

---

---

## Intelligent NL Processing

---

---

### СЕМАНТИЧЕСКИЙ АНАЛИЗ ТЕКСТОВ ЕСТЕСТВЕННОГО ЯЗЫКА: ЦЕЛИ И СРЕДСТВА

Леонид Святогор, Виктор Гладун

**Аннотация:** В данной работе предлагается расширенное толкование понятия «текст естественного языка» и предлагается схема полного освоения его семантического ресурса за счёт «компьютерного понимания» и диалога. Указываются средства достижения указанной цели в процессе семантической обработки текстов – использование трёхуровневой онтологии для извлечения из текста онтологического смысла, а также ввод обратной связи для дополнительного уточнения в диалоге содержания дискурса.

**Ключевые слова:** семантический анализ текста, онтология, смысл, диалог.

**ACM Classification Keywords:** 1.2.7 Natural Language Processing - Text analysis

**Conference:** The paper is selected from XV<sup>th</sup> International Conference “Knowledge-Dialogue-Solution” KDS-2 2009, Kyiv, Ukraine, October, 2009.

---

#### Введение

---

**В начале и в конце семантического анализа естественно-языковых текстов стоит Слово.** Методы анализа разнообразны и зависят от решаемой в прикладной области задачи, и существует не одно направление обработки текстовой информации. В условном разделении можно выделить методы семантической обработки текстов, которые нацелены на «лингвистические преобразования», например – перевод на иностранный язык и обратно; краткий пересказ; конспектирование; тезисное представление; аннотирование и на решение других прагматических задач. С другой стороны, у исследователей искусственного интеллекта интерес к тексту лежит в области «извлечения знаний» – классификация сообщений, ответы на вопросы, контекстный перевод и понимание дискурсов [Sowa, 2002]. Здесь применяются методы концептуального анализа. При этом можно заметить оформление двух проблем: (а) синтез систем представления знаний – онтологий и (б) разработка систем семантического анализа и машинного «понимания» текстов при помощи онтологий.

Проблема (а) решается широким фронтом; из последних, практически успешных работ можно указать на исследование [Палагин и др., 2009], где из корпуса профессиональных текстов автоматически извлекается подструктура знаний в одном из разделов предметной области (ПрО) «Материаловедение». Для синтеза онтологии используются формально-логические и синтаксические средства анализа.

Следует заметить, что «конкурирующим» подходом может служить разработка структур знаний при помощи экспертов и инженеров по знаниям. В этом случае готовой базой для разработчиков онтологий

---

служат учебники, свежие публикации и другие пособия по описанию ПрО [Поспелов, 1988; Гаврилова и Хорошевский, 2001; Ной и МакГиннесс, 2001].

В проблеме (б) наш подход состоит в следующем [Гладун и др., 2008; Святогор и Гладун, 2009].

Если *описание* ситуации, изложенной в тексте, может быть достигнуто чисто лингвистическими средствами, то *понимание* ситуации возможно за рамками лингвистического ресурса текста – мобилизацией когнитивных усилий человека и его индивидуальных знаний. Например, как отмечает Г.С. Поспелов, связное восприятие текста возможно лишь при его понимании.

Аналогично тому, как человеческое понимание рождается при согласовании внешней информации с его ментальной (когнитивной) моделью мира, «компьютерное понимание» может быть достигнуто отображением информации на определённую и формально-заданную систему знаний. Проще говоря, чтобы «понимать» что-то, надо его «узнавать». В машинной обработке текстовой информации роль памяти человека выполняет компьютерная система формальной репрезентации знаний – онтология: именно она позволяет совместить анализ текста с его компьютерным «пониманием». Процедура это достигается достаточно просто: необходимо найти проекцию текста на компьютерную онтологию.

**Говорить о «понимании информации» можно лишь в контексте окружающего её знания.**

Конкретная задача искусственного интеллекта состоит в следующем. Задан текст ЕЯ, или сообщение. То, «о чём говорится», можно назвать *темой, содержанием, интенцией сообщения* или *коммуникативным смыслом*; этот смысл требуется из предъявленного текста *извлечь* [Мальковский, 1985; Штерн, 1998, стр.145].

