

Владимир Сергеевич Можаров

Максим Анатольевич Павленко (д-р техн. наук)

Владислав Николаевич Руденко (канд. техн. наук)

Евгений Александрович Судников

Харьковский национальный университет Воздушных Сил имени Ивана Кожедуба, Харьков, Украина

Кировоградская лётная академия Национального авиационного университета, Кировоград, Украина

Национальный университет обороны Украины имени Ивана Черняховского, Киев, Украина

РАЗРАБОТКА МЕТОДА СИНТЕЗА ЕСТЕСТВЕННО-ЯЗЫКОВЫХ СООБЩЕНИЙ

Основным предметом статьи является рассмотрение вопросов взаимодействия операторов с автоматизированными системами управления воздушным движением. Одним из возможных способов взаимодействия является взаимодействие на естественном ограниченном языке. Такой подход требует разработки метода синтеза естественно-языковых сообщений. В статье проведен анализ методов синтеза естественно-языковых сообщений в экспертных системах, определены и проанализированы основные этапы синтеза, а также определена целесообразность использования того или иного метода в различных задачах. Разработаны методы для устранения проблем, которые встречаются при семантическом синтезе естественно-языковых текстов. Разработка данного метода вызвана тем, что в современных системах управления используются интеллектуальные методы обработки информации и поддержки принятия решений оператором. Разработка предложенного метода позволит разработать систему общения, систему естественно-языковой сигнализации и систему пояснения выработанных решений. Экспериментальная проверка приведенного метода показала существенный рост обоснованности и оперативности принятия решений.

Ключевые слова: оператор, система управления, интеллектуальная система, обработка информации, синтез сообщений, естественный язык, синтаксический разбор.

Введение

Постановка проблемы. В настоящее время интеллектуализация предполагает развитие возможностей вычислительных средств в направлении обеспечения совместного с пользователями решения задач, постоянное расширение доли машины в совместной с человеком деятельности по решению трудно-формализуемых задач. Все это требует разработки и совершенствования не только формально-логического аппарата для представления и обработки знаний, но и средств, обеспечивающих взаимодействие пользователей с системой поддержки принятия решений (СППР). Среди множества средств взаимодействия языковые средства занимают ведущее место, так как только с помощью языка (формального или естественного) можно достичь определенных целей в процессе взаимодействия коммуникантов.

Язык представления знаний для СППР с одной стороны отображает знания о предметной области, как части объективного мира, с другой стороны правила использования закономерностей процессе управления воздушными объектами в заданных ситуациях.

Синтез ограниченного естественного языка (ОЕЯ) конструкций должен обеспечивать оптимальное выполнение задач, стоящих перед пользователем.

Анализ литературы. Анализ существующих семантических моделей таких как “ТЕКСТ-СМЫСЛ” [1], модель концептуальных зависимостей В. Шенка [2], модель “семантик предпочтения” Уилкса [3] показал, что единицей восприятия смысла считается понятие.

Построенные на базе этих моделей системы обработки текстовой информации показали недостаточность такого подхода. Это связано с тем, что ЕЯ текст, как форму отображения знаний о предметной области (ПО), недостаточно рассматривать только на уровне анализа языковой системы и систем представления знания о ПО.

Примером этого могут служить наличие в реальных текстах таких классов лексем как собственные имена, обозначения, сокращения др., которые не подчиняются закономерностям ЕЯ анализа, а непосредственно должны обрабатываться на уровне знаний.

Если рассматривать внутренний язык СППР как некую графическую структуру отображающую знания о ПО, то единицей синтеза следует считать не отдельную формулу или теорию (что в частном случае может быть верно), а дискурс, т.е. совокупность фрагментов выделенных по определенным правилам [4–6].

Дискурсионный подход – процесс связного, строго последовательного, ясного рассуждения, в котором каждая последующая мысль вытекает из

предыдущей; зависит от предыдущей и обуславливает последующую [2].

Решаемая прикладная задача фактически управляет выбором дискурса, который далеко не всегда удовлетворяет отношению строго последовательного порядка в базе знаний. Задачу формирования фрагмента базы знаний (дискурса) удовлетворяющего заданной цели будем называть задачей логического синтеза [44].

