, оказываются тем метапредставлением, которое в дальнейшем можно использовать для восстановления исходной формы текста. Как можно заметить, построение семантической сети целого текста осуществляется статистическими методами, а выявление расширенных предикатных структур предложений текста — лингвистическими, то есть в предлагаемом подходе объединяются лингвистические и статистические методы анализа текстовой информации [Харламов, 2012].

Если поименовать полученные, таким образом, расширенные предикатные структуры предложениями, из которых они были извлечены, то последовательность этих предложений и будет интерпретацией исходного текста. Можно также поименовывать цепочки расширенных предикатных структур понятиями, характеризующими описания, или алгоритмы, этими цепочками представленные. Для реализации автоматического прагматического анализа, и понимания текста используются методы обработки текста на графемати-

ческом, морфологическом и синтаксическом уровнях, а также на семантическом уровне. Задача графематического анализа заключается в разборе текстовой информации на отдельные предложения и базовые элементы (слова, сокращения, цифровые и символные комплексы и т.д.).

Морфологический анализ сводится к автоматическому определению леммы и распознаванию частей речи каждого слова текста (каждому слову ставится в соответствие лексико-грамматический класс). Синтаксический анализ заключается в автоматическом выделении синтаксических элементов предложения — именных групп, терминологических целых, предикативных основ, что позволяет сформировать расширенные предикатные структуры предложений. На этапе семантического анализа предложений текста определяется семантика отдельных фрагментов текстовой информации в виде расширенных предикатных структур. Выявление информационно-логической основы всего текста — проведение семантического анализа целого текста предполагает решение задачи выявления и оценки смыслового содержания текста в виде семантической сети, содержащей ключевые понятия в их взаимосвязях в тексте с их весовыми характеристиками. Прагматический анализ заключается в выявлении сценарной структуры текста — высказываний текста (описательных, или алгоритмических). Эта структура, будучи сформированной на основе семантического представления целого текста, характеризует динамику развития описываемых в тексте ситуаций в виде цепочек расширенных предикатных структур.

Под пониманием текста в рамках данной статьи понимается выявление важных, с точки зрения данного текста, цепочек расширенных предикатных структур текста, которые, таким образом, описывают содержание текста в терминах метаязыка предикатных структур.

## **1. Автоматический анализ текста**

Автоматический анализ текста включает в себя графематический, морфологический, синтаксический, семантический и прагматический уровни обработки.

### **1.1. Графематический уровень обработки**

На графематическом уровне анализа текст очищается от нетекстовой информации, сегментируется на слова и предложения. Единицей графематического анализа является цепочка символов, выделенная с двух сторон пробелами. Выделенная цепочка символов подвергается последовательной обработке эвристическими правилами: отсекаются знаки пунктуации, проверяется наличие гласных внутри цепочки, чередование верхнего и нижнего регистров и т.д.

Полные словоформы анализируются на морфологическом уровне лингвистического анализатора. Цель на этом уровне — разбить все множество словоформ на подмножества по признаку принадлежности к той или иной лексеме и, по возможности, однозначно определить грамматические характеристики.

### **1.2. Морфологический анализ**

Большая часть слов текста представляет неизменный фундамент языка и охватывается словарем в пределах 100 тысяч слов. Другая, более редкая, но не менее важная составляющая лексикона постоянно пополняется и в принципе не имеет чётко очерченных границ, как следствие, прежде всего в части имен собственных и словообразовательных вариантов известных слов, для морфологического анализа используются методы как с декларативной, так и с процедурной ориентацией.

Для декларативного морфологического анализа используется полный словарь всех возможных словоформ для каждого слова. При этом каждая словоформа снабжена полной и однозначной морфологической информацией, куда входят как постоянные, так и переменные морфологические параметры. Задача морфологического анализа сводится



к поиску нужной словоформы в словаре. Если слово не найдено, используются процедурные методы, где каждое слово разделяется на основу и аффикс, и словарь содержит только основы слов вместе со ссылками на соответствующие строки в словаре аффиксов [Дорохина, 2011].

