

Символьное моделирование сообщений и извлечение информации

В. Д. Ильин

ФГУ «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Российская Федерация

Адрес: 119333, Российская Федерация, г. Москва, ул. Вавилова, д. 44-2

vdilyin@yandex.ru

Аннотация

В обзоре представлены основы обновленного подхода к символьному моделированию (s-моделированию) аудио, видео, графических и др. сообщений и решению задач извлечения информации. Значение задач s-моделирования сообщений и извлечения информации определяется их растущей ролью в различных видах деятельности. Доминирование s-моделей сообщений в интеллектуальной деятельности определяется не только их компактностью и выразительностью, но и отсутствием ограничений на типы носителей, применяемых для их хранения. Носителями могут быть память человека, компьютера, смартфона, цифровой камеры или др. Затраты на построение, копирование, передачу, сохранение и накопление s-моделей сообщений (статей электронных энциклопедий, учебников и научных журналов, навигационных карт, видеосообщений, чертежей машин в системах автоматизированного проектирования, медицинских томограмм, записей музыкальных композиций и др.) несопоставимо меньше, чем аналогичные затраты, связанные с несимвольными моделями (макетами судов, зданий и др.). S-моделирование сообщений рассматривается как их отображение в выбранную среду моделирования, выполненное при заданных ограничениях, соответствующих условиям решения задач извлечения информации. Адекватность s-моделей сообщений определяется степенью их соответствия задачам, для решения которых они созданы, и точностью результатов, получаемых при решении этих задач. Предложены обновленные определения понятий s-(символа, кода, сигнала, сообщения, данных, информации). Рассмотрены уточненные s-модели задачи, алгоритма, программы, системы понятий и системы знаний. Задача извлечения информации из сообщения рассматривается как задача интерпретации сообщения на s-модели системы понятий. Приведен краткий обзор работ А. Н. Колмогорова и К. Шеннона, посвященных задачам передачи сообщений. В этих работах говорится о «количестве информации» и рассматриваются задачи, связанные с этим понятием. Понятие «информация» (как результат интерпретации сообщений на моделях систем понятий) там не рассматривается. Приведен пример применения предложенного подхода к анализу задачи перевода с одного языка на другой.

Ключевые слова: символьное моделирование, сообщение, данные, информация, интерпретация сообщения, извлечение информации, модели систем понятий, модели систем знаний

Автор заявляет об отсутствии конфликта интересов.

Для цитирования: Ильин В. Д. Символьное моделирование сообщений и извлечение информации // Современные информационные технологии и ИТ-образование. 2021. Т. 17, № 4. С. 805-813. doi: <https://doi.org/10.25559/SITITO.17.202104.805-813>

© Ильин В. Д., 2021



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Symbolic Modeling of Messages and Information Extraction

V. D. Ilyin

Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow,
Russian Federation

Address: 44 Vavilov St., building 2, Moscow 119333, Russian Federation

vdilyin@yandex.ru

Abstract

The review presents the basics of an updated approach to symbolic modeling (s-modeling) of audio, video, graphic, etc. messages and solving problems of information extraction. The importance of the tasks of s-modeling messages and information extraction is determined by their growing role in various types of activities. The dominance of s-models of messages in intellectual activity is determined not only by their compactness and expressiveness, but also by the absence of restrictions on the types of media used for their storage. The media can be the memory of a person, computer, smartphone, digital camera, etc. The costs of building, copying, transmitting, storing and accumulating s-models of messages (articles of electronic encyclopedias, textbooks and scientific journals, navigation maps, video messages, drawings of machines in computer-aided design systems, medical tomograms, recordings of musical compositions, etc.) are incomparably less than similar costs associated with non-symbolic models (models of ships, buildings, etc.). S-modeling of messages is considered as their mapping into the selected modeling environment, performed under specified constraints corresponding to the conditions for solving information extraction problems. The adequacy of s-models of messages is determined by the degree of their compliance with the tasks for which they were created, and the accuracy of the results obtained when solving these tasks. Updated definitions of the concepts of s-(symbol, code, signal, message, data, information) are proposed. Refined s-models of the problem, algorithm, program, concept system and knowledge system are considered. The task of extracting information from a message is considered as the task of interpreting a message on the s-model of a system of concepts. A brief review of the works of A.N. Kolmogorov and K. Shannon on the problems of message transmission is given. These works talk about the “amount of information” and consider the tasks associated with this concept. The concept of “information” (as a result of interpreting messages on models of concept systems) it is not considered there. An example of the application of the proposed approach to the analysis of the problem of translation from one language to another is given.

Keywords: symbolic modeling, message, data, information, message interpretation, information extraction, models of concept systems, models of knowledge systems

The author declares no conflict of interest.

For citation: Ilyin V.D. Symbolic Modeling of Messages and Information Extraction. *Sovremennye informacionnye tehnologii i IT-obrazovanie = Modern Information Technologies and IT-Education*. 2021; 17(4):805-813. doi: <https://doi.org/10.25559/SITITO.17.202104.805-813>



1. Введение

Извлечение информации из сообщений – постоянно актуальная задача [1-14]. В современных информационных технологиях аудио, визуальные, аудио-визуальные и др. сообщения представлены символами, цифровыми кодами и сигналами. Работа с сообщениями выполняется с помощью программируемых машин (компьютеров, смартфонов и др. компьютерных устройств). В теории *символьного моделирования произвольных объектов (s-моделирования)* программируемые машины названы *s-машинами*¹. S-машины объединены в *человеко-машинную среду поддержки различных видов деятельности (s-среду)*. Инфраструктурным основанием современной s-среды (в наши дни ее называют также *цифровой средой*) служит Интернет.

