

УДК 519.62

Д.В. Преснякова

Харьковский национальный университет радиотехники, Харьков

ЭКВИВАЛЕНТНОСТЬ СЕМАНТИЧЕСКИХ ФУНКЦИЙ

В статье предлагается метод вычисления смысловой эквивалентности двух естественно-языковых форм, представленных в виде семантических функций. Смысл языковых форм ограничивается уровнем лексической семантики, т.е. базой знаний, представленной в виде толкового словаря, где каждое значение слова описывается семантической функцией. В основе метода лежит представление семантической функции ориентированным деревом, где каждая вершина помечена семантической функцией. Метод состоит в обходе деревьев и установления для каждой пары вершин эквивалентности в плане лексической семантики. Операция вычисления эквивалентности может быть использована при автоматическом анализе естественно-языковых текстов, при контекстном поиске информации.

Ключевые слова: *естественный язык, толковый словарь, семантическая функция, семантический словарь, эквивалентность функций, формализация естественного языка.*

Введение

Вычислительная техника предоставляет широкие возможности для хранения и обработки информации, поэтому накопление информации происходит в электронном виде. Поскольку рост информационных ресурсов имеет экспоненциальный характер, то для их обработки создаются системы, ориентированные на анализ и поиск информации. Накопление огромных массивов информации в текстовом виде делает актуальным создание систем, ориентированных на обработку текстовой информации. Диапазон таких систем весьма широк: системы автоматического перевода, естественно-языковые интерфейсы, системы автоматического реферирования. Несмотря на широту диапазона решаемых задач, качество работы всех этих систем прямо зависит от уровня формализации естественного языка, который может предложить современная теория искусственного интеллекта.

Проблема интеллектуального анализа и понимания текстов на естественном языке (ЕЯ) появилась одновременно с созданием компьютеров. Однако, до настоящего времени достичь полного успеха в её решении не удалось. Методы для решения частных задач разработаны в большом количестве, но количество так и не перешло в качество. Полное понимание языка на вычислительной (формальной) базе на данный момент остаётся далеко за пределами современных возможностей и многие фундаментальные проблемы в области автоматической обработки ЕЯ ещё не решены, но развитие современных информационных технологий требует решения частных задач.

Моделирование интеллектуальной деятельности человека по обработке текстовой информации является сверхсложной задачей. Её успешная автоматизация привела бы к значительному повышению

эффективности самих компьютеров, давая возможность человеку общаться с компьютером на естественном языке.

Моделирование естественного языка осуществляется на разных уровнях. Наиболее сложными для моделирования являются уровни, на которых ведётся работа со значениями отдельных элементов и текста в целом. Тесная связь структуры текста и его значения не даёт возможности построить даже простые модели обработки текстовой информации без учета значений элементов, которые составляют текст.

Большое количество современных работ в области искусственного интеллекта направлены на разработку моделей семантики, которые позволяют сделать качественный скачок в семантической интерпретации текстов и улучшить результаты практической работы систем обработки текстовой информации.

Развитие средств формального представления семантики текстов естественного языка и их дальнейшее внедрение в системах обработки текстовой информации способствует развитию систем искусственного интеллекта, усовершенствованию систем анализа текстов с интеллектуальной составляющей.

Постановка задачи. Смысл ЕЯ-форм выявляется в процессе анализа текста на основе определённых грамматических и лексико-семантических форм текста, т.е. на основе его семантических и синтаксических характеристик. Одни из этих характеристик выражены в тексте явно (эксплицитно), другие неявно (имплицитно). Задачей анализа является выявление грамматической и семантической структуры текста. Анализ грамматической структуры текста, направленный на обнаружение его семантической структуры можно рассматривать как разбор многозначных элементов синтак-