Проверка «качества понимания», или релевантности текста извлечённому смыслу, происходит за рамками онтологии, например – экспертной оценкой смысла или по результату принятого решения.

---

### Актуальность и цель работы

---

Машинное понимание языка является парадигмой искусственного интеллекта.

Кардинальным вопросом остаётся один: что мы хотим получить от текста?

Мы хотим получить его формальный (компьютерный) смысл. Для этого рассматривается структура Системы семантического анализа ЕЯ текстов, в которой должны быть предусмотрены и объединены базовые процедуры: грамматический *анализ*, взаимодействие текста с *онтологией*, получение результата – формального *понимания* текста через онтологию и, наконец, – уточнение его *смысла*. В конечном итоге, предлагаемая технология семантического анализа преследует цель – добиться лучшего взаимного понимания автора текста и его потребителя через компьютер, общую базу знаний и родной язык.

---

### Взгляд на семантический ресурс ЕЯ текста

---

Потенциальные возможности текстового документа намного выше тех, что мы используем, и для того, чтобы «понимать» текст, нужно вначале выяснить его потенциальные возможности.

Текст рассматривается не только как вместилище информации – данных, фактов и знаний, которые требуется из него извлечь. *Он представляет собой языковое, информационное и культурное явление*, которое актуально для данного периода существования социума и может быть востребовано потомками. Отсюда следует, что текстовый материал изначально, априори «погружён» в некоторую общечеловеческую систему накопления и интерпретации знаний, в которой он сам был *порождён*. С другой стороны, текст генерируется как индивидуально, так и коллективно и может быть *востребован* также индивидуально или коллективно. Это означает, что содержание («семантическое наполнение»)

материала часто является *многоплановым*, и каждый план имеет свою глубину изложения. Выполняя свою коммуникационную функцию, *текст обязан быть понимаемым пользователем, то есть – должен отображаться в базе знаний потребителя* и взаимодействовать с ней.

Семантический компонент текста ЕЯ давно был зафиксирован лингвистами, которые определили предложение как выражающее законченную мысль. Какую мысль? Как её кратко и неискажённо сформулировать? Поиски ответа продолжаются в рамках когнитивной лингвистики [Штерн, стр.129], психологии [Балл, 2006], искусственного интеллекта [Поспелов, 1988]. В произвольном дискурсе мысль облачена в лексическую оболочку в соответствии с правилами грамматики, и в ряде случаев эту оболочку надо «сбросить». В интересующих нас случаях документ или сообщение несёт актуальную информацию или стабильное знание. **Концентрированное выражение знания мы называем смыслом.** Это концентрированное знание, или смысл, надо извлечь, чтобы потом с ним оперировать. Следовательно, текст нужно хранить и беречь, поскольку он содержит определённый авторский замысел, интеллектуальный ресурс и в социогуманистическом плане является продолжением баз знаний.

Признавая ёмкий семантический ресурс текста, мы приходим к выводу, что во многих приложениях его потребительская ценность быстро не исчерпывается. Извлечение смысла вряд-ли является одноразовой операцией коммуникации. Можно надеяться на то, что раскрыть содержательный ресурс текста полностью (если это возможно!) удастся многократным к нему обращением. В этом состоит **активная функция** документа, которая используется в некоторых приложениях для глубинного семантического анализа [Поспелов, 1988; Штерн, стр.70].

---

### Какие выводы следуют из расширенного толкования понятия «Текст»?

---

Решению задачи полного раскрытия семантического ресурса текста, на наш взгляд, способствует такая Система семантического анализа ЕЯ текстов (Система), которая удовлетворяет следующим требованиям:

*Первое.* Партнёры интеллектуального общения вместе с текстом погружены в единую компьютерную среду онтологического знания.

*Второе.* Предварительная лингвистическая обработка исходного текста (морфологический, синтаксический и семантический анализ предложений) необходима для снятия «лексической оболочки» и выделения термов, несущих содержательную нагрузку.

*Третье.* Результатом компьютерного семантического анализа связного текста должен быть формальный или адаптированный текст ЕЯ, который выражает его смысловое содержание.

*Четвёртое.* Система должна обеспечивать самоконтроль авторского намерения – насколько адекватно он выражает свои мысли.

*Пятое.* Система должна многократно активизировать текст с целью более глубокого проникновения в смысл сообщения.

