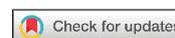


ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

INFORMATION TECHNOLOGIES



УДК 519.711
<https://doi.org/10.37661/1816-0301-2022-19-3-40-49>

Оригинальная статья
Original Paper

Модель представления знаний в системе информационной поддержки принятия решений

С. Ф. Липницкий

*Объединенный институт проблем информатики
Национальной академии наук Беларуси,
ул. Сурганова, 6, Минск, 220012, Беларусь
✉E-mail: lipn@newman.bas-net.by*

Аннотация

Цели. Решается задача математического моделирования знаний в системе информационной поддержки принятия решений. Подход к моделированию основывается на формализации семантических отношений в языке. Формализация таких отношений обеспечивает адекватное представление знаний и служит теоретической базой для алгоритмизации взаимодействия пользователя с системой.

При решении задачи моделирования знаний преследуются три основные цели: построение математической модели семантических отношений, исследование свойств маргинальных синтагм, построение алгоритмов формирования словаря синтагматических структур.

Методы. Используются методы теории множеств, теории графов и математической лингвистики.

Результаты. Разработана математическая модель представления знаний в системе информационной поддержки принятия решений. Формализованы понятия семантических отношений в языке и исследованы свойства маргинальных синтагм. Доказанные свойства таких структур обеспечивают алгоритмизацию информационных процессов в рассматриваемой системе.

Заключение. В качестве реализации предложенной в статье модели разработаны алгоритмы создания словарей синтагм и синтагматических структур. При формировании множества синтагм используются доказанные в статье утверждения о маргинальных синтагмах. Множество синтагматических структур строится с учетом их приведенной типологии. Синтагматические структуры в словаре приобретают статус коммуникативных фрагментов, т. е. устойчивых словосочетаний, из которых синтезируются высказывания языка информационной системы.

Ключевые слова: коммуникативный фрагмент, математическая модель, принятие решений, семантические отношения, синтагма, синтагматическая структура

Для цитирования. Липницкий, С. Ф. Модель представления знаний в системе информационной поддержки принятия решений / С. Ф. Липницкий // Информатика. – 2022. – Т. 19, № 3. – С. 40–49.
<https://doi.org/10.37661/1816-0301-2022-19-3-40-49>

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию | Received 06.04.2022

Подписана в печать | Accepted 26.05.2022

Опубликована | Published 29.09.2022

Model of knowledge representation in the system of information support for decision making

Stanislav F. Lipnitsky

*The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus,
st. Surganova, 6, Minsk, 220012, Belarus
✉E-mail: lipn@newman.bas-net.by*

Abstract

Objectives. The problem of mathematical modeling of the knowledge in the information support system for decision-making is being solved. The approach is based on the formalization of semantic relations in the language. The formalization of such relations provides an adequate representation of knowledge and serves as a theoretical basis for the algorithmization of user interaction with the system.

When solving the problem of knowledge modeling, three main goals are pursued: the construction of a mathematical model of semantic relations, the study of the properties of marginal syntagmas and the construction of algorithms for the formation of a dictionary of syntagmatic structures.

Methods. Methods of set theory, graph theory and mathematical linguistics are used.

Results. A mathematical model of knowledge representation in the information decision support system has been developed. The concepts of semantic relations in the language are formalized and the properties of marginal syntagmas are investigated. The proven properties of such structures provide the algorithmization of information processes in the system under consideration.

Conclusion. As an implementation of the model proposed in the article, algorithms for creating dictionaries of syntagmas and syntagmatic structures have been developed. When forming a set of syntagmas, the assertions about proved in the article marginal syntagmas are used. A set of syntagmatic structures is built taking into account their typology. Syntagmatic structures in the dictionary acquire the status of communicative fragment, i.e. stable phrases to synthesize the information system language.

Keywords: communicative fragment, mathematical model, decision making, semantic relations, syntagma, syntagmatic structure

For citation. Lipnitsky S. F. *Model of knowledge representation in the system of information support for decision making*. Informatika [Informatics], 2022, vol. 19, no. 3, pp. 40–49 (In Russ.).
<https://doi.org/10.37661/1816-0301-2022-19-3-40-49>

Conflict of interest. The author declare of no conflict of interest.

Введение. Система информационной поддержки принятия решений – это программно-лингвистический комплекс, обеспечивающий ведение диалога с пользователем (лицом, принимающим решения). При принятии решений система решает следующие основные задачи: отвечает на вопросы пользователя и задает ему вопросы, снабжает пользователя справочной информацией по существу принимаемого решения, обеспечивает поиск в Интернете аналогов уже принятых решений [1], корректирует запросы пользователя в соответствии с его информационными потребностями, снабжает пользователя рекомендациями и подсказками при описании проблемных ситуаций, синтезирует запросы по описаниям проблемных ситуаций [2–4].

В настоящей статье предлагается подход к моделированию представления знаний в системе информационной поддержки принятия решений. Подход основывается на формализации семантических отношений в языке. Моделирование таких отношений обеспечивает адекватное представление знаний и служит теоретической основой для алгоритмизации диалога системы с пользователем.

