

**МЕТОД ФОРМАЛИЗАЦИИ
КАНОНИЧЕСКОЙ ЛОГИКО-ЛИНГВИСТИЧЕСКОЙ МОДЕЛИ
ЕСТЕСТВЕННО-ЯЗЫКОВОГО ТЕКСТОВОГО СООБЩЕНИЯ
В ДИАЛОГОВЫХ СИСТЕМАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕ-
НИЙ**

к.т.н. О.Я. Лазарева, к.т.н. А.А. Феклистов
(представил д.т.н. Г.В. Певцов)

Предлагается метод формализации канонической логико-лингвистической модели естественно-языкового текстового сообщения, который может быть использован при разработке диалоговых систем поддержки принятия решений.

Постановка проблемы. Повышение качества информационно-аналитической обработки естественно-языковой текстовой информации в диалоговых системах поддержки принятия решений связано с разработкой новых методов формализации текстовых сообщений на естественном языке, в основе которых лежит знаниеориентированный подход [1]. Актуальным направлением является разработка канонических логико-лингвистических моделей текстовых сообщений, ориентированных на логическую обработку в процессе выработки вариантов решения проблем.

Анализ литературы. Проведенный анализ ряда работ [2 – 4] показал, что наиболее эффективным с точки зрения последующей реализации на ЭВМ подходом к формализации таких достаточно сложных информационных структур как естественно-языковые текстовые сообщения является разработка канонических логико-лингвистических моделей (КЛЛМ). Как правило, данные модели характеризуются ограниченным множеством используемых понятий, правил их получения и обработки, достаточным для выполнения на ЭВМ диалоговых задач поддержки принятия решений.

Цель статьи. Целью статьи является представление метода формализации канонической логико-лингвистической модели естественно-языкового текстового сообщения, разрабатываемого на основе знаниеориентированной парадигмы [1]. Основным требованием к данному методу является максимальное (по возможности) соответствие формальной модели логико-лингвистической модели оригинальному представлению естественно-языкового текстового сообщения. Выполнение данного требования

обеспечивается введением пяти правил формирования КЛЛМ, которые составляют суть предлагаемого метода формализации.

1. Правило формирования явного ядерного предиката. Рассмотрим некоторое предложение, выражающее элементарное действие, например: «Иван пишет письмо». Очевидно, что глагол «пишет» выражает действие субъекта «Иван» по отношению к объекту «письмо», что в логике предикатов первого порядка соответствует записи «пишет (Иван, письмо)». Определим данные предикаты как «ядерные предикаты» и представим их в общем виде следующим образом

$$P(a_1, a_2), \quad (1)$$

где a_1 и a_2 – субъектный и объектный аргументы соответственно. Так как отношение между a_1 и a_2 выражено в явном виде (вербализовано), то такие предикаты можно определить как «явные ядерные предикаты». Языковыми выражениями таких предикатов могут выступать, например, такие грамматические классы, как глаголы (*назначать, приземляться*), полные или краткие причастия (*передаваемый, выпускающий, принят, запущен*), деепричастия (*переноса, выделив*), существительные с предикатным значением (*взлет, погрузка, транспортировка*).

2. Правило формирования неявного ядерного предиката. Неявным ядерным предикатом (предикатом-функцией [5]) называется предикат, который формируется при формализации семантических отношений, выраженных в тексте в неявном виде (невербализованных), например, в виде отношений между субъектом и атрибутом. Формальное имя такому предикату назначает когнитолог (специалист по знаниям), а позиции субъекта и объекта занимают элементы, между которыми устанавливается данное отношение. Например: словосочетание «вертолет «Апач»» может быть преобразовано к предикату «имя (вертолет, «Апач»)», где «имя» – это формальное имя, которое вводит когнитолог.

3. Правило декомпозиции многоместного предиката. Обычно предложение на естественном языке имеет достаточно сложную и нерегулярную структуру. В простейшем случае в качестве средства формализации распространенных предложений может быть использован многоместный предикат вида

$$P(a_1, a_2, \dots, a_n), \quad (2)$$

где P – имя главного предиката, a_i – аргумент, имеющий некую функциональную нагрузку (например, субъект действия, объект действия, место, время). Необходимо отметить, что использование многоместных предикатов при логической обработке связано с рядом трудностей [5],

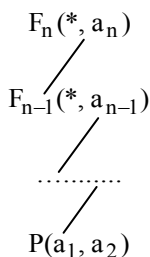
обусловленных сложностью процедуры унификации предикатов, имеющих различную структуру аргументной части. Поэтому выполняется процедура преобразования многоместного предиката к виду взаимосвязанных двухместных предикатов. Рассматривая в выражении (1) некоторый аргумент a_n с учетом его функции по отношению к предикату P , исходное выражение может быть записано как

$$F_n(P(a_1, a_2, \dots), a_n). \quad (3)$$

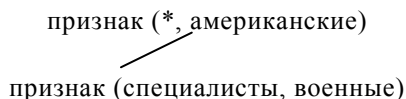
Таким образом, формируется неявный предикат относительно предиката P в зависимости от функции выделенного аргумента. Применяя итеративно правило уменьшения местности исходного предиката, получаем выражение (4), которое полностью эквивалентно исходному (2)

$$F_n(F_{n-1}(\dots(F_3(P(a_1, a_2), a_3), \dots), a_{n-1}), a_n). \quad (4)$$

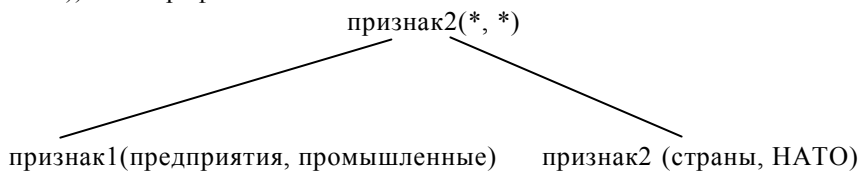
Предикаты F_i ($i \in [3, n]$) отражают семантические отношения, выявленные когнитологом при анализе текста. Полученное выражение можно представить в графической форме в виде древовидной структуры, в которой каждый вложенный предикат связывается с внешним посредством ссылки (виртуального аргумента, который указывает место вложенного предиката):



