

УДК 519.854

С.Ю. Шабанов-Кушнарченко¹, О.В. Калиниченко¹, А.И. Коваленко², А.А. Шматько¹¹ Харківський національний університет радіоелектроніки, Харків² Харківська державна академія культури, Харків

О ФОРМАЛИЗАЦИИ ЗНАНИЙ НА БАЗЕ АЛГЕБРЫ КОНЕЧНЫХ ПРЕДИКАТОВ

Рассмотрена проблема обоснования выбора подхода для формализации смысла текстов на естественном языке. Проанализированы основы связи знаний и смысла текстов на естественном языке; подходы к получению, представлению и формализации знаний. Рассмотрены основы применения аппарата алгебры конечных предикатов и реляционных сетей для формализации смысла текстов естественного языка.

Ключевые слова: смысл текста, предикатная модель, знания, формализация.

Введение

Изучение и формализация интеллекта приводят к задаче алгебраизации логики [1], к понятию универсальной логической алгебры. Эта алгебра представляется в виде двухэтажной конструкции: на первом этаже – мысли, на втором – действия над ними. Носитель универсальной логической алгебры представляет собой множество всех мыслей. В нем выделены особо некоторые мысли, играющие роль базисных элементов. Над множеством всех мыслей заданы базисные операции. Они должны быть выбраны с таким расчетом, чтобы с помощью суперпозиции базисных операций, примененных к базисным элементам, можно было бы выразить любые мысли. В этом случае логическая алгебра будет универсальной. Записи таких суперпозиций называется формулами универсальной логической алгебры. Разные формулы, выражающие одну и ту же мысль, называются тождественными.

Легко представить универсальную логическую алгебру во вполне конкретном виде. Язык, которым мы пользуемся, называемый естественным, является воплощением такой алгебры. Естественный язык (ЕЯ) – это разновидность универсальной логической алгебры, ее конечный фрагмент. Было бы естественно назвать его лингвистической алгеброй [2]. Формулами лингвистической алгебры служат предложения и тексты ЕЯ. Они выражают мысли человека. Операциями лингвистической алгебры служат отдельные мыслительные акты, в результате которых рождаются новые мысли. Процесс мышления представляет собой суперпозицию операций лингвистической алгебры. Предложения, выражающие одну и ту же мысль, – это тождественные формулы лингвистической алгебры. Разные языки, например, русский и английский, представляют собой различные виды лингвистической алгебры, они заданы на одном и том же носителе (поскольку мысли интернациональ-

ны), но отличаются базисными элементами и операциями. Перевод текстов с одного языка на другой означает переход от формул одного варианта лингвистической алгебры к другому при сохранении смысла этих текстов.

Целью работы является анализ возможностей и перспектив применения математического аппарата алгебры конечных предикатов (АКП) для формализации структур ЕЯ и знаний в виде текстов ЕЯ.

Изложение основного материала

Формализация знаний на базе АКП

Приведем основные общепризнанные свойства понятия «знание». Знание – это проверенные общественной практикой полезные сведения, которые могут многократно использоваться людьми при решении тех или иных задач. Информация – это формализованное знание. Явные знания об одном и том же объекте предметной области могут отличаться. Отличия между этими знаниями вызваны неявной неформализованной составляющей.

Знание, понимаемое как «продукт познания людьми предметов и явлений действительности, законов природы и общества», является формой перехода от явления к отображению существенных свойств явления в тексте. Смысл, в свою очередь, преобразует представление о существенном, в общем и целом, в представление о существенном с позиции конкретного индивида.

Важнейшим инструментом, способствующим отображению познаваемых закономерностей, является язык.

Поскольку ЕЯ не формализован, для машинного представления знаний используются модели представления знаний (МПЗ). Известно семь классов моделей знаний: логические, продукционные, фреймовые, сетевые, объектно-ориентированные, специальные и комплексные. Один из подходов в

последнем классе к построению таких моделей предполагает выявление лингвистических отношений на всех уровнях и их представление в виде уравнений АКП.

Решение задач анализа и синтеза текстов ЕЯ сводятся к построению моделей в виде систем уравнений АКП и подстановке известных значений в эти уравнения. Такие модели легко могут быть переведены в тексты ЕЯ, есть возможность эквивалентных преобразований модели.

Для задач автоматической обработки текстов ЕЯ существенным является возможность использования аппарата реляционных сетей (РС). Реляционная сеть – это результат бинаризации системы уравнений АКП произвольной размерности. Основная цель создания РС – переход к модели, которую можно реализовать аппаратно, в виде параллельного программируемого вычислителя.