Моделирование представления знаний. Формализуем понятие языка ведения диалога пользователя с системой информационной поддержки принятия решений. С этой целью введем специальную порождающую грамматику $G = \langle V, N, I, R \rangle$, где V – непустое множество терминальных элементов, $N = \{I, '\}$ – множество нетерминальных элементов, I – начальный символ, а R – схема грамматики, т. е. множество правил вывода вида $\alpha \rightarrow \beta$ (α и β – различные непустые

стые цепочки в словаре $V \cup N$). Схему R грамматики G определим следующим образом: для любого слова $a \in V$ существуют правила вывода $I \rightarrow a'$ и $a' \rightarrow a$; все остальные правила вывода имеют вид $a' \rightarrow a'b'$ или $a' \rightarrow b'a'$, где $a, b \in V$.

В состав нетерминальных символов введен символ «'» (штрих). В связи с этим грамматику G будем называть штрих-грамматикой.

Приведем пример штрих-грамматики:

$$G_{\text{вх.}} = \langle V, N, I, R \rangle,$$

где $V = \{\text{быстрыми, интеллектуальные, информационные, развиваются, темпами, технологии}\}$;

$$N = \{I, '\};$$

$R = \{I \rightarrow \text{быстрыми}', I \rightarrow \text{интеллектуальные}', I \rightarrow \text{информационные}', I \rightarrow \text{развиваются}', I \rightarrow \text{темпами}', I \rightarrow \text{технологии}', \text{быстрыми}' \rightarrow \text{быстрыми}, \text{интеллектуальные}' \rightarrow \text{интеллектуальные}, \text{информационные}' \rightarrow \text{информационные}, \text{развиваются}' \rightarrow \text{развиваются}, \text{темпами}' \rightarrow \text{темпами}, \text{технологии}' \rightarrow \text{технологии}, \text{технологии}' \rightarrow \text{технологии}' \text{ развиваются}', \text{технологии}' \rightarrow \text{информационные}' \text{ технологии}', \text{технологии}' \rightarrow \text{интеллектуальные}' \text{ технологии}', \text{развиваются}' \rightarrow \text{развиваются}' \text{ темпами}', \text{темпами}' \rightarrow \text{быстрыми}' \text{ темпами}'\}$.

Грамматика G порождает, в частности, следующие цепочки:

- интеллектуальные информационные технологии;
- темпами;
- информационные технологии развиваются;
- интеллектуальные информационные технологии;
- интеллектуальные информационные технологии развиваются быстрыми темпами.

Отношение парадигматического подчинения. В лингвистике различают парадигматические и синтагматические семантические отношения. Парадигматические отношения существуют между словами и фразами языка независимо от контекста и объединяют понятия, обозначающие предметы или явления, между которыми имеется постоянная связь. Так, например, этими отношениями связаны слово «жилище» и фраза «двухэтажный загородный дом», слова «мебель» и «стол», понятия «замок» и «фиксирующий элемент». В противоположность парадигматическим синтагматические отношения возникают в тексте, т. е. между словами и словосочетаниями каждого конкретного его предложения. Синтагматическими отношениями связаны, например, понятия «информационная» и «система» в словосочетании «информационная система», слова «домашнее» и «животное» в выражении «домашнее животное».

В качестве парадигматических отношений рассмотрим отношения парадигматического подчинения и парадигматической эквивалентности.

Обозначим через $L(G)$ язык, порождаемый грамматикой G . Рефлексивное и транзитивное бинарное отношение (предпорядок) Δ на множестве $L(G)$ будем называть отношением парадигматического подчинения, если для любых цепочек $\beta, \gamma, \delta \in L(G)$, таких, что справедливы соотношения $(\beta, \delta) \in \Delta$ и $(\gamma, \delta) \in \Delta$, выполняется одно из соотношений: или $(\beta, \gamma) \in \Delta$, или $(\gamma, \beta) \in \Delta$. Если $(\alpha, \beta) \in \Delta$, то будем говорить, что цепочка β парадигматически подчинена цепочке α .

Считаем, что содержательно отношение парадигматического подчинения соответствует родовидовому отношению между словами и фразами языка. Например, фраза «инструмент для обработки отверстий» и слово «сверло» парадигматически подчинены слову «инструмент». В связи с этим естественно было потребовать, чтобы данное отношение обладало свойствами рефлексивности и транзитивности.

В системе информационной поддержки принятия решений отношение Δ представляется в виде словаря парадигматического подчинения (рис. 1).

Цепочка входного языка	Парадигматически подчиненные цепочки входного языка
...	
Автоматизированное рабочее место	Персональный компьютер Программное обеспечение
...	
Информационная технология	Автоматическое реферирование Информационный поиск Распознавание образов
...	