В результате самого общего взгляда на желаемые качества Системы семантического анализа можно сделать вывод, что потенциальные возможности текста реализуются при помощи двух механизмов: **анализа через онтологию и активного диалога.**

---

### Средства обеспечения интеллектуальных функций Системы семантического анализа

---

*Первое требование* – взаимопонимание партнёров коммуникации – обеспечивается единой системой представления общих и профессиональных знаний, накопленных в социуме. В качестве контекстной среды общения предлагается формальная семантическая сеть – **иерархическая трёхуровневая**

**онтология** (см. ниже), сформулированная в работе [Гладун и др., 2008], которая может быть расширена и дополнена спектром любых предметно-ориентированных онтологий [Поспелов, 1988; Гаврилова и Хорошевский, 2001; Палагин и др., 2009].

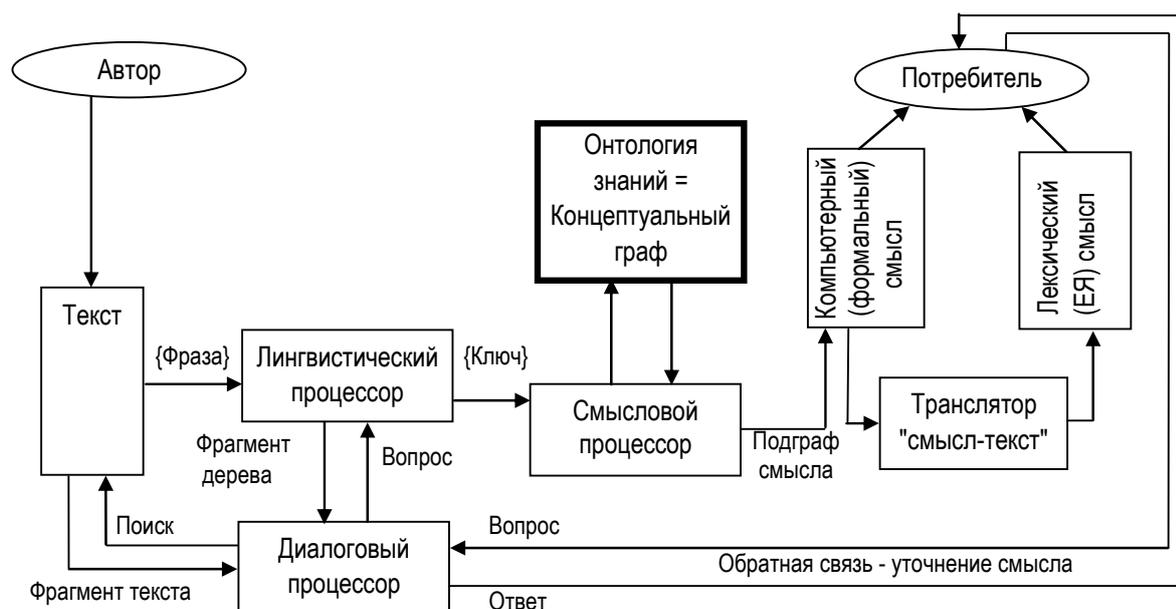
*Второй тезис* – предварительный разбор текста – выполняется **лингвистическим процессором**, ориентированным на семантический анализ обычной текстовой информации [Поспелов, 1988; Штерн, стр.261]. Например, в лингвистической модели «Смысл–Текст» используются пять уровней репрезентации выражения, включая синтаксический, морфологический и семантический уровни [Мельчук, 1999]. В самом простом случае от лингвистического процессора требуются: построить дерево синтаксического разбора, выделить ядерные конструкции предложений, построить отношения, определить «значимые» лексические группы, в частности – ключевые слова текста [Палагин и др., 2009].

*Третье условие* означает, что из текста необходимо извлечь его смысловое содержание. Смысл связного текста, согласно его *определению* в работе [Святогор и Гладун, 2009], формализуется через онтологию – как совокупность подграфов концептуального графа. Задача выявления смысла в некотором текстовом фрагменте (и в целом тексте) возлагается на **смысловой процессор**. Ниже приводится разъяснение онтологического смысла и кратко описывается автоматическая процедура его получения.