На стартовом этапе развития s-среды (в середине 1970-х годов) в лекции, посвященной концептуальному обеспечению построения интеллектуальных систем, американскими исследователями Алленом Ньюэллом и Гербертом Саймоном (Allen Newell and Herbert A. Simon) была предложена гипотеза о том, что «система физических символов имеет необходимые и достаточные средства для общего разумного действия». «Физический» означало (у них) «подчиняющийся законам физики» и «пригодный для реализации в виде инженерных систем». Решающее значение эффективного символического представления сущностей (без которого невозможно развитие науки, техники и других видов интеллектуальной деятельности) было подчеркнуто в рассуждениях этих авторов об интеллектуальных системах [15].

Понятие *символьное моделирование* (как научная дисциплина) было определено в 1989 году в монографии, посвященной методологии автоматизированного конструирования программных систем. «Символьное моделирование – научная дисциплина, изучающая процесс построения символических моделей объектов произвольной природы. Объектами символического моделирования могут быть и процесс формальных рассуждений (в математике), и процесс изобретения (как творческого акта, существующего во всех научных областях, в том числе и в математике, и не поддающегося полной формализации)»².

Несимвольное моделирование служит дополнительным (по отношению к s-моделированию) средством изучения объектов на физических моделях (*физическое моделирование*) и/или прототипах (*прототипирование*). И физические модели, и прототипы создаются на основе s-моделей. Напр., в электротехнике s-модель может быть представлена комплектом электронной документации, включающей файлы специфицированного описания, графических изображений (двумерных и/или трехмерных), электрических схем и др.

В настоящее время для изготовления символических и несимвольных моделей объектов [изучаемых в науке (физике, хи-

мии и др.), создаваемых в технике (в авиастроении, робототехнике и др.), медицине (в имплантологии, томографии и др.), искусстве (в архитектуре, музыке и др.) и др. областях деятельности] применяются технологии моделирования с помощью компьютеров и компьютерных устройств (3D-сканеров, 3D-принтеров и др.).

S-модель объекта изучается как отображение в s-среде, выполненное при заданных ограничениях, соответствующих планируемому применению s-модели (напр., цифровое фото – отображение визуального объекта, сделанное с заданным разрешением; программа решения некоторой задачи – отображение выбранного метода решения, учитывающего заданную совокупность ограничений).

В s-моделировании не предполагается никаких ограничений на виды и типы заменяемых s-символами объектов: объекты могут иметь любую физическую сущность, размещение, происхождение и назначение. S-символы одного вида могут заменять s-символы другого (напр., *визуальные символы* могут заменять *аудиосимволы*).

На каждом этапе развития s-среды постоянно актуальными остаются задачи символического, кодового и сигнального представления сообщений, их преобразования, интерпретации, сохранения, накопления, поиска, обмена и информационной безопасности.

Символьные модели (s-модели) систем понятий и систем знаний, в которых представлены результаты изучения сущностей (объектов исследований); программы, определяющие поведение смартфонов, компьютеров и др. s-машин; веб-страницы и файлы документов – всё это *символьные сообщения (s-сообщения)* и их *s-кодовые*³ эквиваленты.

В современной s-среде люди и управляемые ими роботы формируют s-сообщения, представляя их на языках запросов, программирования и др.; выполняют различные преобразования [из аналоговой формы в цифровую и обратно; из несжатой в сжатую и обратно; из одной формы представления документа в другую (*.doc в *.pdf)]; распознают, используют s-сообщения для конструирования новых s-сообщений (программ, документов и др.); интерпретируют на моделях систем понятий (которые хранятся в памяти интерпретатора также в форме s-сообщений); обмениваются s-сообщениями [используя при этом программно-аппаратно реализованные системы правил (*сетевые протоколы*)⁴] [17]; сохраняют и накапливают s-сообщения (создавая электронные библиотеки, энциклопедии и др. информационные ресурсы), занимаются решением задач поиска и защиты s-сообщений⁵ [16-21].

Интерпретация сообщений на моделях систем понятий – задача (впервые рассмотренная в [22]), методы решения которой во многом определяют темпы продвижения в исследованиях, посвященных не только системам искусственного интеллекта.

¹ Ильин А. В., Ильин В. Д. Основы теории s-моделирования. М.: ИПИ РАН, 2009. 143 с. URL: <https://www.elibrary.ru/item.asp?id=25784971> (дата обращения: 10.09.2021).

² Ильин В. Д. Система порождения программ. М.: Наука, 1989. С. 170-171. URL: <https://www.elibrary.ru/item.asp?id=24889586> (дата обращения: 10.09.2021).

³ Ильин В. Д. Код в информатике [Электронный ресурс] // Большая Российская энциклопедия, 2021. URL: https://bigenc.ru/technology_and_technique/text/3958322 (дата обращения: 10.09.2021).