Применение логического синтеза с одной стороны позволяет синтезировать целенаправленные ЕЯ сообщения и соответственно выдавать пользователю не информацию вообще, а только полезную информацию для решения его задачи, с другой стороны целенаправленный выбор дискурса расширяет возможности семантического и синтаксического методов синтеза. В частности, такой подход дает возможность замены понятий внутри дискурса на соответствующие местоимения, что делает сообщение лаконичным и т.д.

Процессом выделения необходимого фрагмента базы знаний должна управлять целевая установка. Критерием декомпозиции являются знания о решаемой задаче в СППР. Если решается задача объяснения результатов логического вывода – выбираются одни признаки. При пополнении базы знаний – другие признаки, т.е. выбор соответствующего фрагмента осуществляется путем отображения знаний о решаемой задаче на знания о предметной области в терминах внутреннего языка. Результат отображения есть дискурсное значение фрагмента базы знаний.

Затем строится отображение единиц внутреннего языка через соответствующие единицы ЕЯ, которое ставит элементам логической структуры внутреннего языка в соответствии с семантикой понятия ЕЯ. В результате формируется понятийная структура ЕЯ сообщения, в которой понятия связаны логически, но не связаны еще грамматически.

На основе знаний о грамматической структуре ЕЯ производится анализ понятийной структуры и построение грамматической структуры выходного ЕЯ сообщения.

В результате производится три вида отображения знаний на структуры выходного сообщения. Отсюда вытекает необходимость трех этапов обработки данных. Эти этапы преемственны, выходная информация одного является входной информацией для другого.

Целью статьи является повышение качества взаимодействия пользователей с системой поддержки принятия решений АСУ управления воздушного движения при решении трудно формализуемых задач.

Изложение основного материала исследований

Для решения проблемы воспользуемся подходом, который позволяет рассматривать решение задачи синтеза в три этапа:

синтез внутреннего представления ЕЯ - сообщения;

семантическая интерпретация ЕЯ - сообщения; синтез синтаксической и морфологической структуры ЕЯ - сообщения.

Учитывая изложенное заметим, что синтез внутреннего представления структуры ЕЯ-сообщения определяется в основном, понятными методами обработки знаний (результатами выполнения процедур логического вывода и объяснения р), поэтому в дальнейшем более подробно будут рассмотрены методы семантической интерпретации и синтеза синтаксической и грамматической структуры сообщения, считая в качестве исходной информации сформированный на этапе логического вывода дискурс.

Рассмотрим содержание метода семантической интерпретации на примере решения задачи синтеза ЕЯ сообщения, полученного в результате логического вывода при ответе на запрос пользователя. Особенность решения данной задачи в том, что необходимо интерпретировать все элементы ПО. Отличительной особенностью данного метода является его универсальность, за счет использования семантических классов объектов ПО, которые входят в категориальную модель представления знаний. Задача синтеза на этапе семантической интерпретации, включает решение следующих задач:

1) распознавание конструкций внутреннего языка представления базы знаний;

2) интерпретация элементов конструкций (понятий и отношений) внутреннего языка средствами естественного языка.

Согласно модели языка [7] и описанной структуре С-языка [8] предметом распознавания является содержание формулы внутреннего языка, через анализ строк и знаков (символов) заданной формулы. Следовательно, объектом распознавания являются отдельные знаки и строки внутреннего языка.

Определим распознавание знаков в формуле по следующим правилам:

Знак:= символ/ _.

Символ:= буква/ цифра/ специальный символ.

Буква:= Латинская/ русская.

Латинская:= Большая буква (Бб)/ Малая буква (Мб).

Латинская Большая буква (ЛБО):= А/ В/ С/ Д ... /Z.

Латинская малая буква (Лмб):= а/ в/ с/ d ... /z.

Русская:= Большая буква (Бб)/ Малая буква (Мб).

Русская Большая буква (РБб):= А/ Б/ В ... /Я.

Русская малая буква (Рмб):= а/ б/ в ... /я.

Число := цифра/ цифра.

Цифра := 1/ 2/ 3/ 4/ 5/ 6/ 7/ 8/ 9/ 0.

Специальный символ: =, ,, !, ?.