Рис. 1. Пример словаря парадигматического подчинения
 Fig. 1. An example of the dictionary of paradigmatic subordination

Отношение парадигматической эквивалентности. Бинарное отношение Λ на множестве $L(G)$ назовем отношением парадигматической эквивалентности, или синонимии, если для любых цепочек $\alpha, \beta \in L(G)$ соотношение $(\alpha, \beta) \in \Lambda$ выполняется тогда и только тогда, когда справедливы соотношения $(\alpha, \beta) \in \Delta$ и $(\beta, \alpha) \in \Delta$.

Если $(\alpha, \beta) \in \Lambda$, то цепочки α и β будем называть парадигматически эквивалентными, или синонимами. Видно, что отношение Λ симметрично, поскольку тот факт, что $(\alpha, \beta) \in \Delta$ и $(\beta, \alpha) \in \Delta$, влечет за собой выполнение соотношений $(\alpha, \beta) \in \Lambda$ и $(\beta, \alpha) \in \Lambda$. Рефлексивность и транзитивность отношения Λ также следуют из рефлексивности и транзитивности Δ . Таким образом, Λ есть отношение эквивалентности на множестве цепочек входного языка.

При программной реализации отношение Λ представляется в виде словаря парадигматической эквивалентности, или словаря синонимов (рис. 2).

Цепочка	Синонимичные цепочки
...	
Забастовка	Стачка
...	
ПЭВМ	Персональный компьютер ПК
...	
Языкознание	Лингвистика Языковедение Наука о языке
...	

Рис. 2. Пример словаря парадигматической эквивалентности
 Fig. 2. An example of a paradigmatic equivalence dictionary

Отношение синтаксического подчинения. Синтагматическим является отношение синтаксического подчинения. Оно используется при моделировании синтаксической структуры предложений. В естественных языках различают два вида синтаксического подчинения – предикативность и атрибутивность. Примером предикативности является отношение между сказуемым и подлежащим, атрибутивности – между сказуемым и обстоятельством, подлежащим и определением.

Рассмотрим произвольное предложение $\pi = a_1 a_2 \dots a_n$, где a_1, a_2, \dots, a_n – вхождения слов в это предложение. (Принято говорить о вхождениях слов, а не о словах в связи с тем, что одно и то же слово может повторяться в цепочке несколько раз.) Пусть также μ и ν – некоторые непустые непересекающиеся (не имеющие общих вхождений слов) подцепочки предложения π . Определим формально понятие отношения синтаксического подчинения.

Бинарное отношение Ω_π на множестве всех непересекающихся подцепочек предложения π назовем отношением синтаксического подчинения в предложении π языка $L(G)$, если выполняются следующие условия:

– для любых слов a_i, a_j ($1 \leq i \leq n, 1 \leq j \leq n, i \neq j$) предложения π $(a_i, a_j) \in \Omega_\pi$ тогда и только тогда, когда в выводе предложения π из начального символа I имеются цепочки $\alpha a_i' \beta, \gamma a_i' a_j' \delta$ (или $\gamma a_j' a_i' \delta$), причем цепочка $\gamma a_i' a_j' \delta$ выводима из цепочки $\alpha a_i' \beta$ (соответственно цепочка $\gamma a_j' a_i' \delta$ выводима из цепочки $\alpha a_i' \beta$). Здесь $\alpha, \beta, \gamma, \delta$ – цепочки в словаре $V \cup N$. Некоторые из них (возможно, все) могут быть пустыми;

– для подцепочек μ и ν предложения π $(\mu, \nu) \in \Omega_\pi$ тогда и только тогда, когда существует цепочка $a_i a_j$ предложения π , такая, что $(a_i, a_j) \in \Omega_\pi$, и в выводе предложения π из начального символа I цепочка μ получена из цепочки a_i' , а цепочка ν – из цепочки a_j' .

Синтагмы и синтагматические структуры. Под синтагмой понимают сочетание двух синтаксически связанных слов с указанием направления этой связи. Так, например, в синтагме «пароход плывет» синтаксическая связь направлена от слова «пароход» к слову «плывет», а в синтагме «синий горизонт» – от слова «горизонт» к слову «синий». Другими словами, цепочки «пароход» и «горизонт» являются определяемыми членами соответствующих синтагм, а слова «плывет» и «синий» – определяющими.

Кроме двухсловных сочетаний будем различать цепочки, состоящие более чем из двух слов (например, «интеллектуальная информационная система»), т. е. так называемые синтагматические структуры. Определим формально эти понятия.

Пусть по-прежнему $\pi = a_1 a_2 \dots a_n$ – любое предложение языка, а μ и ν – непустые непересекающиеся подцепочки предложения π .

Если для любых слов a_i, a_j ($1 \leq i \leq n, 1 \leq j \leq n, i \neq j$) предложения π при $i < j$ (или $j < i$) справедливо соотношение $(a_i, a_j) \in \Omega_\pi$, то цепочку $a_i a_j$ (или $a_j a_i$) будем называть синтагмой предложения π . При $j \neq i+1$ (или $i \neq j+1$) синтагму $a_i a_j$ (или $a_j a_i$) назовем разделенной, а при $j = i+1$ (или $i = j+1$) – неразделенной. При этом слово a_i будем считать определяемым, а слово a_j – определяющим членом синтагм $a_i a_j$ и $a_j a_i$.