*В четвёртом требовании* предусмотрена сервисная возможность **коррекции текста его автором**. При желании он сравнивает результат автоматического выделения смысла со своими внутренними намерениями.

Наконец, для более глубокого раскрытия смысла, уточнения фактов и других данных (*пятое условие*) предусмотрен режим диалога, который реализуется на естественном языке **диалоговым процессором**. В процессе диалога смысл может существенно измениться, что, в свою очередь, может служить поводом для корректировки онтологии (активность текста).

Как смысловой, так и диалоговый процессоры выполняют интеллектуальную миссию, привлекая внелингвистические знания. К ним следует добавить вспомогательный **транслятор «смысл – текст»**, он (в необходимых случаях) поможет человеку более содержательно истолковать формальный подграф смысла.



**Блок-схема Системы семантического анализа ЕЯ текстов**

На рисунке показана блок-схема Системы семантического анализа ЕЯ текстов, в которой взаимодействуют указанные выше функциональные компоненты.

После установления структуры Системы дадим краткое описание её составных частей, имеющих – в контексте данной работы – принципиальное значение.

---

### **Иерархическая трёхуровневая онтология внешнего мира**

---

В общем случае в искусственном интеллекте понятие «картина мира» появляется как синоним понятий «модель мира», «модель предметной области» [Штерн, стр.156]. В области *Knowledge representation* исследователями предложено немало онтологических систем репрезентации концептуальных знаний о мире (известные онтологии Дж. Совы, Микрокосмос, РуТЕЗ\* и другие). Обладая мощным философским и лексическим потенциалом, общие онтологии скорее относятся к области гносеологических моделей, чем к системам, пригодным для семантического анализа текстов живого языка: взаимодействие онтологий с реальным текстом, так же как и результаты онтологической работы, остаются не определёнными.

В предлагаемой структуре Системы семантического анализа ЕЯ текста онтологии отводится ведущая роль. Как система отображения общественного интеллекта она позволяет интерпретировать текстовую информацию на языке общих знаний и объединяет тройку «автор – текст – потребитель» в единую интеллектуальную среду. Любое явление может быть понято и интерпретировано *только в контексте общепринятого и стабильного знания*.

Базисом Системы служит новая иерархическая трёхуровневая онтология – **ИО\*3** [Гладун и др., 2008]. Она отличается двумя особенностями: **(а)** сетевая структура даёт принципиальную возможность объединить – в рамках единой конструкции – знания высшего уровня абстракции, общедоступные (повседневные и актуальные) знания среднего уровня и профессиональные знания нижнего уровня; **(б)** одновременно она ориентирована на работу с конкретными текстами. Кроме того, показано, что **(в)** результатом извлечения из текста знаний должен быть «*онтологический смысл*». Этот смысл поддаётся строгой формализации и компьютерной обработке [Святогор и Гладун, 2009].

**Иерархическая трёхуровневая онтология ИО\*3** представляет собой семантическую сеть в форме концептуального ориентированного графа.

**«Концептуальный граф** – это способ семантической (понятийной) репрезентации ситуаций и знаний в моделях понимания естественного языка, принятия решений, рассуждений и т.д.» [Штерн, стр.195]. Узлами онтологического графа (*онто-графа*) служат лексические единицы – слова естественного языка, которые трактуются как **категории и понятия**. «Целью категоризации является пояснение нового через известное и структурирование мира при помощи обобщений» [там же, стр.160]. Понятием служит «*денотат некоторой сущности или явления, который кодируется языковым знаком*». Категории и понятия называют **концептами**, которые в лексической семантике выступают в роли элементов «понятийного метаязыка» [там же, стр.191]. По образному определению Дж. Совы «*существует мост между языком и приложением (ПрО). Он не нуждается в полной информации о языке и приложении, но он должен содержать крючки (зацепки), которые привязывают языково-зависимые слова к языково-зависимой грамматике и к языково-независимым, но зависимым от приложений, концептуальным структурам*» [Sowa, 2002].

