

# ОСМЫСЛЕННАЯ ОБРАБОТКА ИНФОРМАЦИИ. СЕМАНТИЧЕСКИЕ И КАТЕГОРИАЛЬНЫЕ АСПЕКТЫ

Александров Сергей Эдуардович  
Фадеев Павел Эдвардович  
Компания “Интеллектуальный Партнер”  
[Http://mujweb.cz/www/context](http://mujweb.cz/www/context)  
E-mail: [intelpart@tut.by](mailto:intelpart@tut.by)

## Аннотация

Предлагается подход к решению проблемы осмысленной обработки информации, основанный на применении категории “элементарный смысл”.

Элементарный смысл понятия представляется в виде устойчивой системы отношений самого понятия с другими понятиями. Для фиксации смысла понятия применен модуль элементарного смысла — универсальная, присущая любому понятию, структура. Предложены принципы и алгоритмы взаимодействия элементарных смыслов между собой. Демонстрируются новые возможности систем осмысленного поиска информации, машинного перевода с учетом контекстов, систем поддержки принятия решений.

На примере коммерческих программных продуктов показано, как, с использованием предложенных механизмов, под конкретную предметную область организуется картина мира элементарных смыслов, и строятся принципы управления такой картиной.

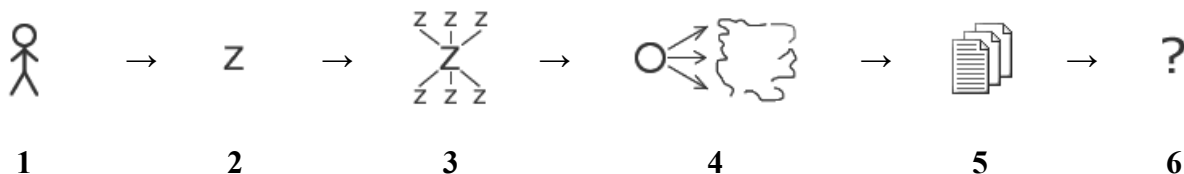
## Информация и ее смысл

По определению Н.Винера, “информация — это обозначение содержания, полученного из внешнего мира в процессе нашего приспособления к нему наших чувств” [5]. Из определения следует, что собственно информация (символы, знаки, обозначения, имена, названия и т.д.) принципиально отличается от ее содержания, смысла.

Указанное различие ярко проявляется во всех гранях взаимодействия человека с компьютером. Например, в процессе автоматического поиска информации — легко искать обозначения, имена, названия, ключевые слова и т.п., однако найти в информационном пространстве то, что требуется по смыслу — уже искусство или удача. Еще одна актуальная область — автоматический перевод с одного языка на другой. Слабый учет смысла и контекста переводимой информации приводит к тому, что любой перевод всегда требует редактуры человеком-специалистом. Таких примеров можно приводить много. Вывод — систематизация, обработка знаний вынужденно сводится к систематизации и обработке обозначений, но этим и ограничивается по определению.

Рассмотрим эту проблему на примере поисковой машины.

Для этого рассмотрим структуру типового процесса поиска информации (здесь под информацией понимается обозначение смысла в виде знаков, символов, ключевых слов):



Где:

1. 1. Пользователь со своими потребностями в осмысленной информации.
2. 2. Запрос пользователя к поисковой машине.
3. 3. Трансформация поисковой машиной запроса пользователя и построение информационного образца для поиска.
4. 4. Собственно поисковой механизм, который идентифицирует образец в информационном пространстве. Информационное пространство, в котором осуществляется поиск.
5. 5. Найденная для пользователя информация.
6. 6. Оценка пользователем предоставленной ему информации и с точки зрения того, содержится ли в ней требуемый ему смысл.

Существуют два полярных принципа работы поисковой машины.

Первый — поиск информации только по ключевым словам. Этот подход используется давно и заключается в установлении соответствия между искомыми символами и символами информационного источника.

Второй — поиск по различным аналогиям. Поиск по аналогиям заключается в том, что в общем случае используются: прямые аналогии; ассоциации; системные отношения; временные связи; подобию по свойствам, по особенностям, по функциям и т.д. Этот ряд можно уточнять до бесконечности.