Если справедливо соотношение $(\mu, \nu) \in \Omega_\pi$, то цепочку $\mu \nu$ (или $\nu \mu$) назовем синтагматической структурой предложения π . Будем также говорить, что μ – определяемый, а ν – определяющий члены синтагматических структур $\mu \nu$ и $\nu \mu$.

Объединение Ω_{Cf} отношений синтаксического подчинения во всех предложениях полного корпуса текстов Cf будем называть отношением синтаксического подчинения в этом корпусе. Синтагмы и синтагматические структуры предложений корпуса Cf назовем его синтагмами и синтагматическими структурами.

Синтаксическое дерево предложения. Если ab – синтагма некоторого предложения π и $(a, b) \in \Omega_\pi$, то будем говорить, что синтаксическая связь направлена от слова a к слову b . Если же $(b, a) \in \Omega_\pi$, то у такой связи противоположное направление. Синтаксические связи между словами предложения представляют обычно в виде ориентированного графа, вершинами которого являются вхождения слов в предложение, а дуги соответствуют синтаксическим связям между ними. Формально понятие синтаксического графа определим следующим образом: ориентированный граф сужения отношения Ω_π на множество всех вхождений слов предложения π назовем синтаксическим графом предложения π ; синтаксическим графом предложения, состоящего из одного слова a , будем считать граф $(\{a\}, \emptyset)$.

Утверждение 1. Синтаксический граф любого предложения языка $L(G)$ является ордеревом (назовем его синтаксическим деревом).

Доказательство. Проведем доказательство индукцией по числу n вхождений слов в предложение. При $n = 1$ и $n = 2$ синтаксическими графами слова и синтагмы являются ордеревья. Предположим, что при $n = k$ синтаксический граф предложения, содержащего k вхождений слов, есть ордеревом. Докажем, что после включения в это предложение еще одного вхождения слова, т. е. при $n = k + 1$, синтаксический граф предложения останется ордеревом. Обозначим включаемое в предложение вхождение слова через b . Тогда в предложении имеется вхождение слова (обозначим его через a), являющееся в силу определения грамматики G определяемым членом синтагмы ab или ba . Если вершину a ордерова с k вершинами соединить с вершиной b дугой (a, b) , то, очевидно, снова будем иметь ордеревом. Утверждение 1 доказано.

Пример синтаксического дерева предложения «Интеллектуальные информационные технологии развиваются быстрыми темпами» представлен на рис. 3.

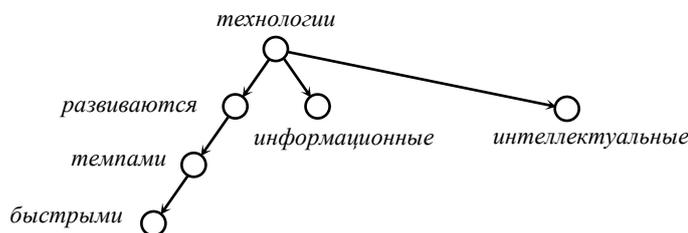


Рис. 3. Пример синтаксического дерева предложения
 Fig. 3. An example of a sentence syntax tree

Синтаксическое дерево предложения является частным случаем более общего понятия – дерева синтаксического подчинения [5] (рис. 4). В рамках рассматриваемой модели данное понятие можно определить следующим образом. Синтаксическое дерево предложения $\pi = a_1 a_2 \dots a_n$ назовем деревом синтаксического подчинения этого предложения, если множеством его вершин является линейно упорядоченное множество $\langle \{a_1, a_2, \dots, a_n\}, \Theta \rangle$. Линейный порядок Θ соответствует естественному порядку слов в предложении, т. е. $i, j = 1, 2, \dots, n-1$ и $r, s = 1, 2, \dots, n$ тогда и только тогда, когда $j < r$. Используя понятия синтаксического дерева и дерева синтаксического подчинения, формализуем свойство проективности предложений языка $L(G)$.

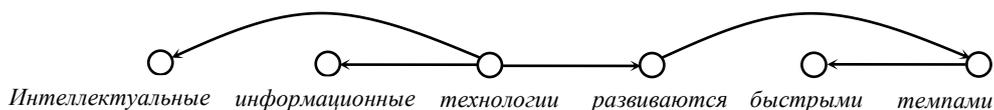


Рис. 4. Пример дерева синтаксического подчинения
 Fig. 4. An example of a syntactic subordination tree

Проективность предложений естественного языка служит критерием их синтаксической правильности. Научные и деловые тексты, как правило, проективны, т. е. состоят из проективных предложений. В художественной литературе, особенно в поэтических произведениях, допустимы непроективные, которые используются в них как художественные средства. В работе [5] показано, что цепочка является проективной, когда в ее дереве синтаксического подчинения отсутствуют пересекающиеся дуги и корень дерева не лежит ни под одной из них. Используя этот наглядный критерий, определим формально понятие проективности в рамках предложенной модели.