**Онтология ИО\*3** организована как «пирамида концептуальных знаний». Концепты обладают разной степенью обобщения. Наиболее абстрактные категории образуют верхний уровень онтологии; в соответствии с парадигмой академика В.И. Вернадского о биосфере и ноосфере, это – *Материя, Вещество, Жизнь, Разум...* Концепты среднего уровня образуют описательный континуум знаний. Они

---

раскрывают значения категорий верхнего уровня через более употребительную в актуальной деятельности общества лексику, например: *Время, Движение, Порядок, Человек, Общество, Организация, Развитие, Управление, Транспорт, Биология, Борьба за существование* и другие. На нижнем уровне пирамиды знаний располагаются концепты двух типов: часть из них обозначают обиходные понятия повседневной жизни, привычные объекты и ситуации (*комната, ложка, верх...*), другие концентрируются вокруг профессиональных знаний ПрО (*концепт, отношение, онтология...*). Множество концептов ПрО может быть пустым.

Узлы онто-графа соединяются формальными и неформальными (*ассоциативными*) связями. Ориентация связей в графе направлена сверху вниз – от концептов более высокого уровня обобщения к концептам, которые их характеризуют.

Контакт текста с онтологией происходит на нижнем уровне **ИО\*3**; соответствующая процедура описана ниже.

Почему разработке «хорошей онтологии» придаётся столь большое значение? Ответ можно найти в принципиально важном утверждении Р. Шенка: **«Метаязыком для внутренней смысловой репрезентации текстов является граф концептуальных зависимостей, который отображает смысловую структуру ситуации»** [Шенк, 1980].

---

#### **Фантом «компьютерного смысла»**

---

Понятие «смысл ситуации» является очень ёмким и трактуется в разных дисциплинах по-разному. Первоначально понятия «смысл» и «мысль» имели ментальный характер – это показывает семантическая близость слов **«СМЫСЛ»** и **«СО-МЫСЛЬ»**. К формализации данного понятия первыми подошли лингвисты: модель И. Мельчука «Смысл–Текст» признаётся классикой лингвистической теории.

В обширной лингвистической литературе имеется много толкований понятия **«СМЫСЛ»**. Смысл соотносится с целями коммуникации [Штерн, стр.185]; трактуется как «структура ситуации» [там же, стр.197]; связывается с категорией «понимания» текста [там же, стр.283] и т.д. Смысл может быть описан совокупностью **денотатов**, освобождённых от эмоциональных, модальных, прагматических, стилистических и других оттенков [там же, стр.81]. Недостаток подобных определений – в их чрезвычайно слабой формализации.

С другой стороны, потребности автоматизации процессов семантического анализа ЕЯ текстов стимулируют развитие вычислительной лингвистики и компьютерных моделей [Демьянков, 1985; Мальковский, 1985], а также компьютерную поддержку этих процессов [Zaboleeva–Zotova & Orlova, 2009]. В конечном итоге методы вычислительной лингвистики сводятся к морфологической интерпретации, операциям над синтаксисом и семантикой, а в практическом плане решают частные задачи. Выделить именно **продукт семантической обработки** – универсальный и стандартизованный результат, который затем можно использовать в совершенно разных приложениях, – такая цель не преследуется.

Фантом «компьютерного смысла текста» ускользает.

Вместе с тем потребности автоматического анализа ЕЯ текстов настоятельно требуют ответа на вышеприведенный кардинальный вопрос: что может служить универсальным продуктом семантической обработки текста? Разделяя позицию Р. Шенка, мы в качестве такого универсального продукта вводим в рассмотрение **«формализованный смысл»**.

На этом пути может материализоваться и оформиться «компьютерный смысл».

---

## Онтологический смысл

---

На помощь в определении понятия «смысл» приходит онтология **ИО\*3**.

Идея состоит в том, что если «пропустить» текст через онтологию, которая является структурой знаний, то на выходе получим **концентрированное знание**, которое коррелирует с текстом. Концептуальный фильтр онтологии даст на выходе концептуальный, или *онтологический смысл*.

В работе [Святогор и Гладун, 2009] приводится формальное определение «онтологического смысла»; здесь оно повторяется тезисно.

Задан концептуальный онтологический граф **ИО\*3**.

**Элементарный смысл** определяется как пара соединённых соседних узлов онтологического графа. Связи не обязательно именуется, они могут лишь фиксировать факт некоторого взаимодействия двух слов (например, *ворона–птица, пассажир–самолёт, развитие–прогресс*). Онто-граф состоит из множества связанных между собою элементарных смыслов, которые вступают в дозволенные комбинации. Связная часть онто-графа, соединяющая два удалённых узла, образует **подграф**; при изменении в нём стрелок на противоположные (снизу – вверх) получается *цепочка подграфа*.