сической структуры и выявление однозначных семантических составляющих текста. При этом сложные семантические и синтаксические конструкции получаются из более простых (составляющих), что требует определения семантических единиц различного уровня и их соотнесение с грамматическими единицами. Предлагается ввести два уровня семантики: лексическую и формальную. Лексическая отражает значение слов и словосочетаний и базируется на толковом словаре и синтаксисе естественно-языковых форм. Формальная семантика отражает объекты и логические связи между ними. Построена модель лексической семантики в виде семантических функций [1] и предложена формализация толкового словаря [2] для интерпретации лексической семантики. Моделью формальной семантики служит модифицированная версия исчисления предикатов. Семантические функции и формулы логики предикатов получаются в результате анализа естественно-языковых и промежуточных форм [3]. Одной из основных операций применяемых в процессе анализа естественного языка, является операция установления эквивалентности двух языковых форм с точки зрения семантики.

Целью данной работы является формулировка метода установления семантической эквивалентности двух семантических функций.

Семантические функции и лексическая эквивалентность

Семантические функции являются конструкциями формального языка, представляющего собой термы. Вопрос эквивалентности двух термов t_1 и t_2 можно рассматривать как вопрос $t_1 = t_2$, т.е. равенство этих термов с точки зрения их лексической семантики.

Структурно семантическая функция может быть представлена в виде ориентированного дерева. Например, семантическая функция

$$F_1(F_2(X_1, X_2), F_3(Y_1, Y_2, Y_3), F_4(Z))$$

может быть представлена деревом на рис. 1.

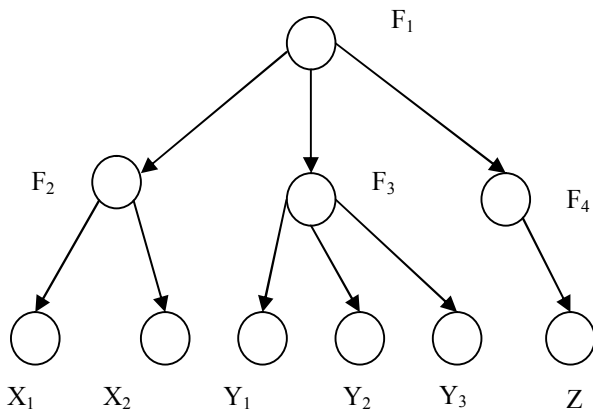


Рис. 1. Дерево синтаксической функции

Вершины дерева на рис. 1 помечены семантическими функциями типа F, по которым можно вычислить их лексические значения на основе семантического словаря.

Можно считать, что с вершинами дерева связаны лексические значения.

Рассмотрим конкретный пример установления эквивалентности семантических функций. Рассмотрим лексическое значение слова «абажур» – V_1 (абажур), которое трактуется как «колпак для лампы» [4]. В свою очередь «колпак» трактуется как «покрышка», а лампа как «осветительный прибор».

Семантические функции для различных лексических значений слова «абажур» могут иметь вид:

$$V_1(\text{абажур}); \tag{1}$$

$$F_{\text{предлог}}(V_1(\text{для}), V_1(\text{колпак}), V_1(\text{лампа})); \tag{2}$$

$$F_{\text{прилагательное}}(V_1(\text{прибор}), V_1(\text{осветительный})). \tag{3}$$

Покажем эквивалентность этих трех семантических функций.

Дерева функций (1), (2), (3) имеет вид (соответственно перечислению) (рис. 2 – 4):



Рис. 2. Дерево_1

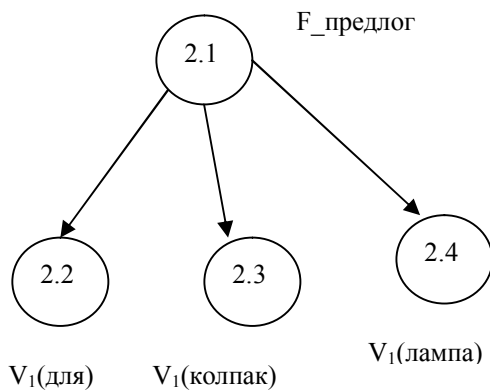


Рис. 3. Дерево_2

Сравнивая Дерево_1 и Дерево_2, расписываем вершину 1.1 в соответствии со значением семантического словаря и получаем функцию (2.), а соответственно и ее дерево как Дерево_2. сравнение полученного дерева и Дерева_2 с рис. 3, устанавливаем их полную эквивалентность.