Произвольную дугу, соединяющую в любом направлении вершины a_i и a_j дерева синтаксического подчинения предложения π , назовем пересекающейся с любой дугой, соединяющей вершины a_r и a_s (также в произвольном направлении), если $1 \leq i < r < j < s \leq n$. Если же $1 \leq i < r < j \leq n$, то будем говорить, что вершина a_r лежит под дугой, соединяющей вершины a_i и a_j . Предложение языка $L(G)$ будем называть проективным, если в его дереве синтаксического подчинения нет пересекающихся дуг и корень дерева не лежит ни под какой дугой; слабо проективным, если в этом ордереве отсутствуют пересекающиеся дуги, но корень его лежит под некоторой дугой, и непроективным во всех остальных случаях.

Утверждение 2. Все цепочки языка $L(G)$ являются проективными предложениями.

Доказательство. Снова применим индукцию по числу n вхождений слов в предложении. При $n = 1$ и $n = 2$ слова и синтагмы соответственно являются проективными предложениями. Пусть при $n = k$ любое предложение языка $L(G)$ проективное. Докажем, что при $n = k + 1$, т. е. после добавления в это предложение одного вхождения слова, проективность не нарушится. Добавим в предложение из k вхождений слов некоторое вхождение слова b . Тогда в предло-

жении должно присутствовать вхождение слова a , такое, что ab (или ba) – синтагма этого предложения. Предположим от противного, что дуга, соединяющая вершины a и b дерева синтаксического подчинения полученного предложения, пересекается с дугой, которая соединяет вершины c и d некоторой синтагмы cd . Тогда в соответствии с определением проективного предложения один из членов синтагмы ab (или ba) должен находиться между (в смысле линейного порядка \prec) вхождениями слов c и d , что противоречит определению схемы R грамматики G . Утверждение 2 доказано.

Свойства маргинальных синтагм. Пусть $\alpha\beta b\gamma$ (или $\alpha\beta\gamma a$) – произвольное предложение языка $L(G)$, где $\alpha, \beta, \gamma \in V^*$ (V^* – множество всех цепочек в словаре V грамматики G), ab (или ba) – синтагма этого предложения с определяемым членом a и определяющим b . Синтагму ab (или ba) назовем маргинальной синтагмой предложения $\alpha\beta b\gamma$ (или $\alpha\beta\gamma a$), если для любого вхождения слова c ($c \neq b$) данного предложения цепочки bc и cb не являются его синтагмами.

Видно, что определяющие члены всех маргинальных синтагм любого предложения языка $L(G)$ являются висячими вершинами его синтаксического дерева. Слово b синтагм ab и ba будем называть маргинальным словом.

Следующее свойство маргинальных синтагм будем использовать при формировании словаря синтагм.

Утверждение 3. Если ρ – произвольная непустая цепочка в словаре V ($\rho \in V^+$), а ab (или ba) – маргинальная синтагма цепочки ρ , причем в схеме R грамматики G имеется правило вывода $a' \rightarrow a'b'$ (или $a' \rightarrow b'a'$), то цепочка σ , полученная из ρ удалением определяющего члена b синтагмы ab (или ba), является предложением языка $L(G)$ тогда и только тогда, когда $\rho \in L(G)$.

Доказательство. Необходимость. Пусть цепочка σ является предложением языка $L(G)$. Тогда необходимость, т. е. выполнение соотношения $\rho \in L(G)$, следует из факта существования в схеме R грамматики G правил вывода $a' \rightarrow a'b'$ (или $a' \rightarrow b'a'$) и $a' \rightarrow a, b' \rightarrow b$.

Достаточность. Пусть имеется синтагма ab с определяемым членом a и определяющим b . Тогда для цепочки ρ существует вывод $D = (I, \alpha, \beta, \dots, \gamma, \mu a'v, \mu a'b'v, \dots, \mu abv, \dots, \rho)$ в грамматике G , где $\alpha, \beta, \gamma, \mu, v \in V^*$ (V^* – множество всех цепочек в словаре V). Поскольку ab – маргинальная синтагма предложения ρ , то в силу ее маргинальности для любого слова c предложения ρ цепочка bc не является синтагмой, т. е. при выводе предложения ρ не используются правила вида $b' \rightarrow b'c'$, а цепочка $\mu a'b'v$ в выводе D получена из цепочки $\mu a'v$ путем применения правила вывода $a' \rightarrow a'b'$. Если цепочку $\mu a'b'v$ исключить из вывода D , то получим вывод цепочки σ из начального символа I . Аналогично рассматривается случай, когда синтагмой предложения ρ является цепочка ba . Утверждение 3 доказано.