Цепочка связанных элементарных смыслов, которая начинается в некотором «активном» узле и заканчивается в вершине онтологии, образует *онтологическую цепочку активного узла*. Цепочка, выделяемая активным узлом на онтологическом графе, трактуется как *смысловая траектория* и называется **онтологическим смыслом активного слова**.

**Комментарий.** Процесс возбуждения смысловой траектории начинается с того, что в предложении из ядерной конструкции выделяется некоторое «ключевое слово». Если оно присутствует в онтологическом графе, то активное слово возбуждает соседний концепт, возбуждение передаётся дальше на высшие уровни онтологии – вплоть до вершины пирамиды. Результатом процесса является цепочка, то есть – *дискретная упорядоченная последовательность взаимосвязанных концептов; она является формальным онтологическим смыслом входного слова в заданной «картине знаний»*.

**Пример.** В ядре предложения выделены ключи: **ворона** и **сыр**. Для них будут построены соответствующие концептуальные цепочки: 1) **ворона** – Птица – Полёт – Движение – Биосфера – Жизнь – Материя и 2) **сыр** – Еда – Жизнь – Материя.

Связи пока-что не интерпретируются – в данной репрезентации онтологического смысла они не имеют значения: необходимо и достаточно зафиксировать *только связь пары объектов*. Далее, поскольку ключи находятся внутри одного предложения, то автоматически будет построена связь, ранее в онтологии отсутствовавшая: **ворона – сыр**. В итоге, линейный формат онтологического смысла будет следующий:

**ворона (птица, полёт, движение, биосфера, жизнь, материя) –  
– сыр (еда, жизнь, материя).**

Семантическое пояснение суммарной цепочки, в случае необходимости, может быть дано позже, а связи могут быть интерпретированы *повторным обращением* к тексту (ворона **имеет** сыр).

В зависимости от целей семантического анализа цепочки можно укорачивать за счёт абстрактных категорий. Кроме того, цепочки могут быть как линейными, так и разветвлёнными, то есть – с присоединением узлов, примыкающих к траектории.

Результатом полного просмотра текста является множество – «пучок смысловых траекторий», который можно трактовать как «семантический портрет текста».

**Онтологический смысл** может быть **целью и результатом** семантического анализа ЕЯ текста благодаря таким свойствам:

– ключевые слова в смысловой цепочке извлекаются непосредственно из текста;

– эти слова помещаются в контекст общих знаний, которые организованы как концептуальная смысловая среда (онтология);

– множество смысловых цепочек даёт краткое, дискретное и формализованное описание текста (фрагмента текста) – «семантический портрет текста» в терминах общих знаний.

Функцию выделения в тексте значимого слова выполняет лингвистический процессор. Функция выделения в онтологическом графе смысловой траектории возлагается на смысловой процессор. Функцию адаптированного представления смысла выполняет транслятор «смысл – текст» (см. рисунок).

Онтологический смысл, извлечённый из текстового документа компьютерной системой, становится **элементом Базы знаний**, которая доступна всем партнёрам коммуникации.

---

### Понимание связного текста

---

Определив формально онтологический смысл, мы можем в прагматическом плане говорить о «компьютерном понимании текста». Под этим будем подразумевать, что: **(а)** компьютер выявляет «замысел автора», интерпретирует его внутри собственной системы знаний и **(б)** производит над смыслом определённые операции, в том числе – даёт ему естественно-языковую грамматическую интерпретацию [Штерн, стр. 60; Мальковский, 1985; Демьянков, 1985].

Предложенный аппарат компьютерного понимания включает триаду **«Онтология – Текст – Смысл»**. Графические структуры триады непосредственно поддерживают процесс понимание текста, потому что семантическая сеть в явном виде способна отвечать на основные вопросы понимания, например: **«Что связывает кошку с мышкой?»**, **«Что общее существует между человеком и птицей?»**, **«Чем отличается человек от птицы?»**. Подчёркнём, что ответы даются не в текстовом лексиконе, а на концептуальном уровне общего знания – на метаязыке онтологии.

