

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin  
(editors)

# **Natural and Artificial Intelligence**

**ITHEA**

**SOFIA**

**2010**

**Krassimir Markov, Vitalii Velychko, Oleksy Voloshin (ed.)**

**Natural and Artificial Intelligence**

ITHEA®

Sofia, Bulgaria, 2010

ISBN 978-954-16-0043-9

First edition

Recommended for publication by The Scientific Council of the Institute of Information Theories and Applications FOI ITHEA

This book is engraved in prof. Zinovy Lvovich Rabinovich memory. He was a great Ukrainian scientist, co-founder of ITHEA International Scientific Society (ITHEA ISS). To do homage to the remarkable world-known scientific leader and teacher this book is published in Russian language and is concerned to some of the main areas of interest of Prof. Rabinovich.

The book is opened by the last paper of Prof. Rabinovich specially written for ITHEA ISS. Further the book maintains articles on actual problems of natural and artificial intelligence, information interaction and corresponded intelligent technologies, expert systems, robotics, classification, business intelligence; etc. In more details, the papers are concerned in: conceptual problems of the natural and artificial intelligent systems: structures and functions of the human memory, ontological models of knowledge representation, knowledge extraction from the natural language texts; network technologies; evolution and perspectives of development of the mechatronics and robotics; visual communication by gestures and movements, psychology of vision and information technologies of computer vision, image processing; object classification using qualitative characteristics; methods for comparing of alternatives and their ranging in the procedures of expert knowledge processing; ecology of programming – a new trend in the software engineering; decision support systems for economics and banking; systems for automated support of disaster risk management; and etc.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

General Sponsor: Consortium FOI Bulgaria ([www.foibg.com](http://www.foibg.com)).

Printed in Bulgaria

**Copyright © 2010 All rights reserved**

© 2010 ITHEA® – Publisher; Sofia, 1000, P.O.B. 775, Bulgaria. [www.ithea.org](http://www.ithea.org); e-mail: [info@foibg.com](mailto:info@foibg.com)

© 2010 Krassimir Markov, Vitalii Velychko, Oleksy Voloshin – Editors

© 2010 Ina Markova – Technical editor

© 2010 For all authors in the book.

© ITHEA is a registered trade mark of FOI-COMMERCE Co.

**ISBN 978-954-16-0043-9**

C/o Jusautor, Sofia, 2010

## ИСТИНА ИЛИ СМЫСЛ? ДВА ПОДХОДА К ОБРАБОТКЕ ЕСТЕСТВЕННО ЯЗЫКОВЫХ ТЕКСТОВ

Леонид Святогор, Виктор Гладун

**Аннотация:** *Выполнен сравнительный анализ двух принципиально разных подходов к семантической обработке текстов и дискурсов: логическая интерпретация и онтологический анализ. В зависимости от семантической сложности текста устанавливается сфера применимости каждого направления. Подробно рассматривается метод онтологического анализа, который опирается на систему знаний и формализованную категорию «смысл». Уточняется понятие «машинное понимание текстов ЕЯ».*

**Ключевые слова:** *логический анализ, онтология, смысл, текст, машинное понимание*

**ACM Classification Keywords:** *1.2.4 Knowledge Representation Formalisms and Methods; 1.2.7 Natural Language Processing – Text analysis*

---

### Введение

Логический и содержательный анализ текстов естественного языка (ЕЯ) не могут быть противопоставлены друг другу: каждый соотносится со своей областью применения. Если логическая обработка, в принципе, устанавливает непротиворечивость высказываний «логическим знаниям» о предметной области (ПрО) и их выводимость из системы аксиом, то в задаче семантического анализа слабо-формализованных и разнообразных по тематике текстов ЕЯ на первый план выдвигается понимание авторского (коммуникативного) замысла, извлечение его из текстового материала и лексическое оформление.

С позиций искусственного интеллекта Д.А. Поспелов выделил пять уровней машинного понимания текстов различной степени семантической сложности. На первом и втором уровне востребуются только те термины, которые содержатся в тексте непосредственно или могут быть получены из текста путём его логического расширения. На третьем уровне для понимания текста нужно привлечь некоторую дополнительную информацию, которую нельзя вывести логическими методами. Четвёртый и пятый уровни требуют, чтобы к интерпретации текста были привлечены такие понятия, которые в тексте отсутствуют, но накоплены в базах знаний о мире, социуме и используются в предметных областях человеческой деятельности.