Утверждение 3 обеспечивает получение проективного предложения языка после исключения из него определяющих членов всех маргинальных синтагм в соответствии со следующей циклической процедурой. На первом шаге этой процедуры в анализируемом предложении являются неразделенные маргинальные синтагмы, на втором шаге из этих синтагм исключаются определяющие члены. Далее процесс повторяется аналогичным образом до получения в каждом предложении единственного его слова.

Утверждение 4. Если $a_1 a_2 \dots a_n$ – синтагматическая структура некоторого предложения π языка $L(G)$, такая, что $a_1 a_n$ – синтагма и для всех $i = \overline{1, n-2}$ цепочка $a_i a_{i+1}$ также является синтагмой предложения π , то слово a_{i+1} есть определяющий член синтагмы $a_i a_{i+1}$, причем синтагма $a_{n-2} a_{n-1}$ маргинальная.

Доказательство. Докажем, что слово a_{i+1} есть определяемый член синтагмы $a_i a_{i+1}$. Предположим от противного, что определяющим членом этой синтагмы является слово a_i . Тогда в дереве синтаксического подчинения предложения π вершина a_{i+1} лежит под дугой, соединяющей вершины a_1 и a_n , т. е. цепочка $a_1 a_2 \dots a_n$ слабо проективная, что противоречит проективности предложения π . Если предположить, что слово a_{n-1} не является маргинальным, то снова получим непроективное предложение. Утверждение 4 доказано.

Аналогично доказываются утверждения 5 и 6.

Утверждение 5. Если $a_1a_2 \dots a_n$ – синтагматическая структура некоторого предложения π языка $L(G)$, такая, что a_1a_n – синтагма и для всех $i = \overline{2, n-1}$ цепочка $a_i a_{i+1}$ также есть синтагма предложения π , то слово a_i является определяющим членом синтагмы $a_i a_{i+1}$, а синтагма $a_2 a_3$ – маргинальной.

Утверждение 6. Пусть $a_1a_2 \dots a_n$ – синтагматическая структура некоторого предложения π языка $L(G)$, такая, что a_1a_n – синтагма и для всех $i = \overline{2, l}, j = \overline{l+1, n-1}$ (l – некоторая постоянная, $1 < l < n$) $a_{i-1}a_i, a_j a_{j+1}$ – синтагмы предложения π . Тогда слово a_i является определяющим членом синтагмы $a_{i-1}a_i$, а слово a_j – определяющим членом синтагмы $a_j a_{j+1}$ этого предложения, причем $a_{l-1}a_l$ и $a_{l+1}a_n$ – маргинальные синтагмы.

С помощью утверждений 4–6 можно определить направление синтаксической связи в синтагмах, когда в предложении имеется так называемое гнездование, т. е. когда одна или несколько вершин дерева синтаксического подчинения предложения лежат под дугой (дугами).

При отсутствии гнездования установить определяемый и определяющий члены синтагмы позволяют следующие два утверждения.

Утверждение 7. Если abc – синтагматическая структура предложения $\pi \in L(G)$, такая, что ab и bc – синтагмы, причем слово b – определяющий член синтагмы ab , а слово a – определяемый, то c – определяющий член синтагмы bc , а b – определяемый.

Доказательство. Действительно, если предположить от противного, что определяющим членом синтагмы bc является слово b , то синтаксический граф предложения π не будет ордером. Полученное противоречие доказывает утверждение 7.

Аналогично можно доказать следующее утверждение.

Утверждение 8. Пусть abc – синтагматическая структура предложения π языка $L(G)$, такая, что ab и bc – синтагмы, причем слово b является определяющим членом синтагмы bc , а слово c – определяемым. Тогда у синтагмы ab слово a – определяющий член, а слово b – определяемый.

Утверждения 3–8 позволяют реализовать алгоритмическую процедуру построения словаря синтагм.

Формирование словаря синтагматических структур. Процесс создания словаря синтагматических структур $Dic_{\mu\nu}$ включает два этапа. На первом этапе формируется множество (словарь) синтагм

$$Dic_{ab} = \{ \langle ab \mid (a, b) \in \Omega_{Cf} \cup \Omega_{Cf}^{-1}; a, b \in V; Dir \},$$

а на втором оно дополняется синтагматическими структурами:

$$Dic_{\mu\nu} = \{ \langle \mu\nu \mid (\mu, \nu) \in \Omega_{Cf} \cup \Omega_{Cf}^{-1}; \nu, \mu \in L(G); Dir \},$$

где элементы множества $Dir = \{ \rightarrow; \leftarrow \}$ указывают направление синтаксической связи.

Алгоритм формирования множества синтагм. На входе алгоритма – полный корпус текстов Cf , на выходе – множество (словарь) $Dic_{\mu\nu}$. Алгоритм работает следующим образом.

1. В каждом предложении полного корпуса текстов Cf выбираются все пары смежных слов, и вся совокупность полученных пар упорядочивается по их неубыванию с целью определения частоты встречаемости этих пар во входном тексте. Те пары, частота которых превышает некоторое пороговое значение, будем считать синтагмами.