Исходя из вышесказанного, можно сформулировать следующее противоречие поиска:

- С одной стороны, если поиск информации базируется на ключевых словах (обозначениях), то поисковая машина не находит информационные источники, связанные с запросом по смыслу.
- С другой стороны, если поиск информации базируется на аналогиях, то это приводит к неограниченному увеличению найденных поисковой машиной информационных источников.

Поиск по ключевым словам по определению оперирует только информационной составляющей и, в этом смысле, обладает принципиальными ограничениями. Поиск по аналогиям потенциально содержит две составляющие — информационную и смысловую, поэтому более перспективен. Разрешение указанного выше противоречия сводится к решению проблемы множественности аналогий.

Проблема множественности аналогий удобно разрешается путем использования *аналогий, присущих естественному языку — субъектно-ориентированных аналогий*.

Так, например, толковый словарь естественного языка описывает аналогии, которые с наибольшей вероятностью подразумевает человек. Для любого понятия, обозначенного словом, толковый словарь дает достаточно полный, присущий этому понятию набор аналогий.

Таким образом, субъектно-ориентированные аналогии — это:

- 1) аналогии, которые с наибольшей вероятностью подразумевает человек;
- 2) аналогии, которые непосредственно присутствуют в естественном языке, поскольку естественный язык является основным средством коммуникации между людьми;
- 3) аналогии, которые “...получены из внешнего мира в процессе нашего приспособления к нему наших чувств” [5].

Для того чтобы при поиске смыслового содержания информации можно было использовать субъектно-ориентированные аналогии, необходимо иметь:

- Аналогии, присущие понятиям естественного языка.
- Принципы связи и взаимодействия различных аналогий между собой.
- Структуру представления понятий языка в виде системы аналогий.

### **Модуль элементарного смысла**

Поскольку отдельное слово естественного языка можно рассматривать как элементарную информационную составляющую языка (элементарное обозначение в языке), то обозначаемое отдельным словом (содержание, стоящее за словом) можно рассматривать как элементарный смысл в естественном языке.

При этом *под смыслом слова (символа, понятия, имени), понимается определенная устойчивая система отношений данного слова с другими словами* [2, 11].

Очень важно, что количество этих отношений конечно, а значит, поддается конкретному учету. Для их фиксации требуется специальная структура — модуль элементарного смысла [2].

Рассмотрим, как должен выглядеть модуль элементарного смысла. Для этого перечислим требования, которым он должен удовлетворять.

Во-первых, модуль элементарного смысла должен опираться на синтаксис естественного языка, для того чтобы иметь возможность из понятий строить словосочетания и предложения.

Это требование указывает на то, что в основе построения модуля элементарного смысла должен лежать естественный язык. Исследования лингвистов показали, что “есть “нечто такое”, что необходимо присуще всякой системе, если она является знаковой, т.е. выполняет коммуникативную и номинативную функции, ... и таким “нечто” являются абсолютные универсалии в сфере семантики и синтаксиса” [3]. При этом, несмотря на многообразие грамматических категорий, существует конечное число, так называемых внутренних семантических падежей, отражающих глубинную связь между словами речи [6, 7, 8]. Приведем некоторые из них [2]:

- субъект,
- действие,
- объект,
- адресат,
- инструмент,

- тема,
- место,
- принадлежность,
- признак действия,
- признак предмета,
- причина,
- цель,
- время,
- условие,
- количество,
- модальность и др.

Таким образом, для обеспечения связи модуля элементарного смысла с естественным языком, необходимо, организовать его так, чтобы он оперировал этими семантическими падежами.

Во-вторых, модуль элементарного смысла должен быть универсальным, для того чтобы отражать все многообразие смысла понятий естественного языка.

В любом естественном языке для выполнения им коммуникативных функций и передачи смысла понятий присутствуют элементы, с помощью которых можно описать: а) взаимодействия, б) структуру предметов или событий, в) время. Иными словами, элементарный смысл понятия достаточно точно описывается, если а) указать взаимодействия рассматриваемого понятия с другими понятиями, б) описать структуру понятия (из чего состоит? во что входит?) и в) указать причинно-следственную временную цепочку, в которую входит понятие (что за чем следует?) [2].