### **1.3. Анализ синтаксического уровня**

Синтактико-семантический анализ предложения проводится в несколько этапов: фрагментация предложения; объединение однородных фрагментов; установление иерархии между фрагментами разных типов; объединение фрагментов в простые предложения; построение внутри фрагментов простых синтаксических групп; выявление предикативного минимума каждого из простых предложений; выделение остальных членов простого предложения, являющихся актантами выявленного предиката, построение синтаксических групп, в которых актант предиката — главное слово.

Синтаксические правила задают отношения между словами (сегментами) в предикативном виде. В зависимости от типа сегментов и типа подчинительного союза с помощью эвристических правил можно реализовать несколько операций объединения над ними: подчинение, однородность, импликацию, присоединение. В результате осуществляется разбиение сложных предложений на простые предложения, связанные сочинительными или подчинительными союзами.

Следующий шаг — построение простых синтаксических групп внутри каждого простого предложения и выделение предикатного ядра. К простым синтаксическим группам относятся группы на атрибутивном уровне, группы с предлогом и сравнительные конструкции. Множество простых предложений русского языка задается перечнем минимальных структурных схем предложений, описывающих предикативный минимум предложения. Во всех сегментах предложения, не являющихся вложенными и однородными, проводится последовательный поиск подходящего шаблона минимальной структурной схемы предложения. В соответствии с найденным шаблоном, каждому главному члену предложения присваивается соответствующее значение.

Далее решается задача получения расширенной предикатной структуры простых предложений и заполнения валентных гнезд предиката [Харламов с соавт., 2012а]. Выделение остальных членов простого предложения (остальных семантически значимых объектов и атрибутов) проводится с помощью последовательного сравнения слов предложения с актантной структурой глагола, для чего используется словарь валентностей глаголов.

### **1.4. Семантический анализ целого текста**

Статистический подход позволяет выявить семантическую структуру целого текста — глобальную внутреннюю структуру текста как однородную (ассоциативную) семантическую сеть, вершинами которой являются ключевые понятия текста (слова и устойчивые словосочетания), а рёбрами — их смысловые (ассоциативные) отношения в тексте, например, опосредованные их попарной встречаемостью в смысловых фрагментах текста. И ключевые понятия, и их связи взвешиваются их численными характеристиками — смысловыми весами.

Для корректной обработки текста, с целью формирования семантической сети, предварительно необходимо осуществить его (текста) развертывание: по возможности, эксплицирование анафор и эллипсисов.

Статистический смысловой анализ текста реализован на основе технологии обработки текстовой информации TextAnalyst [Харламов, 2006], позволяющей автоматически выявлять ключевые понятия в тексте на основе только информации о структуре самого текста (независимо от предметной области). Для этого формируется 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 (например, в предложениях). Частотный портрет, таким образом, содержит информацию о частоте встречаемости понятий и их попарной (в терминах их ассоциативной связи) встречаемости в тексте. Использование хопфилдоподобного алгоритма позволяет перейти от частоты встречаемости к смысловому весу (вес связей при этом остается неизменным).

В результате такой итеративной процедуры перенормировки наибольшие веса получают понятия, связанные с наибольшим числом других понятий с большим весом, то есть те понятия, которые стягивают на себя смысловую структуру текста.

Вследствие такой обработки из ассоциативной частотной сети текста 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, например, предложении).

Иначе семантическую сеть можно представить в виде множества звёздочек  $\langle c_i, c_j \rangle$ :

$$N \cong \langle c_i, c_j \rangle \rangle, \quad (1)$$

где  $\langle c_j \rangle$  — множество ближайших ассоциантов ключевого понятия  $\langle c_i \rangle$  (понятий, напрямую связанных с данным понятием).