2. В корпусе Cf выявляются маргинальные синтагмы, а также направления синтаксической связи между словами синтагм с использованием утверждений 3–8. Маргинальные слова всех найденных синтагм из текста удаляются, а сами синтагмы заносятся в множество $Dic_{\mu\nu}$ и исключаются из списка, полученного на первом шаге работы алгоритма. Затем процедура повторяется. Алгоритм заканчивает работу, когда в тексте будут отсутствовать синтагматические структуры, упомянутые в утверждениях 3–8.

Коммуникативные фрагменты, типология синтагматических структур. Традиционно синтез предложений языка (текста) сначала реализуется как последовательная генерация лексем, затем синтаксических фраз и, наконец, предложений по известным синтаксическим правилам. Однако в соответствии с результатами исследования этой проблемы, изложенными в монографии [6], предложения строятся из готовых, хранящихся в памяти компонентов, названных коммуникативными фрагментами. Эти фрагменты не образуются по синтаксическим правилам, а извлекаются из памяти целиком. В системе информационной поддержки процессов принятия решений будем использовать в качестве таких фрагментов информативные синтагматические структуры.

Обозначим через $\pi = a_1 a_2 \dots a_n$ некоторое предложение языка $L(G)$. Рассмотрим возможные типы синтагматических структур в предложении π (буква x в мнемоническом обозначении типа синтагматической структуры означает вхождение слова, а символы \rightarrow и \leftarrow по-прежнему указывают направление синтаксической связи между вхождениями слов и синтагмами предложения π):

$x \rightarrow (x \leftarrow x)$ – синтагматическая структура $a_i a_{i+1} a_{i+2}$ с определяемым членом a_i и определяющим $a_{i+1} a_{i+2}$, где $1 \leq i \leq n-2$, $a_{i+1} a_{i+2}$ – неразделенная синтагма с определяющим членом a_{i+1} и определяемым a_{i+2} , $a_i a_{i+2}$ – разделенная синтагма с определяемым членом a_i и определяющим a_{i+2} ;

$x \leftarrow (x \leftarrow x)$ – синтагматическая структура $a_i a_{i+1} a_{i+2}$ с определяющим членом a_i и определяемым $a_{i+1} a_{i+2}$, где $1 \leq i \leq n-2$, $a_{i+1} a_{i+2}$ – неразделенная синтагма с определяющим членом a_{i+1} и определяемым a_{i+2} , $a_i a_{i+2}$ – разделенная синтагма с определяемым членом a_{i+2} и определяющим a_i ;

$(x \rightarrow x) \leftarrow x$ – синтагматическая структура $a_i a_{i+1} a_{i+2}$ с определяемым членом $a_i a_{i+1}$ и определяющим a_{i+2} , где $1 \leq i \leq n-2$, $a_i a_{i+1}$ – неразделенная синтагма с определяющим членом a_{i+1} и определяемым a_i , $a_i a_{i+2}$ – разделенная синтагма с определяющим членом a_i и определяемым a_{i+2} ;

$(x \rightarrow x) \rightarrow x$ – синтагматическая структура $a_i a_{i+1} a_{i+2}$ с определяющим членом $a_i a_{i+1}$ и определяемым a_{i+2} , где $1 \leq i \leq n-2$, $a_i a_{i+1}$ – неразделенная синтагма с определяющим членом a_{i+1} и определяемым a_i , $a_i a_{i+2}$ – разделенная синтагма с определяемым членом a_i и определяющим a_{i+2} ;

$(x \rightarrow x) \leftarrow (x \leftarrow x)$ – синтагматическая структура $a_i a_{i+1} a_{i+2} a_{i+3}$ с определяющим членом $a_i a_{i+1}$ и определяемым $a_{i+2} a_{i+3}$, где $1 \leq i \leq n-3$, $a_i a_{i+1}$ – неразделенная синтагма с определяющим членом a_{i+1} и определяемым a_i , а $a_{i+2} a_{i+3}$ – неразделенная синтагма с определяющим членом a_{i+2} и определяемым a_{i+3} , $a_i a_{i+3}$ – разделенная синтагма с определяющим членом a_i и определяемым a_{i+3} ;

$(x \rightarrow x) \rightarrow (x \leftarrow x)$ – синтагматическая структура $a_i a_{i+1} a_{i+2} a_{i+3}$ с определяемым членом $a_i a_{i+1}$ и определяющим $a_{i+2} a_{i+3}$, где $1 \leq i \leq n-3$, $a_i a_{i+1}$ – неразделенная синтагма с определяющим членом a_{i+1} и определяемым a_i , а $a_{i+2} a_{i+3}$ – неразделенная синтагма с определяющим членом a_{i+2} и определяемым a_{i+3} , $a_i a_{i+3}$ – разделенная синтагма с определяемым членом a_i и определяющим a_{i+3} .