Набор строк внутреннего языка определяется следующими правилами:

- П1 str₁ → <Роль: ИО,ТО,КОЗ>;
- П2 str₂ → <Роль: ИО,СИ>;
- П3 str₃ → <Роль: ИО, ТО>;
- П4 str₄ → <Роль: КОЗ>;
- П5 str₅ → <Роль: СИ>;
- П6 str₆ → <Роль: ИО>;
- П7 str₇ → <ИО,ТО>;
- П8 str₈ → <Роль>;
- П9 str₉ → <Роль: ИО, КОЗ>.

Строки формируют формулу. По отношению необходимости и достаточности можно построить дерево всевозможных семантически правильных формул, полный перечень которых представлен на рис. 1.

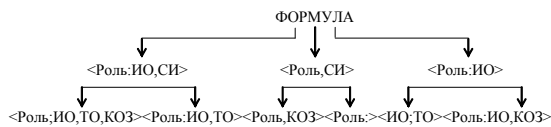


Рис. 1. Структура семантически правильных формул языка взаимодействия

Выходными данными распознавания является информация о семантических классах объектов. Выходные данные: функция, формула. Функция – это команда, которая определяет, что надо сделать с данной формулой записи. Формула – это формализованное описание (или описание на внутреннем языке того, что известно об объекте). Объект объединяет подобъекты.

Функциям := РБб {РЕб} VRБб {Рмб} РБб {Рмб} V VRБб {Рмб} РБб {Рмб} VRБб {Рмб}.

Формула := <начальная строка;
(вторая строка;
:
строка
:
конечная строка)>.

Начальная строка := <Роль: Имя Объекта (ИО), Тип Объекта(ТО),

Код Области Знаний (КОЗ)VRоль: ИО, ТО; VRоль, КОЗ; VRоль; VИО, ТО; V

VRоль: ИО,КОЗ;

Вторая строка := (Роль: Имя Подобъекта (ИПО),

Служебная Информация (СИ);VRоль, СИ;VRоль: ИПО.

Строка := Роль:ИПО, СИ;VRоль, СИ;VRоль: ИПО;

Конечная строка := Роль: ИГО, СИ; VRоль, СИ; VRоль: ИПО)>.

ИО = ИПО := РБб {РБб} VRБб {РБб} Vчисло Рмб {Рмб}

VRмб {Рмб} ЛБб {ЛБб} число

ТО := объект-Рмб {Рмб}.

Роль := Имя Роли (ИР) ИОВИРИПО.

ИР := РБб {Рмб} VRБб.

КОЗ и СИ являются дополнительной информацией и при синтезе ВЯ конструкций не рассматриваются. Выходными данными алгоритма является информация о семантических классах объектов. На основе идентификации исходных данных и правил их объединения, производится распознавание имен объекта, подобъектов и приписывается им соответствующий семантический класс. В результате для нашего примера получаем выражение:

<Объект (имя объекта), семантический класс (тип объекта);

(Роль подобъекта (имя роли (ИР), имя объекта (ИО))):

Роль подобъекта (имя подобъекта (ИПО));

После решения задачи распознавания конструкций внутреннего языка, решаются задачи семантической интерпретации распознанных классов и формирование понятийной структуры фрагмента базы знаний. В качестве исходных данных здесь будем рассматривать объекты, подобъекты и их семантические классы. Подобъекты входят в структуру теории объекта.

ИО = ИПО – это последовательность символов, расположенных между специальными символами: $_ / - / ; /$.

Для нашего примера объектом является: распределение, подобъектами: оператор АСУ УВД, воздушные объекты, воздушный коридор.

После распознавания и приписывания каждому понятию к отношению соответствующего класса, решается задача семантической интерпретации. Для решения этой задачи используется априори заданные таблицы языковой интерпретации всех классов объектов и подобъектов, присутствующих в заданном внутреннем языке.

Фрагменты записи для интерпретации приведенного примера заданы в виде правил:

1. Семантическая интерпретация понятий: <Тип объекта>→<Концепт, грамматическая информация, подобъект>.

2. Семантическая интерпретация отношений: <Имя роли>→>→<Концепт, грамматическая информация, подобъект>

Словарная статья формируется в соответствии с данными таблиц.