Полученные смысловые веса ключевых понятий показывают значимость этих понятий в тексте. В дальнейшем эта информация используется для выявления предложений текста, содержащих наиболее существенную с точки зрения структуры текста информацию в тексте. Для этого на основе смысловых весов понятий, входящих в предложения текста, вычисляются смысловые веса предложений. Предложения затем ранжируются с удалением из текста предложений, имеющих вес ниже порогового.

### 1.5. Прагматический уровень анализа

*Определение.* В данном разделе под предикатом будем понимать тройку  $\langle c_i, r_{ij}, c_j \rangle$ , где  $c_i$  — субъект,  $r_{ij}$  — отношения, размечающие связи субъекта с главным объектом  $c_1$  и другими объектами  $c_j, j > 1$  — актантами предиката. Причём,  $r_{i1} \cong r_p$ , где  $r_p$  — предикатное отношение,  $c_1$  — главный объект:

$$P \cong \langle c_i, r_{ij}, c_j \rangle. \quad (2)$$

Под прагматическим анализом в данной работе будем понимать выявление сценария текста (корпуса текстов), представленного в виде цепочки (цепочек) расширенных предикатных структур, соответствующих предложениям текста (корпуса текстов описывающих предметную область) оставшимся после удаления предложений незначительной смысловой значимости в тексте. Сценарий описывает динамику развития представленной в тексте (корпусе текстов) ситуации. Такая цепочка может быть описательной, или алгоритмической. В первом случае сценарий характеризует восприятие, во втором — действие. Прагматическому анализу обязательно предшествует семантический анализ: до выявления прагматики текста необходимо сформировать семантическую модель предметной области, или семантическую модель текста, на ключевые понятия которой в дальнейшем проецируется входной текст.

Прагматический анализ текста, таким образом, заключается в выявлении цепочек предикатных структур предложений, которые на этапе семантического анализа целого текста



оказались наиболее весомыми в рамках предметной области, к которой относится текст. Степень важности предложений текста определяется с учетом степени важности ключевых слов, которая определяется их ранжированием в рамках семантической сети предметной области на этапе семантического анализа. Количество этих предикатных структур зависит от порога, примененного к смысловому весу предикатных структур (предложений, их содержащих), а порядок этих предикатных структур в цепочках – от порядка следования оставшихся после ранжирования и порогового преобразования предложений в тексте. Такие цепочки полностью характеризуют смысловое содержание текста (корпуса текстов — предметной области).

Цепочки расширенных предикатных структур, представленных их субъект-объектными парами  $W_i = (\langle c_i, r_{ij}, c_j \rangle)$ , соответствуют последовательностям их содержащих предложений текста в порядке их появления в тексте. Множество текстов, описывающих предметную область, порождают множество таких цепочек предикатных структур  $W_i | i = 1..|I|$ .

### 1.6. Понимание текста как поименование цепочек предикатных структур текста предложениями текста, их содержащими

Понимание конкретного текста связано с выявлением предикатных структур  $P_i = \langle c_i, r_{ij}, c_j \rangle$ , характеризующих смысл предложений этого текста, а также цепочек этих предикатных структур  $W_i = (\langle c_i, r_{ij}, c_j \rangle)$ , которые опосредуют смысл отдельных последовательностей предложений текста. Любой текст данной предметной области, порождающий цепочку предикатных структур, таким образом, может быть проинтерпретирован как последовательность предложений текста, их содержащих. Под пониманием текста в данном случае будем понимать проекцию цепочек предикатных структур текста на множество соответствующих цепочек предикатных структур предметной области и поименование этих цепочек соответствующими им предложениями.

Интерпретация в виде последовательности предложений является удобной для представления эксперту и может быть использована для общения с конечным пользователем в случае необходимости удобного интуитивно понятного представления результатов автоматического смыслового анализа текстов.