Потребность в таком переходе на сегодня стала вполне очевидной, т.к. все более полная формализация текстов ЕЯ приводит к быстрому росту числа переменных и уравнений модели, с обработкой которой, особенно в реальном темпе времени, современные последовательные компьютеры не справляются.

Без перехода к параллельным вычислениям возможно только создание приложения уровня ограниченной заданной проблемной областью экспертной системы, не претендующей на полноценную формализацию смысла текстов ЕЯ и, следовательно, на понимание смысла этих текстов.

Основные идеи формализации смысла текстов ЕЯ на базе АКП

Для описания механизма языка эффективным подходом является сравнение естественного языка с языком математики. Математический язык, в отличие от естественного, хорошо формализован, его механизм понятен. Он создавался, в отличие от ЕЯ, не стихийно сам по себе, а целенаправленно отдельными людьми и поэтому весь на виду. Если гипотеза о наличии глубокой аналогии между этими двумя языками оправдается, тогда можно будет найти ответы на многие трудные вопросы, касающиеся механизма ЕЯ, обращаясь за подсказкой к языку математики.

Рассмотрим проблему смысла предложения. Математические тексты, как и тексты естественного языка, состоят из предложений. Предложения, фигурирующие в математических текстах, называют высказываниями. На вопрос: “Что такое смысл высказывания?” ответить нетрудно: как известно, смыслом высказывания является функция зависимости истинностной переменной от предметных переменных этого высказывания. Любое высказывание бывает либо истинным, либо ложным.

Рассмотрим, например, высказывание

$$“x + y = z”, \quad (a),$$

где $x + y = z$ - предметные переменные, под которыми в данном случае понимаются переменные, определенные на множестве $N = \{0, 1, 2, \dots\}$ всех неотрицательных целых чисел.

В таком виде высказывание (a) ни истинно, ни ложно. Но если подставить в него вместо переменных x, y, z какие-нибудь их конкретные значения из множества N , то высказывание (a) обратится в истину или ложь.

Пусть, к примеру, $x = 2, y = 3, z = 5$. Тогда высказывание (a) обращается в тавтологию $2 + 3 = 5$, и говорят, что оно истинно. Если же принять $x = 0, y = 1, z = 2$, то это же высказывание обратится в противоречие $0 + 1 = 2$, в этом случае говорят, что оно ложно. Мы видим, что высказывание (a) заключает в себе некоторую функцию $\xi = f(x, y, z)$, отображающую множество N^3 в множество Σ .

Функции с двоичными значениями, зависящими от значений произвольных аргументов, называются функциями истинности или предикатами. Итак, смыслом математического предложения (высказывания) является предикат

$$\xi = f(x_1, x_2, \dots, x_m),$$

представляющий собою зависимость истинностной переменной ξ от предметных переменных x_1, x_2, \dots, x_m этого высказывания. Любое математическое утверждение представляет собой уравнение вида $f(x_1, x_2, \dots, x_m) = 1$, которое задает вполне определенное отношение, связывающее между собою предметные переменные x_1, x_2, \dots, x_m [3]. Важно отметить, что такая трактовка смысла высказывания в математике общепринята, никаких сомнений в ее правильности ни у кого не возникает. Математик рассматривает каждое высказывание как формулу, а любая формула, всегда выражает какую-то функцию. Именно для выражения функций и были придуманы математиками формулы.

Высказывание же всегда выражает некоторую функцию с двоичными значениями, таким образом, его смыслом (значением) служит вполне определенный предикат.

Мы рассмотрели, что представляет собой смысл математического предложения. Теперь, руководствуясь гипотезой о близком родстве естественного и математического языков, предположим, что такая трактовка смысла применима и к предложениям естественного языка (мы имеем в виду только повествовательные предложения). Иными словами, мы исходим из того, что смыслом каждо-

го предложения естественного языка служит некоторый предикат. А если это так, то к каждому предложению ЕЯ должно быть применимо понятие истинностной переменной с ее двумя значениями “истинно” и “ложно”. Проверяем этот вывод на практике. Берем какое-нибудь предложение ЕЯ, например, “На столе лежит книга”. Взятое само по себе, оно ни истинно, ни ложно. Но если его соотносить с какой-нибудь реальной ситуацией, то предложение, и в самом деле, становится истинным или ложным.