В последних двух вариантах толкование текста требует привлечения широкого контекстного материала: знаний о целях коммуникации, о социальной среде, её моральных и культурных ценностях и других системных характеристик.

**Цель и содержание данной работы** состоят в том, чтобы проанализировать и сравнить между собой два различных подхода к семантической обработке текстов – логическую интерпретацию и онтологический анализ. На основании сравнения проявляется сфера приложения каждого направления. Подробно рассматривается метод онтологического анализа, который опирается на систему знаний и формализованную категорию «смысл». Уточняется понятие «машинное понимание текстов ЕЯ».

---

## ТЕКСТ и ДИСКУРС

---

Объектом семантического разбора является текст или дискурс. Необходимо обозначить рамки этих понятий следующими вопросами: Что такое текст? Каково его функциональное назначение? Что мы хотим из него извлечь? Последний вопрос является кардинальным с прагматической точки зрения.

Для выяснения этих вопросов дадим следующее **определение**. *Текст есть лексическая, грамматически и тематически оформленная, конечная форма передачи замысла автора партнёру по коммуникации – в общей знаковой, языковой и знаниевой среде.*

Современная теоретическая лингвистика опирается на семь критериев текста, сформулированных в 1981 г. Бограндом и Дресслером: 1) связность, 2) целостность, 3) интенциональность, 4) приемлемость, 5) информативность, 6) ситуационность и 7) интертекстуальность [Тичер и др., 2009]. В русле данной работы нас будут интересовать следующие параметры текста: связность, целостность (тематическое единство) и интенциональность (авторский замысел).

В данном исследовании никаких ограничений на тематическое содержание текстов не накладывается. Большинство исследователей считают *функциональным назначением текста* фиксацию и передачу рациональных и культурных знаний в процессе коммуникации людей и сообществ. Текст рассматривают как вербализованную и закодированную средствами языка мысль автора (или его замысел).

На практике сложилось нечёткое разделение текстов на два вида: те, которые функционируют в узкой специализированной предметной области (ПрО) или в документообороте, и другие, неформатные тексты с *широким спектром* семантических отношений, например – дискурсы. Первый вид характеризуется небольшим количеством семантических отношений и достаточно высокой структурируемостью данных. *Дискурс* определяется как «*связный текст в контексте многих конституирующих и фоновых факторов – социокультурных, психологических и т.д.*» [Штерн, 1998, с. 87]. Он характерен локальными (микро) и глобальными (макро) семантическими структурами, причём последние характеризуются сложными, нелинейными схемами их *смыслового взаимодействия*, которые зависят от *авторского замысла*.

Существо дискурса, его суть определяются не столько количеством включённых в ситуацию объектов, сколько переплетением связей между ними. Связи эти имеют преимущественно ассоциативный характер. В многообразии непредсказуемых отношений между семантически оформленными элементами заключена сложность текста для его восприятия.

Под *сложным текстом* мы будем понимать такой, который нельзя представить конечным числом предикативных отношений в силу существования сюжетных, кореферентных и имплицитных связей между отдалёнными частями (смысловыми блоками) текста. Именно референциональные связи «*отражают коммуникативную установку автора, его интенции, фонд знаний партнёра, коммуникативную организацию высказывания, отношение к контексту*» [Штерн, 1998, с. 272]. Связи многообразны и трудно-перечислимы, не всегда могут быть установлены логическими методами, а их совокупность вместе с объектами воссоздаёт фрагмент реальной или ментальной действительности.

Отсюда вытекают трудности принципиального характера, когда мы хотим извлечь из текста его главное содержание, то есть – выявить ту коммуникативную цель, которую преследует автор. Предполагая, что эта цель выражена в тексте некоторым «*СМЫСЛОМ*», и что вербализованный смысл отделён от автора, мы тем самым признаём *объективность* данной категории. «*Смысл существует сам по себе как инвариант, который выясняется (устанавливается) во время перевода или пересказа текста*» [Штерн, 1998, с. 286]. Более того, чётко утверждается, что «*сам текст на естественном языке является*

лишь проводником смысла, а замысел и знания автора лежат во вторичной структуре (смысловой структуре или макроструктуре текста) [Гаврилова, Червинская, 1992, с. 104].