⁴ Ильин В. Д. Компьютерная сеть [Электронный ресурс] // Большая российская энциклопедия, 2021. URL: https://bigenc.ru/technology_and_technique/text/2087809 (дата обращения: 10.09.2021).

⁵ Ильин А. В., Ильин В. Д. S-моделирование объектов информатизации. М.: ИПИ РАН, 2010. 412 с. URL: <https://www.elibrary.ru/item.asp?id=25878776> (дата обращения: 10.09.2021).



1.1. Запись формул и выделение фрагментов текста

Для записи формул и выделения определений, замечаний и примеров используются средства языка *TSM-комплекса* (*TSM: textual symbolic modeling*), разработанного для формализованного описания текстовых *s*-моделей сообщений⁶.

В статье применены следующие средства выделения фрагментов текста:

□ <фрагмент описания> □ ≈ утверждение (определение, аксиома и др.) (здесь и далее символ ≈ заменяет слово «означает»);

◇ <фрагмент описания> ◇ ≈ замечание;

○ <фрагмент описания> ○ ≈ пример.

Курсивом выделены первые вхождения названий понятий и фрагменты описания, к которым автор хочет привлечь внимание.

1.2. Обсуждаемые результаты

В статье представлена часть результатов *методологического обеспечения технологий представления и интерпретации сообщений*. Результаты получены при выполнении научно-исследовательской работы «Моделирование социальных, экономических и экологических процессов» (№ 0063-2016-0005), выполняемой в соответствии с государственным заданием ФАНО России для ФГУ «Федеральный исследовательский центр «Информатика и управление» Российской академии наук.

2. Символы, коды, сигналы

□ *S*-символ – заменитель природного или изобретённого объекта, обозначающий этот объект и являющийся элементом системы построения символьных сообщений (текстов, нотных записей или др.), рассчитанных на восприятие человеком или роботом. □

○ В *s*-моделировании русский алфавит вместе со знаками препинания рассматривается как система текстовых символов для построения сообщений по правилам грамматики русского языка (каждый элемент алфавита является заменителем звука, применяемого в речевых сообщениях); шрифт Брайля для слепых – как система фактурных символов для построения текстовых сообщений, рассчитанных на восприятие осязанием пальцами рук; нотное письмо, система нотных символов – как средство построения графически представленных музыкальных сообщений; система шахматных графических символов – как средство визуального представления шахматных позиций. ○

□ *S*-код – заменитель *s*-символа или символьного сообщения, предназначенный для построения, сохранения, передачи и интерпретации символьных сообщений с помощью *s*-машин. □ ○ Азбука Морзе, коды стандарта Юникод (The Unicode Standard).

В компьютерах, смартфонах и других *s*-машинах *s*-символ представляется в форме *s*-кода, предназначенного для построения, сохранения, передачи и интерпретации символьных сообщений. ◇ Отличительным признаком символьных объектов, существующих в *s*-среде (файлов книг, статей, видеозаписей,

электронных карт, компьютерных программ и др.), являются легко осуществляемые копирование без искажений, распространение и хранение копий (по сравнению с несимвольными объектами: физическими моделями, макетами научно-технических сооружений и др.). ◇

□ *S*-сигнал – физически реализованное представление *s*-символа, рассчитанное на восприятие органами чувств человека (или сенсорами робота), или представление *s*-кода, рассчитанное на прием аппаратными средствами *s*-машин. □

3. Системы понятий и системы знаний

В *s*-моделировании определение системы понятий рассматривается как описание её *s*-модели, сопровождаемое указанием области применимости определения.

□ *S*-модель системы *sc* понятий – <совокупность set^{sc} понятий>, <семейство $rel(set^{sc})$ связей, заданных на set^{sc} >.

В соответствии понятиям системы *sc* поставлены множества их значений.

Область применимости определения задаётся описанием типов:

– корреспондента (на интерпретацию которым рассчитано определение);

– цели, в процессе достижения которой определение целесообразно применять (○ задач, при изучении которых определение может быть полезно ○);

– стадии, на которой имеет смысл использовать определение (○ постановка задачи, создание метода решения ○). □

○ <Система *tr* понятий *треугольник*> ≈ <множество set^{tr} понятий>, <семейство $rel(set^{tr})$ связей, заданных на set^{tr} >.

В *tr* элементами set^{tr} служат стороны треугольника (*a*, *b*, *c*), углы (α , β , γ), периметр *p* и др. Семейство $rel(set^{tr})$ связей включает $p = a + b + c$, $\alpha + \beta + \gamma = \pi$ и др. ○

○ Система $tr^{pi/2}$ понятий *прямоугольный треугольник* может быть определена как специализация *tr*:

$tr^{pi/2} \approx tr [::\alpha = \pi/2]$ (добавлением связи $\alpha = \pi/2$, выделяющей подмножество треугольников, у которых величина одного из углов равна $\pi/2$). ○

Представление связей между понятиями в виде разрешимых задач — необходимое условие построения количественных *s*-моделей систем понятий.

□ *S*-задача – это четвёрка { *Formul*, *Rulsys*, *Alg*, *Prog* }, где *Formul* – постановка задачи; *Rulsys* – множество систем обязательных и ориентирующих требований к решению задачи [23], поставленных в соответствии *Formul*; *Alg* – объединение множеств алгоритмов, каждое из которых соответствует одному элементу из *Rulsys*; *Prog* – объединение множеств программ, каждое из которых поставлено в соответствие одному из элементов *Alg*.