В то же время говорить о понимании авторского замысла, имея дело с многоплановым и лексически избыточным текстовым документом, довольно сложно.

Как указывалось, онтологический смысл, который является продуктом семантического анализа полного текста, реализуется в виде пучка траекторий, мощность которого зависит от длины текста. Траектории активизируются ключевыми словами и «накрывают» весь текст дискретно, но не хаотично. Последовательность развития ситуации сохраняется.

Траектории упорядочены по времени их появления, они семантически взаимосвязаны одной фразой, абзацем, разделом. Отдельные ключевые слова вступают во взаимодействие через общие концепты более высокого уровня. Отдельные траектории пересекаются и частично сливаются, причём их концепты пересекаются в разных комбинациях и с разной частотой. Этот сложный механизм структурно и схематично отражает, в принципе, всю семантическую сложность и **связность текста**, все повороты его тематики. «Неучтёнными» остаются лишь детали, подробности, числовые данные и т.п. Однако уточнение деталей принципиально выходит за рамки выявления смысла и требует другой технологии.

Фундаментальная ценность механизма выявления онтологического смысла кроется в том, что он, создавая графический портрет текста и описывая его метаязыком онтологии, позволяет человеку сложить хотя и самое общее и схематичное, но вполне адекватное представление о ситуации, дать ему концентрированную информацию, возбудить целенаправленные вопросы, отсеять лишние гипотезы.

**В итоге потребитель получает определённую ясность – в чём состоит суть сообщения.**

Что касается человеческого понимания онтологического смысла, формат которого непривычен для (современного) человека, то для его преобразования в грамматическую языковую форму предусмотрен, как указано выше, специальный транслятор «смысл – текст». Принципы трансляции разработаны в лингвистической модели «смысл–текст» [Мельчук, 1999].

---

## Раскрытие смысла в диалоге

---

«Текст обогащает смысл» (эта формула принадлежит С. Васильеву, 1988 г).

Как было показано выше, активная функция текста проявляется в том, что он является, в принципе, неисчерпаемым источником интереса, причём при повторном чтении возможно не только переосмысливание дискурса, но даже изменение онтологии (*Sic!*). Особенно это относится к учебным материалам и высокохудожественным произведениям. Поэтому необходимо обеспечить многократный доступ пользователя к первоисточнику для более полного раскрытия его ресурса. Это возможно в режиме диалога.

В теории репрезентации знаний используют понятие «*страты знаний*» [Гаврилова и Хорошевский, 2001]. Стратификация знаний производится по типу их анализа, при этом различают: *Зачем* – знания, *Кто* – знания, *Что* – знания, *Как* – знания, *Где* – знания, *Почему* – знания и т.д.

Если в упрощённом виде использовать эту методологию для организации диалога, то следует модернизировать лингвистический процессор. В процессе построения дерева синтаксического разбора предложения связи (стрелки) между членами предложения должны быть проиндексированы вопросами, например, такими: *кто, что-делает, где, когда, какой, сколько, зачем, каким-образом* и т.д. Такая индексация создаёт для диалога лингвистическую базу. Вопросы к тексту, сформулированные пользователем, активизируют соответствующие группы слов, которые формируют ответ и тем самым раскрывают «глубинные падежи» ситуации [Поспелов, 1988; Мальковский, 1985; Штерн, стр. 71].

---

## Заключение

---

Предлагаемый взгляд на общие ресурсы текста ЕЯ, на задачи его семантического анализа, способы и результат обработки приводит к единой триаде: «*Онтология – Текст – Смысл*». Задачей семантического анализа текста полагается извлечение концентрированного знания, релевантного замыслу автора. Платформой извлечения знания служит онтология **ИО\*3** – концептуальная система репрезентации общих знаний о мире и предметных областях. Результатом взаимодействия текста с онтологией является онтологический смысл – множество взаимосвязанных подграфов онтологического графа.

Онтологический смысл извлекается из онтологического графа «смысловым процессором» и интерпретируется с помощью транслятора «смысл – текст».

Для более глубокого изучения содержания документа используется «диалоговый процессор», который исследует дерево синтаксического разбора предложения и по заданному вопросу находит в тексте фрагмент, служащий конкретным ответом на вопрос пользователя.