Алгоритм формирования множества синтагматических структур. На входе алгоритма – множество синтагм Dic_{ab} и полный корпус текстов Cf , на выходе – множество (словарь) $Dic_{\mu\nu}$. Алгоритм включает следующие шаги:

1. Последовательно проводится поиск синтагматических структур в соответствии с их типологией, рассмотренной выше. Мнемонические формулы синтагматических структур заполняются синтагмами из множества Dic_{ab} .

2. Синтагматические структуры, информативность которых превышает некоторое пороговое значение, заносятся в множество $Dic_{\mu\nu}$. Информативность произвольной синтагматической структуры $\mu\nu$ вычисляется по формуле

$$I_{Cf}^{\mu\nu} = n_{Cf}^{\mu\nu} / N_{Cf},$$

где $n_{Cf}^{\mu\nu}$ – количество предложений в полном корпусе текстов Cf , в которых присутствует синтагматическая структура $\mu\nu$, а N_{Cf} – количество всех предложений в корпусе Cf .

Заключение. Предложена математическая модель представления знаний в системе информационной поддержки принятия решений. Подход к моделированию основывается на формализации семантических отношений в языке. Формализация таких отношений обеспечивает адекватное представление знаний и служит теоретической базой для алгоритмизации взаимодействия пользователя с системой.

В качестве реализации предложенной в статье модели разработаны алгоритмы создания словарей синтагм и синтагматических структур. При формировании множества синтагм используются доказанные в статье утверждения о маргинальных синтагмах. Множество синтагматических структур строится с учетом приведенной их типологии. Синтагматические структуры в словаре приобретают статус коммуникативных фрагментов, т. е. устойчивых словосочетаний, из которых синтезируются высказывания языка информационной системы.

Список использованных источников

1. Липницкий, С. Ф. Синтез запросов и поиск альтернатив в системе информационной поддержки принятия решений / С. Ф. Липницкий // Проблемы физики, математики и техники. – 2020. – № 2. – С. 91–95.
2. Симанков, В. С. Методологическое обеспечение этапов поддержки принятия решений при синтезе сложных систем / В. С. Симанков, А. Н. Черкасов // Перспективы науки. – 2012. – № 12. – С. 85–89.
3. Моисеенко, Е. В. Информационные технологии в экономике / Е. В. Моисеенко, Е. Г. Лаврушина ; ред. М. А. Касаткина. – М. : Софт, 2009. – С. 120–135.
4. Кравченко, Т. К. Системы поддержки принятия решений / Т. К. Кравченко // Информационные технологии для современного университета ; под общ. ред. А. Н. Тихонова, А. Д. Иванникова. – М. : ГНИИ ИТТ «Информика», 2011. – С. 107–118.
5. Гладкий, А. В. Синтаксические структуры естественного языка в автоматизированных системах общения / А. В. Гладкий. – М. : Наука, 1985. – 144 с.
6. Гаспаров, Б. М. Язык, память, образ. Лингвистика языкового существования / Б. М. Гаспаров. – М. : Новое литературное обозрение, 1996. – 352 с.

References

1. Lipnitskiy S. F. *Synthesis of queries and search for alternatives in the system of information support for decision-making*. Problemy fiziki, matematiki i tehniki [Problems of Physics, Mathematics and Technology], 2020, no. 2, pp. 91–95 (In Russ.).
2. Simankov V. S., Cherkasov A. N. *Methodological support of decision support stages in the synthesis of complex systems*. Perspektivy nauki [Prospects of Science], 2012, no. 12, pp. 85–89 (In Russ.).
3. Moiseenko E. V., Lavrushina E. G. *Informacionnyye tehnologii v jekonomike*. Information Technologies in the Economy. In Kasatkina M. A. (ed.). Moscow, Soft, 2009, pp. 120–135 (In Russ.).
4. Kravchenko T. K. *Decision support systems*. Informacionnyye tehnologii dlja sovremennogo universiteta [Information Technologies for a Modern University]. In Tikhonov A. N., Ivannikov A. D. (eds.). Moscow, Gosudarstvennyj nauchno-issledovatel'skij institut informacionnyh tehnologij i telekommunikacij "Informika", 2011, pp. 107–118 (In Russ.).
5. Gladkiy A. V. *Sintaksicheskie struktury estestvennogo jazyka v avtomatizirovannyh sistemah obshhenija*. Syntactic Structures of Natural Language in Automated Communication Systems. Moscow, Nauka, 1985, 144 p. (In Russ.).
6. Gasparov B. M. *Jazyk, pamjat', obraz*. Lingvistika jazykovogo sushhestvovaniya. Language, Memory, Image. Linguistics of Linguistic Existence. Moscow, Novoe literaturnoe obozrenie, 1996, 352 p. (In Russ.).

Информация об авторе

Липницкий Станислав Феликсович, доктор технических наук, главный научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: lipn@newman.bas-net.by

Information about the author

Stanislav F. Lipnitsky, D. Sc. (Eng.), Chief Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: lipn@newman.bas-net.by