В-третьих, модуль элементарного смысла должен быть конкретным, т.е. иметь некую ограниченную структуру, для того чтобы ее можно было формализовать и оперировать ею.

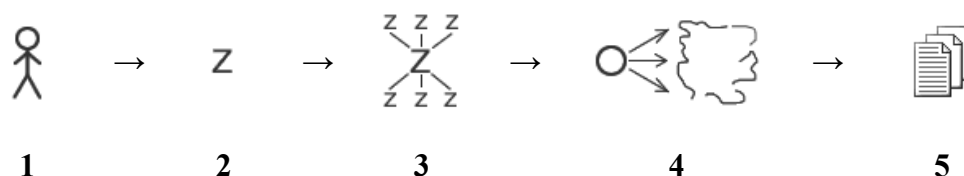
Практика разработки программного обеспечения для обработки естественного языка показала, что существуют два типа модулей элементарного смысла: объект и действие. Такой подход позволяет изначально разделить все понятия на две основные категории: действия и объекты [2, 7].

Для анализа предложений естественного языка и построения модулей элементарных смыслов понятий, а также для автоматического создания словарей элементарных смыслов используется грамматический код, разработанный профессором В.А.Карповым [6].

### **Оперирование модулем элементарного смысла.**

Взаимодействие модулей элементарного смысла между собой в соответствии с взаимодействием слов в языке позволяет представлять тексты в двухуровневом виде. Первый уровень — собственно текст, второй уровень — стоящая за текстом система отношений его элементарных смыслов.

Такое представление информации и ее смысла позволяет организовать процесс работы поисковой машины следующим образом:



Где:

7. 1. Пользователь со своими потребностями в осмысленной информации.
8. 2. Запрос пользователя к поисковой машине.
9. 3. Трансформация поисковой машиной запроса пользователя в систему отношений элементарных смыслов запроса, построение на этой основе поисковой матрицы смыслов.
10. 4. Поисковой механизм, который идентифицирует поисковую матрицу смыслов в информационной среде.
11. 5. Найденная для пользователя информация, соответствующая запросу пользователя по смыслу.

В данной работе [2] элементарная обработка смысла информации проиллюстрирована на примере работы поисковой машины. Однако предложенный подход открывает интересные перспективы для создания коммерческих интеллектуальных устройств самого разнообразного применения. Использование механизмов смысловой обработки информации актуально в следующих областях:

- восстановление по смыслу пропущенной в тексте информации,
- машинный перевод с учетом контекста,
- распознавание и синтез речи с учетом контекстно содержащейся в ней информации,
- осмысленный анализ информации с точки зрения требуемых контекстов,
- управление техническими и информационными системами с учетом человеческого фактора и др.

## Литература

12. 1. Александров С., Фадеев П., Рейтинг-анализ. Учебное пособие; Мн: Тэхналёгія, 1997.
13. 2. Аляксандраў С., Фадзееў П., Шматкантэкставая аўтаматычная апрацоўка вялікіх аб'ёмаў інфармацыі. Прынцыпы, мэтады, зыходныя альгарытмы; Прага: МК Tisk, 2000.
14. 3. Будущее искусственного интеллекта; М: Наука, 1991.
15. 4. Винер Н., Кибернетика, или управление и связь в животном и машине; М: Наука, 1983.
16. 5. Винер Н., Человек управляющий; СПб: Питер, 2001.
17. 6. Карпов В.А. Язык как система, Мн: Вышэйшая школа, 1992.
18. 7. Мартынов В., Универсальный семантический код УСК-4; Мн: Наука и Техника, 1988.
19. 8. Мельчук И.А., Опыт теории лингвистических моделей "смысл-текст"; М: Наука, 1974.
20. 9. Осуга С., Обработка знаний; М: Мир, 1989.
21. 10. Представление и использование знаний, Под ред. Х.Уэно, М.Исидзука; М: Мир, 1989.
22. 11. Приобретение знаний, Под ред. С.Осуги, Ю.Саэки; М: Мир, 1990.
23. 12. Пуанкаре А., О науке, Под ред. Л.С.Понтрягина; М: Наука, 1983.