Сравним теперь Дерево_2 и Дерево_3 путем последовательного просмотра и сравнения их вершин. Две вершины из разных деревьев считаются эквивалентными, если они имеют одинаковые лексические значения.

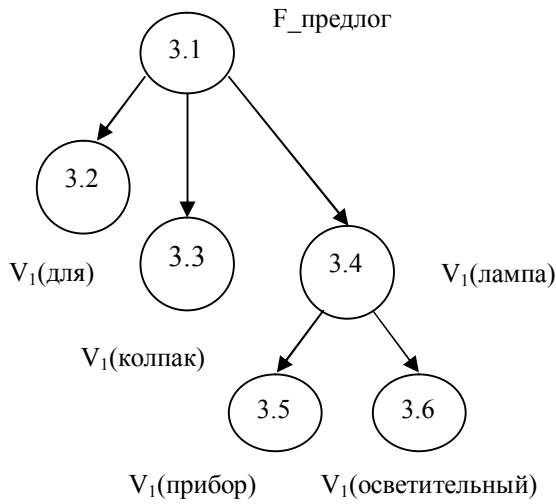


Рис. 4. Дерево_3

Вершины 2.1 и 3.1 эквивалентны, так как помечены одной и той же функциональной константой и имеют одинаковое количество детей. Вершины 2.2 и 3.2 эквивалентны, исходя из одинаковости их пометок. Вершины 2.3 и 3.3 эквивалентны в силу того, что лексическое значение слов «колпак» и «покрышка» эквивалентны по семантическому словарю. Вершины 2.4 и 3.4 эксплицитно не эквивалентны. Но лексическое значение слова «лампа» содержит значение «осветительный прибор», что приводит к трансформации вершины 2.4 в дерево с F_прилагательное в корне. Сравнивая это дерево с поддеревом {3.4, 3.5, 3.6} заключаем их эквивалентность. Следовательно, Дерево_2 эквивалентно Дерево_3.

Информации, содержащейся в семантическом словаре, не всегда бывает достаточно для установления эквивалентности семантических функций. Нужна дополнительная информация о возможной эквивалентности функциональных констант. Рассмотрим конкретный пример.

Пусть имеются семантические функции:

$$F_{\text{предложение}}(X_1, X_2, X_3), \quad (4)$$

где X_1 – глагол, обозначающий действие; X_2 – объект, осуществляющий действие; X_3 – объект, на который направлено действие;

$$F_{\text{генитив}}(Y_1, Y_2), \quad (5)$$

где Y_1 – существительное, обозначающее событие; Y_2 – существительное в родительном падеже или семантическая функция F, в которой существительное Y_2 в родительном падеже является главным словом.

Функциональная константа F_предложение из (4) и F_генитив из (5) могут быть эквивалентными, если существительное Y_1 выражает событие как результат действия X_1 .

Для примера сравним смысл двух словосочетаний «программа стирает память» и «память стирается программой». Если не учитывать некоторые нюансы (к примеру, время действия), то смысл этих предложений одинаков, и их семантические функции должны быть

эквивалентными. Семантические функции для вышеприведенных предложений будут соответственно:

$$F_{\text{предложение}}(V(\text{стирать}), V(\text{программа}), V(\text{память})); \quad (6)$$

$$F_{\text{генитив}}(V(\text{стирание}), F_{\text{творительный_падеж}}(V(\text{память}), V(\text{программа}))). \quad (7)$$

Дерево для формулы (6) представлено на рис. 5, а для формулы (7) – на рис. 6.