Если считать, что текст служит, образно говоря, «контейнером» мысли и смысла, то необходимо подчеркнуть, что данный контейнер «погружён в море знаний». Это означает, что толкование текста невозможно без привлечения терминов и понятий, которые в самом тексте не упоминаются, но существуют в обиходе коммуникации; они называются *контекстным знанием*. Сложный текст может быть осмыслен, то есть – понят только в контексте широких языковых знаний об окружающей его среде.

---

## ЛОГИКА

---

### Отношение логики к смыслу и смысла – к логике

Будем исходить из предпосылки, что в классической логике значение истинности любого сложного высказывания, которое построено из простых (аксиоматических) высказываний и логических связок *и*, *или*, *не* – есть функция истинности простых высказываний, причём аргументы и функции заданы на множестве значений (*истина*, *ложь*). Центральной проблемой теории является проверка *выводимости* утверждений и их логической *непротиворечивости*. Хотя сами по себе логические высказывания инвариантны к содержательной (смысловой) стороне терминов, предполагается, что от бессмысленных высказываний должны уберечь «правильные» аксиомы, поскольку на стадии их формирования закладывается семантическая связь с внешним миром, а также «корректные» формальные правила вывода (причём обязаны также соблюдаться «правила применения правил»). Всё это, однако, при конечной системе аксиом и при достаточно длинном выводе не убергает – в момент перехода с логического языка на естественный – от появления *«смысловых парадоксов»*.

Несмотря на это, в логике встречается мнение, что в процессе формального вывода сохраняется *«связь по смыслу»* [Одинцов, 1996; Рубашкин, 1989]. Рассмотрим пример. Возьмём две аксиомы, не вызывающие никаких возражений: *«В огороде – бузина»* и *«В Киеве – дядько»*. Объединив их импликацией *если – то*, получим безукоризненно истинное высказывание. Однако порождённая здесь логическая истина демонстрирует как раз факт отсутствия смысла. Разумеется, здесь недостаёт аксиомы, связывающей объекты двух посылок, но мы это обнаружили уже за рамками логики («слишком поздно»). В данном примере «связь по смыслу» вывела нас за пределы смыслового поля.

Проблема, однако, заключается не в том, что между *истиной* и *смыслом* нет *однозначного соответствия*, а в том, что такого соответствия нельзя требовать. *Смысл вообще не является компетенцией логических исчислений*, и в логике это понятие не имеет определения. Смысл, понимаемый как совокупность неформальных связей между частями дискурса, *не измеряется двоичной переменной*. Считать смыслом *имя предиката* не корректно – это имя получает смысловую интерпретацию только в естественно-языковом выражении. В свою очередь, *логические связки и, или, не* также не имеют никакой другой интерпретации, кроме «математической» – это просто табличные числовые функции. На пути превращения логики Аристотеля в формальную систему, как отмечает Д.А. Поспелов, *«исчезла логика и победила математика»* [Поспелов, 1982, с. 31].

Не всё благополучно обстоит также и с аксиоматикой. Попытка придать семантику даже простому аксиоматическому утверждению обходится дорого: приходится уточнять понятия, затем – уточнять вторичные понятия, и этот процесс лавинообразно нарастает, вовлекая в себя всё новые пласты знания [Рубашкин, 1989]. Именно поэтому *«Логик пока не идёт дальше рассмотрения узкоспециализированных понятийных систем»* [там же, с. 185].

Однако мы вправе требовать связи истинности со смыслом в ином – этическом плане. Как отмечают психологи, человек *«выделяет особый смысл, который связан с ориентацией на истину как на*

бытийную ценность и является нормативным по отношению к профессиональной деятельности учёного» [Балл, 2006, с. 110].

Сформулируем кратко итог. **Истинность есть мера выводимости высказываний, а смысл есть мера знания о мире.** С позиций искусственного интеллекта можно считать, что категория смысла универсальнее истины.

#### Что мы хотим извлечь из текста?

В конечном итоге, все методы семантического анализа направлены на формализацию и обработку знаний. Однако при решении задач семантической обработки текстов вырисовывается принципиальная разница не столько в методах, сколько в целевой установке обработки знаний: если логический аппарат ориентирован на вывод и получение *новых* знаний, то семантический анализ текстов нацелен на извлечение *готовых* знаний и их перекрёстный анализ. При этом, как указывалось, *истинность или ложность обнаруженных в тексте конкретных знаний не являются целью семантической обработки*: приоритетная цель заключается в установлении смысловых отношений между лексическими элементами, кодирующими знание.