Применение аппарата ссылок позволяет преобразовать исходное представление КЛЛМ в сетевое представление, которое более наглядно отображает взаимосвязи вложенных предикатов. Это правило также применяется для формализации именных словосочетаний, содержащих более двух слов. Если они распространяются словами, связанными с помощью однородных синтаксических отношений, то в этом случае такое словосочетание представляется как многоместный предикат с именем функции зависимости. Аналогично правилам уменьшения местности явного предиката получаем совокупность вложенных двухместных неявных предикатов. Например: «американские военные специалисты» – признак (признак (специалисты, военные), американские)» или графически:



В случае неоднородности синтаксических связей формируются вложенные предикаты. Например, словосочетание «промышленные предприятия стран НАТО» может быть представлено следующим образом: «признак2 (признак1 (предприятия, промышленные), признак2 (страны, НАТО))» или графически:



4. Правило формирования неявных аргументов. Наиболее естественной функциональной нагрузкой аргументов a_1 и a_2 представляется соответственно субъект и объект действия. Любой из этих аргументов может отсутствовать. В этом случае в модели он может представляться «неявным аргументом», который может быть либо восстановлен по контексту, либо заменен аргументом-переменной или ссылкой. Например, в словосочетании «принято решение» отсутствует субъект действия. При формализации получаем «принять (x, решение)». Замена субъектного аргумента переменной с одной стороны позволяет не нарушать структуры предиката, а с другой стороны предполагает возможность означивания пропущенного аргумента при анализе межфразового контекста. При формализации отношений, в которых отсутствует объект (например, фразы с непереходными глаголами), объектный аргумент вообще может опускаться. Например, словосочетание «спортсмен бежит» представляется как «бежать (спортсмен)». Если обозначить отсутствующий аргумент как «x», то получим предикат «бежать (спортсмен, x)».

Аппарат неявных аргументов является одной из отличительных особенностей предлагаемой КЛЛМ и позволяет формализовать такие естественно-языковые явления как безличные и неопределенно-личные предложения, а также анафорические отношения, выраженные словами-заместителями.

5. Правило формирования предикатов-связок. Наряду с традиционными логическими связками «не», «и», «или», «импликация», «эквиваленция» в КЛЛМ используются логико-лингвистические связки (ЛЛС). Они позволяют формализовать предикаты, выраженные в тексте союзами, союзными словами, знаками препинания. Для унификации процедур логической обработки ЛЛС делятся на функциональные группы, соответствующие трем логическим связкам «и», «или», «не». Преобразование логических и логико-лингвистических связок к предикатному виду осуществляется следую-

шим образом. Пусть p и q предикаты. Тогда выражения (p «и» q), (p «а также» q), («как» p , «так и» q), («не только» p , «но и» q) приобретает вид «и (p , q)»; выражение («не» p , «а» q) – «и (не (p), q)»; выражение («или» p , «или» q) – «или (p , q)»; выражение («ни» p , «ни» q) – «не (p , q)». Аналогичным образом ЛЛС используются и для аргументов предикатов.

Выводы. Предложенный в работе метод формализации канонической логико-лингвистической модели естественно-языкового текстового сообщения ориентирован на максимальное (по возможности) соответствие формальной модели исходному естественно-языковому представлению текстового сообщения. Выполнение данного требования достигается путем введения следующих правил: правило формирования явного ядерного предиката, правило формирования неявного ядерного предиката, правило декомпозиции многоместного предиката, правило формирования неявных аргументов и правило формирования предикатов-связок. Представленные правила отличаются от известных в настоящее время методов обработки предикатных структур, например, правила декомпозиции многоместного предиката в двухместные, предложенного в [5]. Предлагаемый метод формализации канонической логико-лингвистической модели естественно-языкового текстового сообщения может быть использован при разработке диалоговых систем поддержки принятия решений.

ЛИТЕРАТУРА

1. Рось А.А., Замаруева И.В., Феклистов А.А., Назаренко С.Г., Рось В.А., Лазарева О.Я., Лукашин С.О. Концептуальный проект информационных систем нового поколения // Сборник научных трудов VI-й международной конференции "Знание – Диалог – Решение" (KDS-97). Т. 1. – Ялта. – 1997. – С. 37 – 46.
2. Поспелов Д.А. Логико-лингвистические модели в системах управления. – М.: Энергоиздат, 1981. – 232 с.
3. Мельчук И.А. Опыт теории лингвистических моделей "Смысл-Текст". Семантика. Синтаксис. – М.: Наука, 1974. – 314 с.
4. Шенк Р. Обработка концептуальной информации. – М.: Энергия, 1980. – 360 с.
5. Логический подход к искусственному интеллекту: от классической логики к логическому программированию: Пер. с франц. / А. Тейз, П. Грибомон, Ж. Луи и др. – М.: Мир, 1990. – 432 с.

Поступила 30.01.2004

ЛАЗАРЕВА Ольга Ярославовна, канд. техн. наук, ст. научный сотрудник научного центра. В 1981 году окончила ХГУ. Область научных интересов – математическая лингвистика.

ФЕКЛИСТОВ Андрей Александрович, канд. техн. наук, ст. научный сотрудник, начальник лаборатории научного центра. В 1993 году окончил ХВВКИУ РВ. Область научных интересов – искусственный интеллект.