Постановка задачи *Formul* – пара { *Mem*, *Rel* }, где *Mem* – множество понятий задачи, на котором задано разбиение $Mem = Inp \cup Out$ ($Inp \cap Out = \emptyset$) и совокупность *Rel* связей между понятиями, определяющая бинарное отношение $Rel \subset Inp * Out$. Множество *Mem* называем памятью задачи, а *Inp* и *Out* – её входом и выходом, значения которых предполагается соответственно задавать и искать. □

⁶ Ильин В. Д. Символьное моделирование [Электронный ресурс] // Большая российская энциклопедия, 2021. URL: https://bigenc.ru/technology_and_technique/text/4010980 (дата обращения: 10.09.2021).



□ S-алгоритм – система правил решения задачи (соответствующая одному из элементов *Rulsys*), позволяющая за конечное число шагов поставить в однозначное соответствие заданному набору данных, принадлежащему *Inp*, результирующий набор, принадлежащий *Out*. □

□ S-программа – реализованный (на языке программирования высокого уровня, машинно-ориентированном языке и/или в системе машинных команд) s-алгоритм, представленный в форме сообщения, определяющего поведение s-машинного решателя задачи с заданными свойствами. Существует в сим-вольном, кодовом и сигнальном воплощениях, связанных отношениями трансляции]. □

◊ В общем случае множества *Rulsys*, *Alg* и *Prog* могут быть пуста: числа их элементов зависят от степени изученности задачи. ◊

Для каждого элемента из *Rulsys*, *Alg* и *Prog* задано описание применения. Описания применения элементов *Rulsys* включают спецификацию типа *решателя задачи* (автономная s-машина, сетевая кооперация s-машин, кооперация человек-s-машина или др.); требование к информационной безопасности и др. Описания применения элементов из *Alg* включают данные о допустимых режимах работы решателя задачи (автоматический локальный, автоматический распределенный, интерактивный локальный или др.), о требованиях к полученному результату и др. Описания применения программ включают данные о языках программирования, операционных системах и др.

◊ Каждая программа сопровождается ссылками на наборы тестовых примеров. ◊

□ S-модель системы *sk* знаний представлена триадой $\langle ca \approx s\text{-модель системы } sc \text{ понятий } \rangle$, $\langle set^{lnq} \approx s\text{-модель совокупности языков сообщений, интерпретируемых на } ca \rangle$, $\langle set^{intr} \approx s\text{-модель совокупности интерпретаторов на } ca \text{ сообщений, составленных на языках из } set^{lnq} \rangle$. □

Интерпретация сообщения на *ca*:

1. построение выходного сообщения по заданному входному (сообщения представлены на языках из совокупности set^{lnq});
2. анализ выходного сообщения (требуются ли изменения в *ca*);
3. если требуется, то изменение *ca*; если нет – завершение.

○ Онлайн-сервис построения маршрутов основан на системе навигационных знаний. ○

4. Сообщение, данные, информация

□ S-сообщение – конечная упорядоченная совокупность s-символов, рассчитанная на распознавание и интерпретацию получателем, или её s-код, удовлетворяющий требованиям решения базовых задач s-(представления, преобразования, распознавания, конструирования, интерпретации, обмена, сохранения, накопления, поиска и защиты) в s-среде. □

○ S-модели систем понятий и систем знаний, в которых представлены результаты изучения некоторых сущностей (объектов исследований); программы, определяющие поведение s-машин; веб-страницы и файлы документов – всё это s-сообщения. ○

□ Файл – поименованная единица хранения компьютерного кода сообщения (данных или программы) на накопителе (SSD,

жестком диске или др.) компьютера или компьютерного устройства (смартфона, цифровой камеры или др.). □

□ S-данные – s-сообщение, необходимое для решения некоторой задачи или совокупности задач, представленное в форме, рассчитанной на распознавание, преобразование и интерпретацию решателем (программой или человеком).

Специализация s-сообщения (*s-message*) по параметру получатель s-сообщения (*s-recipient*), значением которого является решатель s-задачи (*s-solver*): $s\text{-data} \approx s\text{-message} [::s\text{-recipient} = s\text{-solver}]$. □

□ S-информация – результат интерпретации сообщения на s-модели системы понятий. Для извлечения информации из сообщения необходимо иметь:

– принятое сообщение, представленное в форме, рассчитанной на распознавание и интерпретацию получателем сообщения;

– хранящиеся в памяти модели систем понятий, среди которых – необходимая для интерпретации принятого сообщения;

– механизмы поиска необходимой модели, интерпретации сообщения, представления результата интерпретации в виде сообщения и записи его в память. □

○ Результат интерпретации сообщения m_a , представленного на языке *a*, полученный переводчиком (человеком или роботом) в виде сообщения m_b на языке *b*, – информация, извлечённая из сообщения m_a . ○

○ Экранное представление веб-страницы, рассчитанное на восприятие человеком, – результат интерпретации сообщения, полученного браузером от веб-сервера. ○

5. О работах А. Н. Колмогорова и К. Шеннона

В работах К. Шеннона [24] и А. Н. Колмогорова [25] определены понятия *сообщение* и *информация* введены с целью решения задач количественной оценки объема кода некоторого сообщения и изменения предсказуемости исхода опыта.