В зависимости от степени формализации текста и сложности мы можем извлечь из него либо ту информацию, которая в нём содержится – в терминах непосредственно составляющих, либо другую информацию, которая раскрывает содержимое терминами контекстного знания. В последнем случае необходимо построить систему контекстного знания – онтологию. Как будет показано ниже, целью семантической обработки является извлечение из текста и дискурса онтологического смысла. Тем самым категория «смысл», извлечённая из ментальной сферы, приобретает определённость и конструктивность, что позволяет вести обработку машинными средствами. Однако, как справедливо указывает *И.Б. Штерн*, «*рассматривая понимание естественного языка как оперирование смыслом, прежде всего надо разрешить принципиальные проблемы, связанные с определением сущности и структуры смысла*» [Штерн, 1998, с. 286]. Один из возможных способов решения этих проблем предложен в следующих разделах.

Краткий вывод. В тех текстах, где данные могут быть структурированы и ограничены по множеству объектов и отношений (например, в рамках ПрО), целью диалога или поиска служит извлечение данных. В сложных текстах требуется определить смысл как цель коммуникации.

В следующем разделе рассмотрен пример логико-предикативной модели представления знаний, используемой для машинного диалога. Он необходим для сопоставления с онтологической моделью.

#### Принципы логической обработки знаний

Типичные черты предикатно-логического подхода присутствуют в поисковой системе, подробно и ясно описанной в работе [Одинцов, 1996]. Предметной областью служит организация производства, снабжения и сбыта в металлургической отрасли Украины (фрагмент отрасли).

Информация о ПрО представлена в *базе знаний* в виде семантических структур **фактов**, причём данные структуры явно определяют стандартные связи между объектами. Фактом называют *локализованную минимальную совокупность понятий естественного языка, отражающих связь между классами объектов, классами и их элементами, а также между элементами различных классов*. Факты называются также суждениями или высказываниями. Классы объектов перечислены в *базе данных*. Семантическая структура факта представлена деревом, в узлах которого находятся понятия, а дуги отражают семантические отношения между ними. Одновременно такая графическая структура транслируется в предикатную форму.

Запрос, который формулируется на естественном языке с незначительными ограничениями на синтаксис, преобразуется через предикатную форму в семантическую структуру, подобную структурам базы знаний. После этого производится **распознавание** структуры запроса в базе знаний. При совпадении структура базы знаний преобразуется в предикат, а последний транслируется в ответ на естественном языке.

Поисковая система поддерживает составные и сложные вопросы. Она может не только продублировать вопрос с выдачей искомым данных, но и представить прогноз относительно ресурсов, выполнения плана и другие знания.

Как утверждает автор рассматриваемой работы, предикат типа: **«Находится(предприятие (А), город(В))»** «связывает понятия «предприятие» и «город» в единое, имеющее смысл, целое» [Одинцов, 1996, с. 19]. С этим можно согласиться, если рассматривать данную ситуацию как «функциональное отношение между предикатом и актантами, которое отражает семантическую структуру данной ситуации» [Штерн, 1998, с. 265]. И при этом добавить, что *смысл целого* проверяется за пределами логики.

В описанной системе логика предикатов, дополненная структурами знаний, служит естественным аппаратом для проверки истинности выводимых утверждений: предметная область в понятийном плане ограничена, данные структурированы, предикатные отношения перечислимы (сведены в таблицу), синтаксис упрощён. Однако, «возникают огромные технические трудности, которые перерастают в принципиальные препятствия, как только модель мира становится достаточно объёмной, а число тем, по которым человек хочет вести диалог с ЭВМ – велико» [Поспелов, 1982, с. 192].

В новейших работах логико-семантического направления наблюдается стремление больше *сблизить логику с семантикой*. Для этого, например, в работе [Никитченко, 2008] в теорию вводятся «модели миров». Успех в данном направлении, по нашему мнению, может быть достигнут, если номинативные структуры логики будут связаны с системой репрезентации знаний, а выводимость будет означать обнаружение формальных смысловых связей. Тогда семантическая нагрузка может быть возложена в моделях миров на интенциональные и экстенциональные связи, а прерогативой логики останется вывод и синтез сложных семантических понятий из простых.