## 2. Пример прагматического анализа текста

Для иллюстрации описанного в работе способа понимания текста рассмотрим некоторый русскоязычный текст по курсу физики [http://www.kodges.ru/Т.И. Трофимова. Курс физики, Москва: Высшая школа, 2001]:

«Глава 2. Динамика материальной точки и поступательного движения твердого тела.

§ 5. Первый закон Ньютона: всякая материальная точка (тело) сохраняет состояние покоя или равномерного прямолинейного движения до тех пор, пока воздействие со стороны других тел не заставит её изменить это состояние. Первый закон Ньютона выполняется не во всякой системе отсчета, а те системы, по отношению к которым он выполняется, называются инерциальными системами отсчета...».

После сегментации текста в процессе графематического анализа на слова и смысловые фрагменты и после проведения морфологического анализа слов проводится в несколько этапов синтактико-семантическая обработка предложения. В том числе, в предложениях текста выявляются и раскры-



ваются анафорические ссылки. После разрешения анафор текст примера будет выглядеть так:

«Первый закон Ньютона: всякая материальная точка (тело) сохраняет состояние покоя или равномерного прямолинейного движения до тех пор, пока воздействие со стороны других тел не заставит **точку** изменить состояние **покоя или равномерного прямолинейного движения**. Первый закон Ньютона выполняется не во всякой системе отсчета, а те системы, по отношению к которым **закон** выполняется, называются инерциальными системами отсчета».

Результат анализа для нашего примера приведён в таблице 1.

Таблица 1

Предикативный минимум простых предложений, входящих в состав предложений исходного текста

№ предл.	Составляющие простые предложения	Шаблон минимальной структурной схемы предложения	Предикативный минимум (субъект-предикат)
1	Первый закон Ньютона	Существительное в именительном падеже	закон <b>NULL</b>
1	Всякая материальная точка (тело) сохраняет состояние покоя или равномерного прямолинейного движения	Существительное в именительном падеже+спрягаемая форма глагола	точка сохраняет
1	До тех пор пока воздействие со стороны других тел не заставит точку изменить состояние покоя или равномерного прямолинейного движения	Существительное в именительном падеже+спрягаемая форма глагола+инфинитив	воздействие не заставит изменить
2	Первый закон Ньютона выполняется не во всякой системе отсчета	Существительное в именительном падеже+спрягаемая форма глагола	закон выполняется
2	А те системы, по отношению к которым закон выполняется, называются инерциальными системами отсчета	Существительное в именительном падеже+спрягаемая форма глагола	системы называются

Выделение остальных членов расширенной предикатной структуры предложения осуществляется последовательным сравнением слов предложения с актантной структурой глагола из словаря валентностей глаголов. Заполненные валентные гнезда для предикатов текста примера приведены в таблице 2.

Таблица 2

Заполнение валентных гнезд для предикатов текста примера

№ предл.	Предикат	1. Субъект	2. Объект	3. Адресат	4. Инструмент	5–7. Локативы
1	сохраняет	точка	<b>состояние</b>	–	–	–
1	<b>не</b> заставить	воздействие	–	–	–	–
1	изменить	–	<b>состояние</b>			
2	выполняться	закон				системе
2	называться	системы			системами	

Далее, на основе статистического семантического анализа целого текста, выявляется множество ключевых понятий текста. Осуществляется построение ассоциативной сети текста путем объединения ключевых понятий с учётом их попарной встречаемости в тексте. И в процессе итеративной процедуры осуществляется перенормировка частот



встречаемости ключевых понятий в смысловый вес ключевых понятий, являющихся вершинами семантической сети. Смысловые веса предложений вычисляются как нормированные суммы смысловых весов составляющих предложения ключевых понятий. Пример смысловых весов предложений приведён в таблице 3.