Если смыслом предложения ЕЯ, действительно, служит предикат, то в предложениях должны присутствовать не только истинностные, но также и предметные переменные. Проверяем, так ли это на самом деле. Берем предложение “На столе лежит книга”.

Проведем несложный мысленный эксперимент. Носителю ЕЯ предлагается предложение: “На столе лежит книга”. Слово “стол” указывает на некоторый предмет (назовем его первым); слово “книга” указывает на второй предмет. Затем человеку, - носителю ЕЯ, задается вопрос об истинности этого предложения в данной конкретной ситуации. Предметы выбираются произвольно: они не обязательно должны быть столом и книгой. Т.е., на самом деле, в предложении речь идет не только о предметах, но и о предметных переменных. Обозначим последние символами x и y .

Запишем предложение в более развернутой форме так, чтобы в нем явно появились предметные переменные: «Предмет x есть стол, предмет y есть книга, предмет u лежит на предмете x ». Очевидно, что в такой, более подробной, записи предложение выражает ту же самую мысль, что и его краткая исходная запись.

Таким образом, предложение ЕЯ является, как и математическое высказывание, носителем некоторого предиката, представляющего собой зависимость истинностной переменной предложения от его предметных переменных. Этот предикат мы и примем в качестве формального эквивалента смысла предложения ЕЯ.

Фраза “На столе лежит книга” включает в себе много смысловых вариантов. Один из них мы уже рассмотрели ранее. Рассмотрим другой вариант. Он отличается от первого тем, что теперь предъясняется та же фраза, но указывается только первый предмет. Если этот предмет – не стол, то носитель ЕЯ отреагирует отрицательным ответом, а если – стол, то посмотрит, лежит ли на нем хотя бы одна книга. Если – да, он скажет, что высказывание истинно, а если – нет, то – ложно. Полный смысловой перевод фразы теперь получается несколько иным: “Предмет x есть стол, и существует предмет u такой, что предмет u есть книга, и

предмет u лежит на предмете x ”. Есть и третий вариант, когда указывается только второй предмет. Полный смысл фразы в этом случае запишется в виде: “Предмет u есть книга, и существует предмет x , такой, что предмет x есть стол, и предмет u лежит на предмете x ”. Если же не указывается ни одного из предметов, тогда носитель ЕЯ будет вынужден обратиться к четвертому смысловому варианту: “Существуют предметы x и u такие, что предмет x есть стол, предмет u есть книга, и предмет u лежит на предмете x ”.

Двигаться в направлении полной формализации записи высказывания можно, постепенно заменяя в ней элементы ЕЯ на математические выражения. Когда этот процесс исчерпается, мы получим полностью формализованную запись высказывания.

Сделаем еще один шаг в этом направлении. Рассмотрим высказывание: “Предмет x есть стол, предмет u есть книга, предмет u лежит на предмете x ”.

Оно легко расчленяется на три высказывания:

“Предмет x есть стол”,

”Предмет u есть книга”,

“Предмет u лежит на предмете x ”.

Каждое из этих высказываний имеет свой смысл, а значит, – и свой предикат.

Обозначим предикаты высказываний:

первого – символом $\text{стол}(x)$,

второго – $\text{книга}(y)$,

третьего – $\text{лежит на}(y, x)$.

Записи стол , книга , лежит на теперь выступают в роли имен каких-то предикатов, а каких именно – мы еще не знаем. Ясно, что предикат исходного высказывания можно представить в виде конъюнкции этих трех предикатов. Смысл исходного высказывания теперь запишется в виде $\text{стол}(x) \wedge \text{книга}(y) \wedge \text{лежит на}(y, x)$.

Предикат $\text{лежит на}(y, x)$ можно подвергнуть дальнейшему расчленению.

Введем предикат $\text{лежит}(y)$, понимаемый в смысле “Предмет u лежит”, и предикат $\text{на}(y, x)$, соответствующий высказыванию “Предмет u расположен на предмете x ”.

Ясно, что предикаты $\text{лежит на}(y, x)$ и $\text{лежит}(y) \wedge \text{на}(y, x)$ совпадают.

Допускает дальнейшее расщепление и предикат $\text{лежит}(y)$.

Его можно представить в виде конъюнкции таких предикатов:

настоящее время(y);
лежать(y).

Первый предикат означает: “Предмет y наблюдается в текущий момент времени”, второй: – “Предмету y присуще состояние лежания” (безотносительно ко времени).