Концепт – описание типа объекта языковыми средствами.

Грамматическая информация содержит сведения о требуемом падеже для зависимого подобъекта.

На основе идентификации входных данных и информации, правил их преобразования, строится их понятийная структура выходного текстового сообщения.

<Концепт, грам. инф., подобъект>

<Концепт, грам. инф., подобъект>

Следующая рассматриваемая задача – метод синтеза синтаксической и морфологической структуры естественно-языкового сообщения.

Особенность решения данной задачи в том, что в дискурсе собрана вся информация об объекте, которая может соответствовать одному или нескольким языковым предложениям.

Отличительной особенностью данной модели является то, что она позволяет выделить эти предложения в форме простых нераспространенных предложений. На основе комплексного использования синтаксических и морфологических методов обработки языка.

В качестве исходных данных грамматического синтеза будем рассматривать выходную информацию этапа семантической интерпретации.

Объектом синтеза является понятийная структура сообщений. Предметом синтеза на данном этапе являются грамматические связи между элементами понятийной структуры

Выходными данными являются одно или несколько ЕЯ предложений. На данном этапе обработки конструкции решаются следующие задачи:

- распознавание словоформ по словарю словоформ;
- формирование предложений;
- грамматическое согласование словоформ внутри предложения;
- синтез словоформ по парадигматическому словарю;
- графическое оформление выходного сообщения.

В качестве примера возьмем типовую фразу: “Оценить распределение самолетов по воздушным коридорам”. Словоформам предложения в морфологической зоне словаря будет приписана следующая информация:

- ОЦЕНИТЬ - /Ч.3/.
- РАСПРЕДЕЛЕНИЕ: - /Ч.1(СГ), 1.2, 2.3, 3.1, 4.4/.
- САМОЛЕТОВ- /Ч.1.(А), 1.2, 2.3, 3.2, 4.2/.
- ПО - /Ч.7, 4.3/.
- ВОЗДУШНЫМ - /Ч.1, 1.2, 2.3, 3.2, 4.3/.
- КОРИДОРАМ - /Ч.1(А), 1.2, 2.2, 3.1, 4.2/.

Если при решении задачи анализа был использован словарь словоформ, то рационально его использовать в задаче синтеза переопределив правила выбора грамматической информации. В соответствии с моделью производится поиск и сравнение словоформ с содержимым словаря и выбор нужной информации, сохраняя существующую, т.к. информация о падеже уже определена на этапе семантической интерпретации, то в словаре выбирается информация о части речи, роде и числе, остальная информация считается несущественной.

Далее решается задача разбиения текста на предложения. Предложения формируются на основе следующих правил:

П1 Предложение (ПР):= группа подлежащего (ГР.ПОДЛ.) \wedge группа сказуемого (ГР.СКАЗ.) \wedge группа дополнения (ГР.ДОП.) \vee ГР.

ПОДЛ. \wedge ГР.СКАЗ. \vee ГР.ПОДЛ. - (тире) ГР.ПОДЛ.

П2 ГР.ПОДЛ.: = ГР.П. \wedge С. (И.П.) \vee С. (И.П.) \vee С.(И.П.) \wedge С. (И.П.) \vee Гр.П.

\wedge С.(И.П.) \wedge С. (И.П.) \vee С. (И.П.) \wedge Число.

П3 ГР. П.: = П. \wedge Число.

П4 ГР. СКАЗ.: = Г. \wedge С.(В.П.) \vee Р. \vee - (тире).

П5 ГР. ДОП.: = ГР.С. \vee ГР.П. \wedge ГР.С.

П6 ГР. С: = ПР. + С. (И.П.) \vee С. (И.П.) \vee С. (Р.П.).

В соответствии с моделью далее производится расстановка слов в предложении в соответствии с правилами.

Если Предложение:= ГР.ПОДЛ. : \wedge ГР.СКАЗ. \wedge ГР.ДОП. \vee ГР.

ПОДЛ. \wedge ГР.СКАЗ.

====> ГР.ПОДЛ. \wedge ГР.СКАЗ. \wedge ГР.ДОП.