Так, для оценки изменения предсказуемости исхода опыта *b* в зависимости от исхода опыта *a* применяется разность энтропий $I(a, b) = H(b) - H_a(b)$, где $H(b)$ и $H_a(b)$ – энтропия исхода опыта *b* при неизвестном и известном исходе опыта *a* соответственно. При этом $I(a, b)$ рассматривается как приращение предсказуемости исхода опыта *b*, если известен исход опыта *a*. Заметим, что содержание опытов *a* и *b* и типы возможных исходов предполагаются заранее известными. Предполагается также, что знание исхода опыта *a* поможет в предсказании исхода опыта *b*. Другими словами, все известно, кроме исхода опыта.

В упомянутой работе [24] К. Шеннон определил основную задачу связи, как «точное или приближенное воспроизведение в некотором месте сообщения, которое было выбрано из некоторого множества возможных сообщений и отправлено из другого места». Он предложил рассматривать эту работу как математическую теорию связи. В коммуникационной модели К. Шеннона определены основные элементы, присущие любой коммуникационной системе, а теория связи К. Шеннона представляет собой методологическое обеспечение технологий кодирования, передачи, декодирования и приема сообщений. К. Шеннон разделяет задачи передачи сообщений и определения их смыслового значения: «семантические аспекты связи



не имеют отношения к технической стороне вопроса): «...часто сообщения имеют значение, т.е. находятся в соответствии с некоторой системой с определенной физической или умозрительной сущностью». Обратим внимание на следующее утверждение К. Шеннона: «Если множество возможных сообщений конечно, то число сообщений или любую монотонную функцию от этого числа можно рассматривать как меру информации, создаваемой выбором сообщения из этого множества, в предположении, что все сообщения равновероятны». Сам по себе выбор сообщения не создает информации. Нетрудно представить, что одно и то же сообщение, может иметь различающиеся результаты интерпретации. Полученные экземпляры могут зависеть и от оценки получателем сообщения его отправителя (как источника сведений), и от того, какие будут применены модели систем понятий и методы интерпретации на выбранных моделях.

6. Возможен ли перевод с одного языка на другой без потери смысла [18]

Человек-переводчик, получив сообщение на языке *A* (*входное сообщение*), сначала интерпретирует его на имеющихся в его памяти *моделях систем понятий*, чтобы понять смысл.

Если требуемых моделей нет или вместо них ошибочно выбраны какие-то другие, или допущены ошибки при интерпретации – смысл сообщения останется недоступным.

♦ Известно, что человек не в состоянии понять сообщения, составленные на его родном языке, если модели систем понятий, необходимые для интерпретации сообщений, отсутствуют в его памяти. ♦

Как-то поняв смысл сообщения, человек-переводчик приступает к формированию сообщения на языке *B* (*выходного сообщения*), стремясь передать *как-то понятый смысл*. Для этого ему необходимо найти в языке *B* *модели систем понятий*, соответствующие тем, которые использовались им (переводчиком!) при интерпретации *входного сообщения*.

♦ Не исключено, что в языке *B* таких систем понятий просто нет или они далеки от смыслового соответствия тем, на основе которых переводчик интерпретировал *входное сообщение*. ♦

6.1. О сообщениях эквивалентных по смыслу

Известно, что лучший перевод может сделать автор сообще-

ния (которому, как никому другому, ясен смысл). При этом автор (обладатель смысла) должен одинаково хорошо владеть языками *A* (на котором составлено *входное сообщение*) и *B* (языком *выходного сообщения*).

♦ И в этом случае сообщения на языках *A* и *B* не обязательно будут эквивалентны по смыслу. Прежде всего, потому, что *эквивалентность сообщений по смыслу трудно определима даже в научно-технических предметных областях* (где принято строго определять системы понятий, на которых надлежит интерпретировать каждое сообщение). ♦

6.2. Необходимые условия перевода без потери смысла

Чтобы без потери смысла *входному сообщению* на языке *A* поставить в соответствие *выходное сообщение* на языке *B* необходимо, чтобы выполнялись три следующих условия:

1. системы понятий предметных областей, к которым относится сообщение, эквивалентны по смыслу в языках *A* и *B*;
2. переводчик (человек или *s*-машинная система) способен найти эквивалентные по смыслу системы понятий, определённые на языках *A* и *B*;
3. переводчик способен составить сообщение на языке *B*, эквивалентное по смыслу сообщению на языке *A*.

6.3. Возможен ли перевод без потери смысла

Есть основания полагать, что к одновременному выполнению необходимых условий перевода без потери смысла можно только приближаться. Чем нетривиальнее смысл сообщения, тем меньше оснований для одновременного выполнения условий (1-3).

7. Заключение

В обзоре рассмотрен обновленный подход к *s*-моделированию сообщений и решению задач *извлечения информации*. Приведены уточненные определения понятий *s*-(*символа, кода, сигнала, сообщения, данных, информации*). Уточнены также *s*-*модели задачи, алгоритма, программы, системы понятий и системы знаний*. Задача извлечения информации из сообщения рассмотрена как задача интерпретации сообщения на *s*-модели системы понятий. Приведен пример применения предложенного подхода к анализу задачи перевода с одного языка на другой.