Таблица 3

Смысловой вес предложений (фрагмент)

	Предложение	Смысловой вес
1	Динамика материальной точки и поступательного движения твердого тела	67
2	Динамика является основным разделом механики, в ее основе лежат три закона Ньютона, сформулированные им в 1687 г.	15
3	Законы Ньютона играют исключительную роль в механике и являются (как и все физические законы) обобщением результатов огромного человеческого опыта.	15
4	Первый закон Ньютона: всякая материальная точка (тело) сохраняет состояние покоя или равномерного прямолинейного движения до тех пор, пока воздействие со стороны других тел не заставит ее изменить это состояние.	99
...		

Предложения текста ранжируются по смысловому весу путём сравнения их смыслового веса с заранее заданным пороговым значением. Удаляются предложения, которые имеют смысловый вес ниже порогового (порог, например, равен 50).

Оставшиеся предложения могут быть использованы для построения квазиреферата (сценария) текста в виде цепочки предложений текста, соответствующих выявленным расширенным предикатным структурам текста. Фрагмент такого квазиреферата выделен жирным шрифтом ниже на фоне исходного текста:

**«Глава 2. Динамика материальной точки и поступательного движения твердого тела.**

§ 5. Первый закон Ньютона. Масса. Сила.

Динамика является основным разделом механики, в её основе лежат три закона Ньютона, сформулированные им в 1687 г. Законы Ньютона играют исключительную роль в механике и являются (как и все физические законы) обобщением результатов огромного человеческого опыта. Их рассматривают как систему взаимосвязанных законов и опытной проверке подвергают не каждый отдельный закон, а всю систему в целом. **Первый закон Ньютона: всякая материальная точка (тело) сохраняет состояние покоя или равномерного прямолинейного движения до тех пор, пока воздействие со стороны других тел не заставит её изменить это состояние.** Стремление тела сохранять состояние покоя или равномерного прямолинейного движения называется инертностью. Поэтому первый закон Ньютона называют также законом инерции.

Механическое движение относительно, и его характер зависит от системы отсчета. **Первый закон Ньютона выполняется не во всякой системе отсчета, а те системы, по отношению к которым он выполняется, называются инерциальными системами отсчета. Инерциальной системой отсчета является такая система отсчета, относительно которой материальная точка, свободная от внешних воздействий, либо покоится, либо движется равномерно и прямолинейно».**



Ниже, в Таблице 4 представлены расширенные предикатные структуры оставшихся предложений в порядке их следования в тексте.

Таблица 4

Расширенные предикатные структуры оставшихся предложений в порядке их следования в тексте

№ предл.	Предикат	Субъект	Связь	Объект	Связь	Атрибут	Адресат Инструмент Локативы	Связь	Атрибут
1	NULL	динамика	R_G	точка	R_A	материальный			
			R_G	движение	R_A	поступательный			
					R_G		тело	R_A	твердый
2	1.NULL	Закон Ньютона			R_A	первый			
	2.сохранять	точка			R_A	всякий			
					R_A	материальный			
			R_O	состояние			покой		
							движение	R_A	прямолинейный
								R_A	равномерный
	3.не заставить изменить	воздействие	R_G	сторона	R_G		тело	R_A	другой
			R_O				точка		
			R_O				состояние	R_A	это
...									

Таблица 5

Обозначение связей в предикатной структуре предложения, использованных в таблице 4

Обозначение связи	Синтаксическая группа
<b>Валентные гнезда предиката</b>	
R_S	Предикат-Субъект
R_O	Предикат-Объект
R_I	Предикат-Инструмент
R_L	Предикат-Локатив
<b>Связи на атрибутивном уровне</b>	
R_A	Объект-Признак объекта
R_A_P	Действие-Признак действия
<b>Синтаксические группы актантов</b>	
R_G	Генитивное определение в постпозиции

### Обсуждение результатов

Рассмотренный пример прагматического анализа (на фоне всех остальных этапов анализа от графематического до семантического) показывает, что использованный подход является достаточно трудоёмким, так как требует осуществления полного лингвистического анализа текста до синтактико-семантического анализа отдельного предложения, включительно. В сравнении с этим, статистический семантический анализ целого тек-