После решения этой задачи, решается задача грамматического согласования словоформ внутри предложений по правилам синтаксической связи.

Среди синтаксических связей необходимо выделить следующие: согласование, управление и примыкание.

В заключении решается задача графического оформления выходного сообщения. При этом предложение должно начинаться с новой строки с большой буквы, заканчиваться точкой.

Сформированное таким образом ЕЯ выражение выдается пользователю для дальнейшего анализа.

Выводы и перспективы дальнейших исследований

Введение этапа синтеза внутреннего представления ЕЯ-сообщения позволяет формировать дискурс, удовлетворяющий заданной цели. Применение синтеза внутреннего представления ЕЯ-сообщения с одной стороны позволяет синтезировать целенаправленные ЕЯ сообщения и соответственно выдавать пользователю не информацию вообще, а только полезную информацию для решения конкретной задачи, с другой стороны дискурс расширяет возможности семантического и синтаксического методов синтеза. В дальнейшем планируется разработка внутренней и внешней структуры естественно-языковых конструкций, формирования ограниченного словаря и определения требований к предложениям и тексту, формируемым с использованием предложенного подхода.

Литература

1. Мельчук И. А. Опыт теории лингвистических моделей “Смысл – Текст” / И. А. Мельчук – М.: Школа “Языки русской культуры”, 1999. – 346 с. 2. Shank R.

Conceptual information processing / R. Shank – North-Holland Publishing Company, 1988. – p. 237. 3. Wilks Y. Close Engagements with Artificial Companions: Key Social,

Psychological, Ethical and Design Issues / [Электронный ресурс]. – Режим доступа: <http://ssrn.com/abstract=1308500>. 4. Гаврилова Т. А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В. Ф. Хорошевский. – СПб.: Питер, 2001. – 384 с. 5. Темнова Е. В. Современные подходы к изучению дискурса / Язык, сознание, коммуникация: Сб. статей / Отв. ред. В. В. Красных, А. И. Изотов. – М.: МАКС Пресс, 2004. – Вып. 26. С. 24-32. 6. Математические основы эргономических исследований: монография / П. Г. Бердник, Г. А. Кучук, Н. Г. Кучук, Д. Н. Обидин, М. А. Павленко, А. В. Петров, В. Н. Руденко,

О. И. Тимочко. – Кропивницкий : КЛА НАУ, 2016. – 248 с. 7. Судаков Б. Н. Формальное представление семантики естественно-языковых текстов в экспертных системах / Б. Н. Судаков, М. Ю. Ярмак // Вестник Нац. техн. ун-та “ХПИ”: сб. науч. тр. Темат. вып.: Проблемы усовершенствования электрических машин и аппаратов. Теория и практика. – Харьков: НТУ “ХПИ”. 2012. – № 52 (958). С. 184-190. 8. Судаков Б. Н. Подход к разработке модели внутреннего языка для системы поддержки принятия решений / Б. Н. Судаков, А. В. Першин // Системи обробки інформації. – Харьков: ХУ ПС. 2006. Випуск 3 (52). С. 135-144.

РОЗРОБКА МЕТОДУ СИНТЕЗУ ПРИРОДНО-МОВНИХ ПОВІДОМЛЕНЬ

Володимир Сергійович Можаров

Максим Анатолійович Павленко (д-р техн. наук)

Владислав Миколайович Руденко (канд. техн. наук)

Євген Олександрович Судніков

*Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна
Кіровоградська льотна академія Національного авіаційного університету, Кіровоград, Україна
Національний університет оборони України імені Івана Черняхівського, Київ, Україна*

Основним предметом статті є розгляд питань взаємодії операторів з автоматизованими системами управління повітряним рухом. Одним з можливих способів взаємодії є взаємодія на природній обмеженій мові. Такий підхід вимагає розробки методу синтезу природно-мовних повідомлень. У статті проведено аналіз методів синтезу природно-мовних повідомлень в експертних системах, визначено та проаналізовано основні етапи синтезу, а також визначена доцільність використання того чи іншого методу в різних завданнях. Розроблено методи для усунення проблем, які зустрічаються при семантичному синтезі природно-мовних текстів. Розробка даного методу викликана тим, що в сучасних системах управління використовуються інтелектуальні методи обробки інформації та підтримки прийняття рішень оператором. Розробка запропонованого методу дозволить розробити систему спілкування, систему природно-мовної сигналізації і систему пояснення вироблених рішень. Експериментальна перевірка наведеного методу показала істотне зростання обґрунтованості та оперативності прийняття рішень.