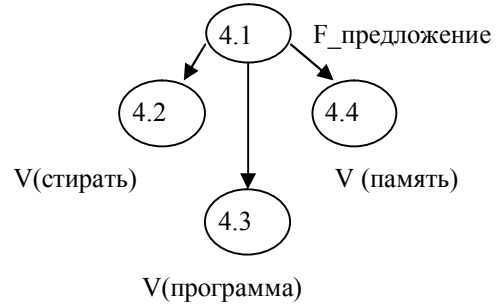


Рис. 5. Дерево_4

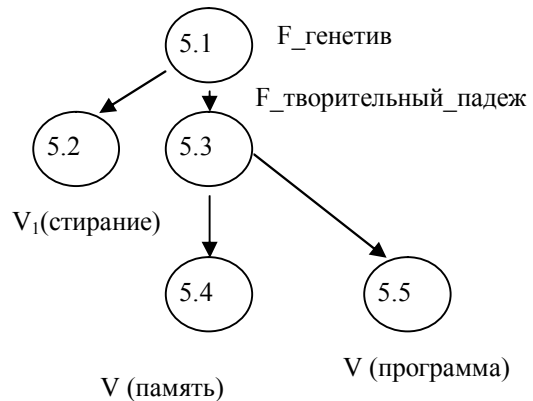


Рис. 6. Дерево_5

Доказать эквивалентность Дерево_4 и Дерево_5, используя вышеуказанную методику в чистом виде, т.е. сравнивая текущие вершины двух деревьев, затруднительно. В этом случае, когда сравниваются F_предложение – действие и F_генитив с F_творительный_падеж в качестве второго аргумента и событием в качестве первого аргумента, F_генитив предварительно должно быть преобразовано в глагольную форму, т.е. в F_предложение. Дерево_5 должно так же быть заменено на новое дерево, соответствующее семантической функции, полученной из F_генитив. Такое преобразование переводит формулу (7) в (6), дерево которой совпадает с Дерево_4.

Следовательно, чтобы установить эквивалентности, выводимые из формул типа (6), (7) необходимо сформулировать методы эквивалентных преобразований семантических функций. Иными словами, нужно иметь методы вычисления значений семантических функций, и эти значения представлены другой семантической функцией типа F.

Метод вычисления эквивалентности семантических функций

Сформулируем теперь метод вычисления эквивалентности двух семантических функций, для чего введем следующие обозначения. СФ1 и СФ2 – семантические функции; Д1 и Д2 – деревья для функций СФ1 и СФ2 соответственно; ТекД1 и ТекД2 – пара обрабатываемых вершин (текущих) из соответствующих деревьев; СемСл – семантический словарь.

Метод вычисления семантических функций.

Вход: СФ1, СФ2.

Выход: «да», «нет».

1. Преобразование СФ1 в Д1, СФ2 в Д2.
2. Обход Д1 и Д2 в глубину с фиксацией пары вершин в ТекД1 и ТекД2.
3. Если ТекД1 = ТекД2 (как термы), то выполнить пункт 2.
4. Если ТекД1 = 'V(α)' и ТекД2=F_1, то преобразование Д1: ТекД1 становится корнем поддерева, который обозначен F_2, где F_2 – семантическая функция, являющаяся значением V_i(α) в СемСл. Указатели на вершины ТекД1 и ТекД2 не изменяются. Выполнить пункт 3.
5. Если ТекД1 = F и ТекД2 = 'V(α)', аналогично пункту 4.
6. Если ТекД1 = 'V_i(α)' и ТекД2 = 'V_j(β)', то преобразовать Д1 и Д2. ТекД1 становится корнем для СФ F_1, ТекД2 – корнем поддерева для СФ F_2, где F_1 и F_2 суть значения V_i(α) и V_j(β) соответственно в СемСл. Указатели в ТекД1 и ТекД2 не меняются. Выполнить пункт 3.
7. Если ТекД1 = 'F_1' и ТекД2 = 'F_2', то выбрать дерево для преобразования и изменить его в соответствии с правилами вычисления значения для функции типа F.