---

## ОНТОЛОГИЯ И СМЫСЛ

---

### Онтологическая картина мира

Наша задача заключается в том, чтобы **перевести понятие «смысл» из ментальной или подразумеваемой категории – в конструктивную**. Это возможно сделать лишь в том случае, если смысл будет включён в формальную систему репрезентации знаний – семантическую сеть; наиболее пригодной для наших целей является **онтология** [Святогор, Гладун, 2009].

Известны онтологии двух типов – абстрактные и предметные онтологии. *Абстрактные онтологии* (Дж. Совы, СУМО, Микрокосмос) создаются с философских позиций, чтобы ответить на вопрос: какие категории познания создаёт человек для понимания универсума – природы, космоса, материи. Однако на уровнях межличностного и социального общения человек, стремясь быть понятным, предпочитает структурировать мир иначе. В этом случае для передачи коммуникативных знаний подходят *предметные онтологии*, которые лучше передают ситуации и события в окружающей среде. Предметная онтология способна сочетать в себе, в терминологии Ю.Д. Апресяна, «научную» и «наивную» картины мира.

Предметная лексическая онтология представляет собой семантическую сеть, в которой узлам соответствуют языковые понятия, описывающие объекты или состояния – *концепты*, а связи отображают различные *отношения* между ними. В соответствии с законами познания внешнего мира сеть

имеет *иерархическую* архитектуру, где чем больше уровень общности понятия, тем более высокое место в структуре оно занимает. «На стадии концептуализации эта иерархическая структура превращается в пирамиду знаний» [Гаврилова, Червинская, 1992, с. 44]. Формальные концепты и отношения сети имеют под собой мощную *систему интерпретации* на естественном языке.

Для репрезентации глобальной системы всеобщих знаний, ориентированной на анализ текстов с произвольной тематикой, разработана предметная онтология **ИО\*3**; она подробно и вместе с примерами описана в работе [Гладун и др., 2008].

### **Трёхуровневая иерархическая онтология ИО\*3**

Основными принципами построения трёхуровневой иерархической онтологии – **ИО\*3** являются:

- структурирование мира на основе материалистической парадигмы академика *В.И. Вернадского*;
- введение *трёх уровней* иерархии («слоёв знаний»), внутри которых развиваются многочисленные подуровни связанных концептов;
- выбор в качестве *концептов*: имён, именующих понятия, объекты, ситуации, состояния и атрибуты, и сложившихся в языковой среде общего знания и общения;
- выбор в качестве *концептуальных связей*: формальных, атрибутивных и ассоциативных отношений;
- задание *интерпретирующей системы* на множестве декларативных знаний, хранящихся в учебниках, энциклопедиях и толковых словарях;
- замыкание концептов на *словарь* естественного языка.

Верхний уровень иерархии образует дерево, вершиной которого служит глобальный концепт *Материя*. Далее он раскрывается через понятия: *Косное вещество*, *Живое вещество*; последнее разделяется на *Биосферу* и *Ноосферу*, и далее деление продолжается, переходя к концептам среднего уровня. Здесь находятся понятия, которые сложились в науке для изучения природы, человека и общества по отраслям знаний. Наконец, общенаучные понятия раскрываются на нижнем уровне через конкретные характеристики и отношения бытового уровня, которые могут быть спрогнозированы в корпусе текстов. Кроме того, на третьем уровне формулируются и подключаются к среднему уровню концепты предметных областей, которые могут, в принципе, представлять любые области знаний и деятельности человека.

Построение пирамиды онтологии подчинено чёткому принципу репрезентации знаний через *интенционалы* и *экстенционалы*. Каждое понятие в семантике рассматривается в единстве его этих двух категорий. Это значит, что каждый интенционал должен быть раскрыт соответствующим ему по семантическому отношению экстенционалом. Практически это означает, что каждый концепт, выступающий в роли интенционала или экстенционала, имеет *значение* из множества **имён** объектов или ситуаций, которые имеют *денотат* в реальном мире и лексический *смысл* в естественном языке [Шевченко, 2003]. В онтологии **ИО\*3** в роли экстенционала могут выступать: *части понятия* (при классификации), его *признаки* (при атрибуции), *значения* (при уточнении понятия), *примеры* (в качестве представителей класса), а также *ассоциативные понятия* (по усмотрению эксперта). Очевидно, что при движении по вертикали пирамиды (вниз) экстенциональные компоненты становятся интенциональными по отношению к собственным определителям. Экстенционалы (имена концептов) взаимодействуют со словарём естественного языка.