Список использованных источников

- [1] Cardie C. Empirical Methods in Information Extraction // AI Magazine. 1997. Vol. 18, no. 4. P. 65-79. doi: <https://doi.org/10.1609/aimag.v18i4.1322>
- [2] Califf M. E., Mooney R. J. Bottom-Up Relational Learning of Pattern Matching Rules for Information Extraction // Journal of Machine Learning Research. 2003. Vol. 4. P. 177-210. URL: <https://www.jmlr.org/papers/volume4/califf03a/califf03a.pdf> (дата обращения: 10.09.2021).
- [3] Siefkes C., Siniakov P. An Overview and Classification of Adaptive Approaches to Information Extraction // Journal on Data Semantics IV. Lecture Notes in Computer Science; S. Spaccapietra (ed.). Vol. 3730. Springer, Berlin, Heidelberg, 2005. P. 172-212. doi: https://doi.org/10.1007/11603412_6
- [4] Altinel B., Ganiz M. C. Semantic text classification: A survey of past and recent advances // Information Processing & Management. 2018. Vol. 54, issue 6. P. 1129-1153. doi: <https://doi.org/10.1016/j.ipm.2018.08.001>
- [5] Vo D.-T., Al-Obeidat F., Bagheri E. Extracting temporal and causal relations based on event networks // Information Processing & Management. 2020. Vol. 57, issue 6. Article number: 102319. doi: <https://doi.org/10.1016/j.ipm.2020.102319>



- [6] Contrastive Information Extraction With Generative Transformer / N. Zhang [и др.] // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2021. Vol. 29. P. 3077-3088. doi: <https://doi.org/10.1109/TASLP.2021.3110126>
- [7] Joint Extraction of Retinal Vessels and Centerlines Based on Deep Semantics and Multi-Scaled Cross-Task Aggregation / R. Xu [и др.] // IEEE Journal of Biomedical and Health Informatics. 2021. Vol. 25, issue 7. P. 2722-2732. doi: <https://doi.org/10.1109/JBHI.2020.3044957>
- [8] Liu X., Cheng J., Zhang Q. Multi-Stream Semantics-Guided Dynamic Aggregation Graph Convolution Networks to Extract Overlapping Relations // IEEE Access. 2021. Vol. 9. P. 41861-41875. doi: <https://doi.org/10.1109/ACCESS.2021.3062231>
- [9] Abdollahi A., Pradhan B., Alamri A. VNet: An End-to-End Fully Convolutional Neural Network for Road Extraction From High-Resolution Remote Sensing Data // IEEE Access. 2020. Vol. 8. P. 179424-179436. doi: <https://doi.org/10.1109/ACCESS.2020.3026658>
- [10] LSTM-Based End-to-End Framework for Biomedical Event Extraction / X. Yu [и др.] // IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2020. Vol. 17, no. 6. P. 2029-2039. doi: <https://doi.org/10.1109/TCBB.2019.2916346>
- [11] Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine / J. Wang [и др.] // International Journal of Remote Sensing. 2015. Vol. 36, issue 12. P. 3144-3169. doi: <https://doi.org/10.1080/01431161.2015.1054049>
- [12] GA-ORB: A New Efficient Feature Extraction Algorithm for Multispectral Images Based on Geometric Algebra / R. Wang [и др.] // IEEE Access. 2019. Vol. 7. P. 71235-71244. doi: <https://doi.org/10.1109/ACCESS.2019.2918813>
- [13] Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions / D. Baviskar [и др.] // IEEE Access. 2021. Vol. 9. P. 72894-72936. doi: <https://doi.org/10.1109/ACCESS.2021.3072900>
- [14] Ilyin A. V., Ilyin V. D. Towards a Normalized Economic Mechanism Based on E-services // Agris On-line Papers in Economics and Informatics. 2014. no. 3. P. 39-49. URL: <https://online.agris.cz/archive/2014/03/04> (дата обращения: 10.09.2021).
- [15] Newell A., Simon H. Computer science as empirical inquiry: symbols and search // Communications of the ACM. 1976. Vol. 19, issue 3. P. 113-126. doi: <https://doi.org/10.1145/360018.360022>
- [16] Ilyin V. D. Symbolic Modeling (S-Modeling): an Introduction to Theory // Artificial Intelligence Trends in Systems. CSOC 2022. Lecture Notes in Networks and Systems; R. Silhavy (ed.). Vol. 502. Springer, Cham, 2022. doi: https://doi.org/10.1007/978-3-031-09076-9_54
- [17] Cerf V., Kahn R. A Protocol for Packet Network Intercommunication // IEEE Transactions on Communications. 1974. Vol. 22, no. 5. P. 637-648. doi: <https://doi.org/10.1109/TCOM.1974.1092259>
- [18] Ilyin A. V., Ilyin V. D. Interval Planning the Supplies of Scarce Product // Contemporary Engineering Sciences. 2015. Vol. 8, no. 31. P. 1495-1498. doi: <https://doi.org/10.12988/ces.2015.59263>
- [19] Жирнов В. В., Солонская С. В. Метод преобразования символьных радарных отметок малозаметных подвижных объектов на основе эффекта Тальбота // Радиотехника: Всеукр. межвед. науч.-техн. сб. Харьков: ХНУРЭ, 2021. Вып. 205. С. 129-137. doi: <https://doi.org/10.30837/rt.2021.2.205.14>
- [20] Shvalov D. V., Kravchenko V. A., Shirapov D. Sh. Automated Logic-Mathematical Modeling of Railway Automation Devices Technical Condition // 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon). Vladivostok, Russia: IEEE Press, 2019. P. 1-7. doi: <https://doi.org/10.1109/FarEastCon.2019.8934943>
- [21] Kravchenko V. A., Shirapov D. Sh. Logic-Functional Modeling of Nonlinear Radio Engineering Systems // 2018 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon). Vladivostok, Russia: IEEE Press, 2018. P. 1-6. doi: <https://doi.org/10.1109/FarEastCon.2018.8602769>
- [22] Ильин В. Д., Соколов И. А. Информация как результат интерпретации сообщений на символьных моделях систем понятий // Информационные технологии и вычислительные системы. 2006. № 4. С. 74-82. URL: <https://www.elibrary.ru/item.asp?id=12830934> (дата обращения: 10.09.2021).
- [23] Ilyin A. V., Ilyin V. D. The technology of interactive resource allocation in accordance with the customizable system of rules // Applied Mathematical Sciences. 2013. Vol. 7, no. 143. P. 7105-7111. doi: <http://dx.doi.org/10.12988/ams.2013.311649>
- [24] Shannon C. E. A Mathematical Theory of Communication // The Bell System Technical Journal. 1948. Vol. 27, no. 3. P. 379-423. doi: <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- [25] Колмогоров А. Н. Три подхода к определению понятия «Количество информации» // Проблемы передачи информации. 1965. Т. I, Вып. 1. С. 3-11. URL: <http://mi.mathnet.ru/rus/ppi/v1/i1/p3> (дата обращения: 10.09.2021).