Компьютерное понимание текста достигается за счёт: 1) погружения текста в единую среду знаний – онтологию, 2) формального представления смысла в памяти компьютера и 3) возможности операций над онтологическим смыслом.

---

## Возможные применения

---

Предлагаемую новую информационную технологию можно использовать для формирования Баз данных, архивирования электронных документов, их индексирования, классификации и поиска в Интернет. В виртуальных лабораториях возможно на её основе создавать интеллектуальные банки данных, работающие в единой среде знаний.

Данная технология ориентирована на автоматическое извлечение **метаданных** из текстовых документов. При соответствующей доработке она может служить в системах автоматического реферирования научных публикаций, а в перспективе – для осмысленного интерпретирования мультимедийных документов.

---

## Благодарности

---

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария [www.ithea.org](http://www.ithea.org) и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина [www.aduis.com.ua](http://www.aduis.com.ua).

---

## Литература

---

- [Sowa, 2002] J.F. Sowa. Concepts in Lexicon: Introduction. Architectures for Intelligent Systems. – IBM Systems Journal, vol. 41, No. 3, 2002. – pp. 331 – 349.
- [Zaboleeva–Zotova & Orlova, 2009] A. Zaboleeva–Zotova, Y. Orlova. Computer Support of Semantic Text Analysis of a Technical Specification on Designing Software. – International Book Series, Number 9. Intelligent Processing. Supplement to the International Journal “Information Technologies & Knowledge” Volume 3 / 2009. – ITHEA, Sofia, 2009. – p. 29.
- [Балл, 2008] Психология в рациогуманистической перспективе. Избранные работы. – К.: «Основа», 2006. – 401 с.
- [Гаврилова и Хорошевский., 2001] Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – СПб.: Питер, 2001. – 384 с.
- [Гладун и др., 2008] В. Гладун, В. Величко, Л. Святогор. Структурирование онтологии ассоциаций для конспектирования естественно–языковых текстов – International Book Series, Number 2. Advanced Research in Artificial Intelligence. Supplement to the International Journal “Information Technologies & Knowledge” Volume 2 / 2008. – ITHEA, Sofia, 2008. – p. 153.
- [Демьянков, 1985] Демьянков В.З. Основы теории интерпретации и её приложения в вычислительной лингвистике. – М.: Изд-во Моск. ун-та, 1985. – 76 с.
- [Мальковский, 1985] М.Г. Мальковский. Диалог с системой искусственного интеллекта. – М.: Изд-во Моск. ун-та, 1985. – 213 с.
- [Мельчук, 1999] Мельчук И.А. Опыт теории лингвистических моделей «Смысл – Текст». – М.: Школа «Языки русской культуры», 1999. – 346 с.
- [Ной и МакГиннесс] Natalya F. Noy, Deborah L. McGuinness:  
[http://protege.stanford.edu/publications/ontology\\_development/ontology101.html](http://protege.stanford.edu/publications/ontology_development/ontology101.html)
- [Палагин и др., 2009] А. Палагин, С. Крывый, В. Величко, Н. Петренко. К анализу естественно-языковых объектов – International Book Series, Number 9. Intelligent Processing. Supplement to the International Journal “Information Technologies & Knowledge” Volume 3 / 2009. – ITHEA, Sofia, 2009. – p. 36.
- [Поспелов, 1988] Г.С. Поспелов. Искусственный интеллект – основа новой информационной технологии. – М.: Наука, 1988. – 279 с.
- [Святогор и Гладун., 2009] Л. Святогор, В. Гладун. Определение понятия «Смысл» через онтологию. Семантический анализ текстов естественного языка. – International Book Series, Number 9. Intelligent Processing. Supplement to the International Journal “Information Technologies & Knowledge” Volume 3 / 2009. – ITHEA, Sofia, 2009. – p.53.
- [Шенк, 1980] Шенк Р. Обработка концептуальной информации. Пер. с англ. – М.: Энергия, 1980.
- [Штерн, 1998] І.Б. Штерн. Вибрані топіки та лексікон сучасної лінгвістики. Енциклопедичний словник. – К.: «АртЕк», 1998. – 335 с.

---

## Информация об авторах

---

**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)

**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)