Для того, чтобы придать пирамиде знаний конструктивные, процедурные свойства, воспользуемся понятием «графа концептуальных зависимостей», введенного *Р. Шенком*. Уникальным свойством данной конструкции является тот факт, что «**граф концептуальных зависимостей служит метаязыком для внутренней смысловой репрезентации текстов, который отображает смысловую структуру ситуации**» [Штерн, 1998, с. 197].

Иерархическую трёхуровневую онтологию – **ИО\*3** будем считать ориентированным *концептуальным графом*, все пути на котором ведут из любого концепта к глобальной вершине *Материя*.

Относительно количества концептов и внутренних связей онтологии никаких ограничений не делается: её структура является *расширяемой* за счет добавления новых знаний, в том числе – разных ПрО.

Концептуальный онтологический граф **ИО\*3** имеет следующие преимущества, которые выделяют его среди других систем репрезентации знаний:

- граф служит инструментом *формализации* понятия «смысл текста естественного языка»;
- он позволяет свести процедуру семантического анализа текста к *извлечению смысла* и реализует данную процедуру *стандартными графическими операциями*;
- допускает *машинное представление* и машинную обработку текстовых знаний.

Семантические процедуры извлечения из текста его глубинного содержания будем в дальнейшем называть *онтологическим анализом*. Результатом онтологического анализа служит *онтологический смысл*. В следующем разделе показано, каким образом онтологический анализ (выделение смысла) может быть реализован алгоритмически.

### **Онтологический смысл: формальное представление**

В искусственном интеллекте задача извлечения знаний из текстов конкретизируется как «*понимание и выделение смысла текста*» [Гаврилова, Червинская, 1992, с. 104]. Как было показано выше, фундаментальной основой для выделения смысла может служить онтология. Однако без формального определения смысл остаётся объектом ментального рассмотрения в психологии и литературоведении. Попытка построить лингвистическую модель «*Смысл – Текст – Смысл*» [Мельчук, 1999] не привела к конструктивным результатам, поскольку, используя формальные семантические, синтаксические и морфологические структуры, не опиралась на систему знаний. Поэтому нашей первой задачей было дать формальное определение этой категории мышления и сделать её *вычислимой*.

Полное рассмотрение дано в работе [Святогор, Гладун, 2009]; здесь приводятся только основные определения.

Определение 1. *Конструктом смысла* является пара концептов, связанных между собой формальным, атрибутивным или ассоциативным отношением.

Определение 2. Онтологический смысл строится из конструктов смысла.

**Онтологическим смыслом некоторого концепта является подграф концептуального графа онтологии ИО\*3, определяемый начальным и финальным концептами.** Указанный подграф называется *смысловой траекторией* начального концепта.

Определение 3. Значением начального концепта смысловой траектории служит *ключевое слово* текста; значением финального – вершина онтологии: *Материя*.

Определение 4. Совокупность смысловых траекторий всех ключевых слов текста формально определяет *онтологический смысл текста*.

Таким образом, смысловой анализ начинается с извлечения из текста очередного ключевого слова. Основанием для такой операции служит определение, предложенное в работе [Гаврилова, Червинская, 1992, с. 109]: «*Набор ключевых слов – это набор опорных точек, по которым развёртывается текст при кодировании в память и осознаётся при декодировании, это семантическое ядро цельности*». Со своей стороны, мы опираемся на гипотезу, что ключевые («значимые») слова являются основным семантическим инструментом автора текста для кодирования своего замысла.

Таким образом, **формальный онтологический смысл** получается в результате проекции дискретизированного текста на онтологию **ИО\*3**. Результатом отображения в сети каждого ключевого слова служит **смысловая траектория** – дискретная упорядоченная цепочка взаимосвязанных концептов, которая «считана» с подграфа онтологического графа. Задание онтологического смысла в таком формате пригодно для компьютерного представления и «машинного понимания».