Поступила 10.09.2021; одобрена после рецензирования 09.11.2021; принята к публикации 05.12.2021.

Об авторе:

Ильин Владимир Дмитриевич, ведущий научный сотрудник Вычислительного центра им. А.А. Дородницына РАН, ФГУ «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (119333, Российская Федерация, г. Москва, ул. Вавилова, д. 44-2), доктор технических наук, профессор, ORCID: <https://orcid.org/0000-0002-9761-082X>, vdilyin@yandex.ru

Автор прочитал и одобрил окончательный вариант рукописи.



References

- [1] Cardie C. Empirical Methods in Information Extraction. *AI Magazine*. 1997; 18(4):65-79. (In Eng.) doi: <https://doi.org/10.1609/aimag.v18i4.1322>
- [2] Califf M.E., Mooney R.J. Bottom-Up Relational Learning of Pattern Matching Rules for Information Extraction. *Journal of Machine Learning Research*. 2003; 4:177-210. Available at: <https://www.jmlr.org/papers/volume4/califf03a/califf03a.pdf> (accessed 10.09.2021). (In Eng.)
- [3] Siefkes C., Siniakov P. An Overview and Classification of Adaptive Approaches to Information Extraction. In: Spaccapietra S. (ed.) *Journal on Data Semantics IV. Lecture Notes in Computer Science*. Vol. 3730. Springer, Berlin, Heidelberg; 2005. p. 172-212. (In Eng.) doi: https://doi.org/10.1007/11603412_6
- [4] Altinel B., Ganiz M.C. Semantic text classification: A survey of past and recent advances. *Information Processing & Management*. 2018; 54(6):1129-1153. (In Eng.) doi: <https://doi.org/10.1016/j.ipm.2018.08.001>
- [5] Vo D.-T., Al-Obeidat F., Bagheri E. Extracting temporal and causal relations based on event networks. *Information Processing & Management*. 2020; 57(6):102319. (In Eng.) doi: <https://doi.org/10.1016/j.ipm.2020.102319>
- [6] Zhang N., et al. Contrastive Information Extraction With Generative Transformer. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2021; 29:3077-3088. (In Eng.) doi: <https://doi.org/10.1109/TASLP.2021.3110126>
- [7] Xu R., et al. Joint Extraction of Retinal Vessels and Centerlines Based on Deep Semantics and Multi-Scaled Cross-Task Aggregation. *IEEE Journal of Biomedical and Health Informatics*. 2021; 25(7):2722-2732. (In Eng.) doi: <https://doi.org/10.1109/JBHI.2020.3044957>
- [8] Liu X., Cheng J., Zhang Q. Multi-Stream Semantics-Guided Dynamic Aggregation Graph Convolution Networks to Extract Overlapping Relations. *IEEE Access*. 2021; 9:41861-41875. (In Eng.) doi: <https://doi.org/10.1109/ACCESS.2021.3062231>
- [9] Abdollahi A., Pradhan B., Alamri A. VNet: An End-to-End Fully Convolutional Neural Network for Road Extraction From High-Resolution Remote Sensing Data. *IEEE Access*. 2020; 8:179424-179436. (In Eng.) doi: <https://doi.org/10.1109/ACCESS.2020.3026658>
- [10] Yu X., et al. LSTM-Based End-to-End Framework for Biomedical Event Extraction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2020; 17(6):2029-2039. (In Eng.) doi: <https://doi.org/10.1109/TCBB.2019.2916346>
- [11] Wang J., Song J., Chen M., Yang Z. Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine. *International Journal of Remote Sensing*. 2015; 36(12):3144-3169. (In Eng.) doi: <https://doi.org/10.1080/01431161.2015.1054049>
- [12] Wang R., Zhang W., Shi W., Wang X., Cao W. GA-ORB: A New Efficient Feature Extraction Algorithm for Multispectral Images Based on Geometric Algebra. *IEEE Access*. 2019; 7:71235-71244. (In Eng.) doi: <https://doi.org/10.1109/ACCESS.2019.2918813>
- [13] Baviskar D., Ahirrao S., Potdar V., Kotecha K. Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions. *IEEE Access*. 2021; 9:72894-72936. (In Eng.) doi: <https://doi.org/10.1109/ACCESS.2021.3072900>
- [14] Ilyin A.V., Ilyin V.D. Towards a Normalized Economic Mechanism Based on E-services. *Agris On-line Papers in Economics and Informatics*. 2014; (3):39-49. Available at: <https://online.agris.cz/archive/2014/03/04> (accessed 10.09.2021). (In Eng.)