Ключові слова: оператор, система управління, інтелектуальна система, обробка інформації, синтез повідомлень, природна мова, синтаксичний розбір.

DEVELOPMENT OF THE METHOD OF SYNTHESIS OF NATURAL LANGUAGE MESSAGES

Volodymyr S. Mozharov

Maksym A. Pavlenko (Doctor of Technical Sciences)

Vladyslav M. Rudenko (Candidate of Technical Sciences)

Yevhen O. Sudnikov

*Kharkiv National University of Air Forces named after Ivan Kozhedub, Kharkiv, Ukraine
Kirovograd Airborne Academy of National Aviation University, Kirovograd, Ukraine
National Defence University of Ukraine named after Ivan Cherniakhovsky, Kyiv, Ukraine*

The main subject of the article is the consideration of the interaction of operators with automated air traffic control systems. One of the possible ways of interaction is interaction in a natural limited language. This approach requires the development of a method for synthesizing natural language messages. The analysis of methods for synthesizing natural language messages in expert systems is made, the main stages of synthesis are determined and analyzed, and the expediency of using this or that method in various problems is determined. Methods have been developed to eliminate problems that occur in the semantic synthesis of natural language texts. The development of this method is due to the fact that modern management systems use intelligent methods of processing information and supporting decision-making by the operator. The development of the proposed method will allow to develop a communication system, a system of natural language alarm and a system for explaining the solutions developed. Experimental verification of the above method has shown a significant increase in the validity and efficiency of decision-making.

Keywords: operator, control system, intellectual system, information processing, message synthesis, natural language, syntactic analysis.

References

1. Melchuk I.A. (1999), Experience of the theory of linguistic models “Meaning – Text”. [Опыт теории лингвистических моделей “Смысл – Текст”], Moscow: School “Languages of Russian Culture”, 346 p. 2. Shank R. (1988),

- Conceptual information processing, North-Holland Publishing Company, p. 237. **3. Wilks Y.** Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues, [Electronic resource]. Access mode: <http://ssrn.com/abstract=1308500>.
- 4. Gavrilova T.A.** (2001), Knowledge bases of intellectual systems. [*Bazy znaniy intellektual'nyh sistem*], T.A. Gavrilova, V.F. Khoroshevsky, St. Petersburg: Peter, 384 p.
- 5. Temnova E.V.** (2004), Modern approaches to the study of discourse. [*Sovremennye podhody k izucheniju diskursa*], Language, consciousness, communication: Sat. articles, Otv. Ed. V.V. Krasnykh, A.I. Izotov, Moscow: MAX Press, Issue. 26. pp. 24-32. **6. Mathematical** foundations of ergonomic studies: monograph. [*Matematicheskie osnovy jergonomicheskikh issledovanij*], P.G. Berdnik, G.A. Kuchuk, N.G. Kuchuk, D.N. Obidin, M.A. Pavlenko, A.V. Petrov, V.N. Rudenko, O.I. Timochko, Kropivnicki: KLA NAU, 2016, 248 p. **7. Sudakov B.N.** (2012), Formal representation of the semantics of natural language texts in expert systems. [*Formal'noe predstavlenie semantiki estestvenno-jazykovykh tekstov v jekspertnyh sistemah*], M. Yu. Yarmak, Vestnik Nat. tech. un-ta "KhPI": Sat. sci. tr. Subject. Issue: Problems of improving electrical machines and apparatus. Theory and practice. Kharkov: NTU "KhPI", No. 52 (958), pp. 184-190.
- 8. Sudakov B.N.** (2006), Approach to the development of the internal language model for the decision support system. [*Podhod k razrabotke modeli vnutrennego jazyka dlja sistemy podderzhki prinjatija reshenij*], A.V. Pershin, Sistemi obrobki informatsii, Kharkov: HU PS. Issue 3 (52), pp. 135-144.