Онтологический смысл отличается от других известных определений, а тем более – от интуитивного понимания смысла, тем, что он *задан в многомерной координатной системе представления знаний о внешнем мире и имеет – в фиксированной онтологии – однозначное графическое представление.*

---

## МАШИННОЕ ПОНИМАНИЕ

---

### Понимание и связность

Как отмечалось выше, многие исследователи в области искусственного интеллекта определяют **понимание реалии** как «*постижение её смысла через конституирующие её контексты*» [Штерн, 1997, с. 92]. «*Процесс понимания текста заключается в извлечении смысла, заложенного в текст, и именно текст является носителем этого смысла ... Рассматривая понимание естественного языка как оперирование смыслом, прежде всего необходимо разрешить принципиальные проблемы, связанные с определением сущности и структуры смысла*» [Штерн, 1998, с. 285].

На пути разрешения этих принципиальных проблем понимания естественного языка нами предложено понятие онтологического смысла. Подчеркнём ещё раз, что онтологический смысл не есть отображение высказывания на множество *ноль* и *единица*. Это есть *отображение высказывания на связанное множество именованных концептов, репрезентующее систему знаний* – онтологию.

В самом общем случае **понимание** является целью коммуникации и прерогативой двух интеллектов. На ментальном уровне процесс понимания означает возбуждение в памяти человека моделей реальности (языковых и образных), нахождение аналогий или закрепление новых связей; на машинном уровне факт «компьютерного понимания» означает возбуждение на графе онтологии соответствующего подграфа смысла и связывание его с другими подграфами. В том и другом случае реконструируется некоторая языковая ситуация (реалия).

Соединение онтологии со смыслом позволяет соединить *семантику и прагматику текста в единый процедурный комплекс*. Действительно, онтология знаний отвечает за связь текста с внеязыковой действительностью, а смысловая траектория отвечает за кодирование онтологического смысла в машинной памяти. Такой синтез семантики и прагматики позволяет нам дать следующее определение:

***машинным пониманием текста называется процесс построения семантических траекторий, возбуждаемых текстом на графе онтологии, которая во внутренней памяти машины выполняет роль системы репрезентации знаний о внешнем мире.***

Таким мы видим результат онтологического анализа ЕЯ тестов.

Ещё один полезный результат от введения онтологического смысла состоит в уточнении термина **связность текста**. Связный текст есть функция связанной мысли. Однако, будучи отделён от мысли, текст, особенно дискурс, имеет дискретную тематическую или сюжетную структуру. В литературоведении под связностью (например, повести) понимают возможность наблюдать за развитием выделенного сюжетного фрагмента в контексте полного сюжета, отслеживая определённую тему на фоне других. Связность дискурса реализуется авторскими приёмами, чаще всего через имена действующих лиц, место и время действия. На уровне формальной логики связность текста полностью достигается переносом

слова из предыдущего предложения в последующее, т.е. связность умозаключений является, в некотором смысле, синонимом выводимости.

Благодаря формализации смысла в системе знаний можно ввести конструктивное понятие **связность текста по онтологическому смыслу**. Два ключевых слова в тексте порождают две смысловые траектории, которые либо *сходятся* в одну точку – на вершине пирамиды, либо *пересекаются* на одном из трёх уровней онтологии **ИО\*3**. Первый общий концепт этих траекторий говорит о том, **что** связывает два ключевых слова на концептуальном уровне. Именно этот общий концепт *именует ту ситуацию*, в которой ключи вступают в первое взаимодействие. Чем выше в пирамиде происходит замыкание траекторий, тем менее связным является изложение.

Краткие выводы. На уровне графических структур мы соединили семантику с прагматикой, «заставив» **машину понимать смысл**. «Понимание» и «смысл» соотносятся как синонимы. Связность текста устанавливается через пересечение траекторий смысла.

---

### **Заключение. ЗАЧЕМ НАМ НУЖЕН ОНТОЛОГИЧЕСКИЙ СМЫСЛ ?**

---

Не отрицая ментальное происхождение смысла, которое следует из утверждения, что *«смысловым содержанием фраза наполняется в психическом процессе»* [Шевченко, 2003, с. 25], мы можем дополнить: **«смысловое содержание фразы выявляется в онтологическом процессе машинного понимания текста»**.

Какие последствия может иметь введение «онтологического смысла» для теории и практики?