Соединяя конъюнкцией все введенные предикаты, записываем рассматриваемое высказывание в виде:

$$\text{стол}(x) \wedge \text{книга}(y) \wedge \text{настоящее время}(y) \wedge \\ \wedge \text{лежать}(y) \wedge \text{на}(y, x).$$

Двигаться дальше в анализе смысловой структуры данного высказывания мы пока не будем.

Чтобы это осуществить, надо будет проникнуть в смысловую структуру слов “стол”, “книга”, “настоящее время”, “лежать” и “на”. Для этого можно обратиться к толковому словарю, выразив смысл каждого из этих слов с помощью приведенных в нем определений, в роли которых выступают словосочетания, составленные из других слов. Когда этот процесс оборвется, нужно собрать слова, оставшиеся невыраженными, и связать их друг с другом системой высказываний, воспринимаемых носителями ЕЯ как истинные.

Эта система высказываний выполняет роль аналога системы аксиом, определяющих первичные понятия математики. А в языке она абстрактно определяет смысл слов, которые не удается выразить с помощью прямых определений через другие слова.

Выводы

В статье рассмотрены основные свойства понятий «знание» и «смысл текста ЕЯ».

Показаны преимущества алгебрологического подхода над другими существующими подходами в решении проблемы полной формализации смысла текста на ЕЯ.

Приведены основные идеи формализации смысла текстов ЕЯ на базе алгебры конечных предикатов.

Продолжая аналогию ЕЯ с языком математики, можно сказать, что аппарат алгебры конечных предикатов развивается по мере накопления знаний о структурах ЕЯ, – так же, как и в аппарате математики создавались новые разделы и структуры под требования современных ей технологий и задач.

Так, через определенное время после создания основ АКП появилась потребность во втором этаже аппарата АКП и была разработана алгебра предикатных операций; с ростом сложности формальных моделей ЕЯ были созданы методы декомпозиции предикатов, на основе которых разработано новое направление развитие аппарата АКП – реляционные сети.

Список литературы

1. Бондаренко М.Ф. Теория интеллекта [Текст]: учеб. / М.Ф. Бондаренко, Ю.П. Шабанов-Кушнарченко. – Х.: СМІТ, 2006. – 576 с.
2. Лингвистическая алгебра как аппарат формализации смысла предложений естественного языка [Текст] / И.А. Лецинская, В.А. Лецинский, Л.Г. Петрова, С.Ю. Шабанов-Кушнарченко // Системы обработки інформації. – 2011. – Вып. 2. – С. 34-38.
3. Дударь З.В. Отношения как предмет формульного описания [Текст] / З.В. Дударь, Р.В. Мельникова, Ю.П. Шабанов-Кушнарченко // Проблемы бионики. – Х.: ХТУРЭ, 1997. – Вып. 48. – С. 94-103.
4. Нонака И. Компания – создатель знания. Зарождение и развитие инноваций в японских фирмах [Текст] / И. Нонака, Х. Такеучи. – М.: Олимп-Бизнес, 2003. – 320 с.

Поступила в редколлегию 25.03.2015

Рецензент: д-р техн. наук, проф. С.Ф. Чалый, Харьковский национальный университет радиоэлектроники, Харьков.

ПРО ФОРМАЛІЗАЦІЮ ЗНАНЬ НА БАЗІ АЛГЕБРИ СКІНЧЕННИХ ПРЕДИКАТІВ

С.Ю. Шабанов-Кушнарченко, О.В. Калиниченко, А.І. Коваленко, О.А. Шматько

Розглянуто проблему обґрунтування вибору підходу для формалізації змісту текстів на природній мові. Проаналізовано основи зв'язку знань і змісту текстів природною мовою, підходів до отримання, подання і формалізації знань. Розглянуто основи застосування апарату алгебри скінчених предикатів і реляційних мереж для формалізації змісту текстів природної мови.

Ключові слова: зміст тексту, предикатна модель, знання, формалізація.

ABOUT KNOWLEDGE FORMALIZATION BASED ALGEBRA OF FINITE PREDICATES

S.Yu. Shabanov-Kushnarenko, O.V. Kalynychenko, A.I. Kovalenko, A.A. Shmatko

The problem of justification of the choice of approach for the formalization of the meaning of texts in the natural-governmental language. Analyzed the communication basis of knowledge and meaning of natural language texts, ap-ing receivable, presentation and formalization of knowledge. The basics of using the device al-algebra of finite predicate and relational networks to formalize the meaning of natural language texts.

Keywords: meaning of the text, the predicate model, knowledge, formalization.