Дадим оценку сложности алгоритма, построенного по вышеописанному методу. Строка 1 метода, где строится дерево по семантической функции, оценивается как O(n), где n – количество вершин дерева. Строка 2 определяет цикл обхода деревьев с

п вершинами и оценивается как O(n). В теле этого цикла работает цикл по преобразованию деревьев. Преобразование состоит в построение поддеревьев с m вершинами, что оценивается как O(m). Общая сложность алгоритма имеет оценку O(m*n).

Выводы

В работе предложен метод установления смысловой эквивалентности двух семантических функций, которые отражают лексическую семантику естественно-языковых форм. Операция вычисления эквивалентности широко используется в процессе анализа естественно-языковых форм для получения их лексических значений в виде семантических функций. Эта операция может быть использована в процессе логического вывода в модели формальной семантики. Операция вычисления эквивалентности может использоваться и в исключительно прикладном аспекте. Например, при поиске информации с использованием контекста.

Список литературы

1. Дюбо Г.Ф. Модель поверхностного смысла естественно-языкового языка на базе семантических функций / Г.Ф. Дюбо, Д.В. Преснякова // Бионика интеллекта: научн.-техн.журнал. – 2007. – № 1(66). – С. 103-106.
2. Дюбо Г.Ф. Формализация семантики естественно-языкового языка с применением лямбда-исчисления / Г.Ф. Дюбо, Д.В. Преснякова // Бионика интеллекта: научн.-техн.журнал. – 2007. – № 2(67). – С. 41-46.
3. Дюбо Г.Ф. Формальная семантика и анализ естественно-языкового языка / Г.Ф. Дюбо, Д.В. Преснякова // Восточно-Европейский журнал передовых технологий. – Х., 2008. – № 4/2 (34). – С. 48-53.
4. Ожегов С.И. Словарь русского языка: Ок.57 тыс. слов / С.И. Ожегов; под ред.чл.-корр.АН СССР Н.Ю. Шведовой. – 17-е изд., стереотип. – М.: Рус. яз., 1985. – 797 с.

Поступила в редколлегию 25.03.2009

Рецензент: д-р техн. наук, проф. С.Ю. Шабанов-Кушнарченко, Харьковский национальный университет радиотехники, Харьков.

ЕКВІВАЛЕНТНІСТЬ СЕМАНТИЧНИХ ФУНКЦІЙ

Д.В. Преснякова

У статті пропонується метод обчислення смислової еквівалентності двох природно-мовних форм, представлених у вигляді семантичних функцій. Сенси мовних форм обмежуються рівнем лексичної семантики, тобто базою знань, представленою у вигляді тлумачного словника, де кожне значення слова описується семантичною функцією. У основі методу лежить представлення семантичної функції орієнтованим деревом, де кожна вершина помічена семантичною функцією. Метод полягає в обході дерев і встановлення для кожної пари вершин еквівалентності в плані лексичної семантики. Операція обчислення еквівалентності може бути використана при автоматичному аналізі природно-мовних текстів, при контекстному пошуку інформації.

Ключові слова: природна мова, тлумачний словник, семантична функція, семантичний словник, еквівалентність функцій, формалізація природної мови.

EQUIVALENCE OF SEMANTIC FUNCTIONS

D.V. Presnyakova

The method of calculation of semantic equivalence of two naturally-linguistic forms, presented as semantic functions is offered in the article. Sense of linguistic forms is limited to the level of lexical semantics, i.e. by the base of knowledges, presented as an explanatory dictionary, where every value of word is described a semantic function. In basis of method presentation of semantic function lies the oriented tree, where every top is noted a semantic function. A method consists of tree traversal and establishment for every pair of tops of equivalence in the plan of lexical semantics. The operation of calculation of equivalence can be used for the automatic analysis of naturally-linguistic texts, at a context information retrieval.

Keywords: human language, explanatory dictionary, semantic function, semantic dictionary, equivalence of functions, formalization of human language.