В теоретическом плане – открыта научная методология глубинного анализа естественно-языковых текстов сложной семантики через синтез онтологического смысла. **Смысл становится вычисляемой семантической функцией текста и дискурса**. В поле знаний об искусственном интеллекте разработана цельная концепция: «текст – знания – смысл», а в инженерии знаний сделан шаг по **«превращению искусства в ремесло»**.

В практическом плане – извлечение из текста онтологического смысла позволяет разгрузить пользователя сети Интернет от рутинной работы по анализу потоков ненужной (нерелевантной запросу) информации. «Смысловой фильтр» не только «отсеет мусор», но представит текст в сжатом виде, сохранив только его смысл. Благодаря этому в системах поддержки принятия решений руководитель или менеджер получает возможность быстро вникнуть в суть проблемы – **не читая сообщение полностью, а просмотрев его «смысловой портрет»**.

В перспективе пользователь сможет построить собственную онтологическую сферу его интересов и в рамках **«персональной онтологии» знаний** оперировать с актуальной информацией на смысловом уровне.

---

### **Благодарности**

---

Статья напечатана при финансовой поддержке проекта **ITHEA XXI** Института Информационных теорий и Приложений FOI ITHEA Болгария ([www.ithea.org](http://www.ithea.org)).

---

### **Литература**

---

- [Тичер и др., 2009] С. Тичер, М. Мейер, Р. Водак, Е. Веттер. Методы анализа текста и дискурса \* Пер. с англ. – Х.: Изд-во Гуманитарный Центр, 2009. – 356 с.
- [Штерн, 1998] Штерн І.Б. Вибрані топіки та лексикон сучасної лінгвістики. Енциклопедичний словник. – К.: «АртЕк», 1998. – 336 с.
- [Гаврилова, Червинская, 1992] Гаврилова Т.А., Червинская К.Р. Извлечение и структурирование знаний для экспертных систем. – М.: Радио и связь, 1992. – 200 с.

- [Одинцов, 1996] Одинцов Б.Е. Пресуппозиция в ассоциативных сетях и машинное понимание текстов. – Львов, 1996. – Препринт. Львовский госуниверситет им. И. Франко. – 60 с.
- [Рубашкин, 1989] Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах. – М.: Наука, 1989. – 192 с.
- [Поспелов, 1982] Поспелов Д.А. Фантазия или наука: на пути к искусственному интеллекту. – М.: Наука, 1982 – 224 с.
- [Балл, 2006] Балл Г.А. Психология в рациогуманистической перспективе: Избранные работы. – К.: Изд-во Основа, 2006. – 408 с.
- [Никитченко, Шкільняк, 2008] Нікітченко М.С., Шкільняк С.С. Математична логіка та теорія алгоритмів: підручник. – К.: Видавничо-поліграфічний центр «Київський університет», 2008. – 528 с.
- [Святогор, Гладун, 2009] Л. Святогор, В. Гладун. Определение понятия «Смысл» через онтологию. Семантический анализ текстов естественного языка. – International BookSeries, Number 9. IntelligentProcessing. Supplement to the International Journal "Information Technologies & Knowledge" Volume 3 / 2009. – ITHEA, Sofia, 2009. – p. 53.
- [Гладун и др., 2008] В. Гладун, В. Величко, Л. Святогор. Структурирование онтологии ассоциаций для конспектирования естественно-языковых текстов. – International BookSeries, Number 2. Advanced Research in Artificial Intelligence. Supplement to the International Journal "Information Technologies & Knowledge" Volume 2 / 2008. – ITHEA, Sofia, 2008. – p. 153.
- [Шевченко, 2003] Шевченко А.И. Актуальные проблемы теории искусственного интеллекта. – «Наука і освіта», 2003.
- [Мельчук, 1999] Мельчук И.А. Опыт теории лингвистических моделей «Смысл – Текст». – М.: Школа «Языки русской культуры», 1999. – 346 с.
- [Штерн, 1997] И. Штерн. Интродуктивные модели гуманитарных знаний: концептуальне гештальты versus понятия \ KDS–97. Шестая Международная конференция «Знания – Диалог – Решение». Сборник научных трудов в двух томах. Том 1. – Ялта, 1997. – Стр. 89.

---

### **Информация об авторах**

---

**Гладун Виктор Поликарпович** - *Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mail: [aduis@rambler.ru](mailto:aduis@rambler.ru)*

**Святогор Леонид Александрович** - *Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mail: [aduis@rambler.ru](mailto:aduis@rambler.ru)*