- [15] Newell A., Simon H. Computer science as empirical inquiry: symbols and search. *Communications of the ACM*. 1976; 19(3):113-126. (In Eng.) doi: <https://doi.org/10.1145/360018.360022>
- [16] Ilyin V.D. Symbolic Modeling (S-Modeling): an Introduction to Theory. In: Silhavy R. (ed.) *Artificial Intelligence Trends in Systems. CSOC 2022. Lecture Notes in Networks and Systems*. Vol. 502. Springer, Cham; 2022. (In Eng.) doi: https://doi.org/10.1007/978-3-031-09076-9_54
- [17] Cerf V., Kahn R. A Protocol for Packet Network Intercommunication. *IEEE Transactions on Communications*. 1974; 22(5):637-648. (In Eng.) doi: <https://doi.org/10.1109/TCOM.1974.1092259>
- [18] Ilyin A.V., Ilyin V.D. Interval Planning the Supplies of Scarce Product. *Contemporary Engineering Sciences*. 2015; 8(31):1495-1498. (In Eng.) doi: <https://doi.org/10.12988/ces.2015.59263>
- [19] Zhyrnov V., Solonskaya S. Metod preobrazovaniya simvol'nykh radarnykh otmetok malozametnykh podvizhnykh ob'yektov na osnove effekta Tal'bota [Method for transforming symbolic radar marks of low-noticeable moving objects based on the Talbot effect]. *Radiotekhnika: All-Ukr. Sci. Interdep. Mag. No. 205*. KNURE, Kharkiv; 2021. p. 129-137. (In Russ., abstract in Eng.) doi: <https://doi.org/10.30837/rt.2021.2.205.14>
- [20] Shvalov D.V., Kravchenko V.A., Shirapov D.Sh. Automated Logic-Mathematical Modeling of Railway Automation Devices Technical Condition. *2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*. IEEE Press, Vladivostok, Russia; 2019. p. 1-7. (In Eng.) doi: <https://doi.org/10.1109/FarEastCon.2019.8934943>
- [21] Kravchenko V.A., Shirapov D.Sh. Logic-Functional Modeling of Nonlinear Radio Engineering Systems. *2018 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*. IEEE Press, Vladivostok, Russia; 2018. p. 1-6. (In Eng.) doi: <https://doi.org/10.1109/FarEastCon.2018.8602769>
- [22] Ilyin V.D., Sokolov I.A. *Informatsiya kak rezul'tat interpretatsii soobshcheniy na simvol'nykh modelyakh sistem ponyatiy* [Information as a result of message interpretation based on s-model of systems of concepts]. *Informacionnye tekhnologii I I vichislitel'nye sistemy = Journal of Information Technologies and Computing Systems*. 2006; (4):74-82. Available at: <https://www.elibrary.ru/item.asp?id=12830934> (accessed 10.09.2021). (In Russ., abstract in Eng.)



- [23] Ilyin A.V., Ilyin V.D. The Technology of Interactive Resource Allocation in Accordance with the Customizable System of Rules. *Applied Mathematical Sciences*. 2013; 7(143):7105-7111. (In Eng.) doi: <http://dx.doi.org/10.12988/ams.2013.311649>
- [24] Shannon C.E. A Mathematical Theory of Communication. *The Bell System Technical Journal*. 1948; 27(3):379-423. (In Eng.) doi: <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- [25] Kolmogorov A.N. *Tri podkhoda k opredeleniyu ponyatiya "Kolichestvo informatsii"* [Three approaches to the definition of the concept "quantity of information"]. *Problemy peredachi informatsii* = Problems of Information Transmission. 1965; 1(1):3-11. Available at: <http://mi.mathnet.ru/rus/ppi/v1/i1/p3> (accessed 10.09.2021). (In Russ.)

Submitted 10.09.2021; approved after reviewing 09.11.2021; accepted for publication 05.12.2021.

About the author:

Vladimir D. Ilyin, Leading scientist of the Dorodnitsyn Computing Centre of the Russian Academy of Sciences, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences (44 Vavilov St., building 2, Moscow 119333, Russian Federation), Dr.Sci. (Tech.), Professor, **ORCID:** <https://orcid.org/0000-0002-9761-082X>, vdilyin@yandex.ru

The author has read and approved the final manuscript.