ста реализуется сравнительно просто. Однако вычислительная сложность такого подхода к прагматическому анализу текстов оправдана. С одной стороны, это позволяет представить текст в виде компактного метаописания в виде цепочек расширенных предикатных структур, которое может быть раскрыто в обратную сторону в некоторое текстовое представление, близкое по смыслу к исходному тексту, то есть реализовать понимание текста. Это может быть эффективно использовано, например, для формирования синтезируемого ответа в системах речевого диалога, а также для сопоставления с аналогичным метапредставлением на другом языке в системах контекстного перевода. С другой стороны, это же метаописание может быть интерпретировано эксперту как последовательность предложений текста (квазиреферат), и такое представление весьма удобно пользователю, как обычный естественно-языковой текст.

Кроме того, представление текста в виде цепочек расширенных предикатных структур позволяет улучшить смысловое сравнение текстов путем выявления степени пересечения множеств таких цепочек двух сравниваемых текстов. То же можно сказать о классификации текстов: сравнение таких множеств цепочек для исходного текста и рубрик (классов текстов, описывающих предметные области) позволяет отнести текст к одному или нескольким классам в зависимости от степени пересечения соответствующих множеств цепочек.

### Заключение

В статье представлен метод прагматического анализа текста с использованием объединенного полного лингвистического и статистического семантического подходов к анализу текста. Представленный метод позволяет представить текст минимальным множеством прагматических цепочек — цепочек расширенных предикатных структур предложений, несущих максимальный смысл этого текста. Цепочки расширенных предикатных структур представляют собой некоторое компактное метаописание смысла текста, которое может быть интерпретировано предложениями естественно-языкового текста. Такое описание может быть использовано для автоматического смыслового сравнения текстов и для автоматической классификации текстов.

### Литература

1. [Харламов с соавт., 2012а] Харламов А.А., Ермоленко Т.В., Дорохина Г.В., Гнидько Д.С. Метод выделения главных членов предложения в виде предикатных структур, использующий минимальные структурные схемы. Речевые технологии. № 2. 2012.
2. [Харламов с соавт., 2012б] Бондаренко Е.А., Каплина О.А., Харламов А.А. Предикатные структуры в системе машинного распознавания текста Речевые технологии. № 4. 2012.
3. [Харламов, 2006] Харламов А.А. Нейросетевая технология представления и обработки информации (естественное представление знаний). М.: Радиотехника, 2006. 89 с.
4. [Мартынов, 1977] Мартынов. Универсальный семантический код / Минск: Наука и техника, 1977. 192 с.
5. [Голенков с соавт., 2011] Голенков В.В., Гулякина Н.А. Графодинамические модели параллельной обработки знаний: принципы построения, реализации и проектирования // Труды Международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» (OSTIS'2011), Минск: 2011

6. [Харламов, 2012] *Харламов А.А.* Способ автоматизированного восстановления сценария текста на естественном языке. Заявка на патент на изобретение № 2012153432 от 12 декабря 2012 г.
7. [Дорохина, 2011] *Дорохина Г.В.* Автоматическое выделение синтаксически связанных слов простого распространенного неосложненного предложения / Г.В. Дорохина, Д.С. Гнидько // «Сучасна інформаційна Україна: інформатика, економіка, філософія»: матеріали доповідей конференції, 12–13 травня 2011 року, Донецьк, 2011. Т. 1. 34–38 с.
8. *Hopfield, J.J.* Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. 79, 1982. P. 2554–2558.

### Сведения об авторах

***Харламов Александр Александрович*** —

доктор технических наук, старший научный сотрудник Института высшей нервной деятельности и нейрофизиологии РАН,

***Ермоленко Татьяна Владимировна*** —

кандидат технических наук, научный сотрудник отдела распознавания речевых образов Института проблем искусственного интеллекта МОНМС и НАН Украины (г. Донецк).