

УДК 004.04

Цветков В.Я.

Государственный научно-исследовательский институт информационных и телекоммуникационных технологий «Информика» (г. Москва)

ОБРАБОТКА ИНФОРМАЦИИ С УЧЁТОМ СЕМАНТИЧЕСКОГО СОДЕРЖАНИЯ

V. Tsvetkov

State Research Institute of Information Technologies and Telecommunications «Informica» (Moscow)

SEMANTIC PROCESSING

Аннотация. В статье рассмотрены общие вопросы обработки информации. Объектом анализа является информационное сообщение. Для обозначения содержания сообщения введено понятие контента. В статье описана новая модель обработки информационного сообщения (информации). Эта модель включает первичную и вторичную информацию. Процесс обработки информационного сообщения включает первичную обработку, алгоритмическое преобразование, интерпретацию. Модель включает множество уровней обработки. В статье показано различие между уровнями обработки и этапами обработки на каждом уровне, различие между интерпретацией и алгоритмической обработкой в аспекте изменения содержания информации, а также различие между интерпретацией и алгоритмической обработкой в аспекте однозначности и неоднозначности результата обработки. Результат обработки информации (знания, смысл) может быть получен через множество путей из исходной первичной информации. Модель позволяет получать различные смысловые значения из исходного сообщения.

Ключевые слова: информация, обработка, интерпретация, знания.

В качестве информационного объекта, над которым будем производить анализ, выберем сообщение. Это достаточно исследованный объект как в теории передачи информации, так и в лингвистике [2]. Сообщение имеет какие-то существенные признаки, что в совокупности определяет сущность объекта [4], называемого сообщением. Сообщение имеет внутреннее содержание, которое обозначим как контент.

Определим сообщение как информационный объект, представляющий единое целое, обладающий контентом, передаваемый от процесса к процессу, от «клиента» к «клиенту» или от референта к получателю. В табл. 1 приведены термины и их условные обозначения, используемые в данной работе.

Abstract. The article deals with some general issues of information processing. The object of the analysis is an information message. A new model of message (information) processing is described. It includes primary and secondary information processing. The message processing includes primary processing and algorithmic transformation interpretation. The model includes multiple levels of processing. The article shows the difference between the levels of handling and processing steps at each level. The difference is shown between interpretation and algorithmic processing in the aspect of changes in information content and between interpretation and algorithmic processing in the aspect of uniqueness and ambiguity of the result of processing. The ways of obtaining the results of information processing (knowledge, sense) can be numerous. This model allows of obtaining different senses from the original message.

Key words: information processing, interpretation, knowledge.

Таблица 1

Базовые термины и их обозначения

Русский	Английский	Немецкий	Латынь	Обозначения
Сообщение	message	Nachricht	nuntius	M
Сведения	data	Daten	notitia	D
Знания	knowledge	Wissen	scientia	K
Правило обработки	rule processing	Regel Verarbeitung	лат. regula processui	Rp
Методы интерпретации	methods of interpretation	Methoden der Interpretation	modos interpretationem	Mi
Уровень объекта	the level of the object	die Höhe des Objektes	in gradu obiectum	Lo
Уровень обработки	the level of processing	Die Ebene der Verarbeitung	in gradu processus	Lp
Состояние объекта	state of the object	Zustand der Immobilie	conditio bonis	So
Этап обработки	processing stage	Verarbeitungsstufe	processui scaena	Ps

Будем рассматривать процессы обработки, интерпретации и преобразования информации на примере информационного сообщения. Во всех случаях будем говорить об изменении качества объекта.

Преобразование меняет качество, но не меняет сущность объекта, интерпретация меняет сущность или порождает новую сущность. Для того чтобы провести анализ обработки информации, используем табличную модель, которая отражает различные уровни и этапы обработки.

По одному уровню, то есть по горизонтали, происходит обработка-декодирование,

обработка-преобразование и интерпретация. Иногда декодирование отсутствует, поэтому этот этап объединим в один этап обработки.

В табл. 2 приведена многоуровневая модель обработки и интерпретации информации. Уровни направлены «сверху вниз» и обозначены цифрами в левом столбце. Сущностями одного уровня являются: сообщение, сведения 1(данные), сведения 2 (знания, смысл).

На нумерованных строках показаны процедуры преобразования сущностей от одного уровня к другому.

Таблица 2

Модель обработки и интерпретации информации.

Уровень	Сообщения	Правила обработки	Сведения 1 (данные)	Правила интерпретации	Сведения 2 (знания, смысл)
1	M1	Rpm1	D1	Mi1	DI1
	Rpm1		Rpd1'		Rpdi1
2	M2	Rpm2	D2	Mi2	DI2
	Rpm2		Rpd2'		Rpdi2
3	M3	Rpm3	K3	Mi3	DI3
	Rpm3		Rpd3'		Rpdi3
4	M4	Rpm4	K4	Mi4	KI4
	Rpm4		Rpd4'		Rpdi4
5	M5	Rpm5	K5	Ьш5	KI5

Правила алгоритмической обработки подразумевают решения задач первого рода [3], то есть обработку известными методами по известным алгоритмам. Эту обработку можно назвать объективной, поскольку она не зависит от субъекта и окружения.

Методы интерпретации включают элемент субъективности, поскольку они включают язык, языковое окружение, интеллект и опыт субъекта, осуществляющего интерпретацию. В данном случае субъектом может быть не только человек, но компьютерная программа (например, переводчик текстов) или система искусственного интеллекта.

При использовании алгоритма одни и те же исходные данные дают один и тот же результат при обработке разными субъектами.

При интерпретации разные субъекты дают разные результаты. Разумеется, в основном результаты могут совпадать, но в частностях и деталях могут различаться. Мало того, интерпретация сообщения одним и тем же субъектом по мере изменения его интеллекта или накопления опыта также будет давать разные результаты.

Исходное $M1$ сообщение, или информация, называется первичным. Оно также называется первичной сущностью [4]. Дополнительно существует первичная обработка сообщений $Rpm1$. Она направлена сверху вниз и переводит $M1$ в $M2$ ($M1 \rightarrow M2$). Используя обозначения общей теории систем, запишем этот процесс как:

$$Rpm1: M1 \rightarrow M2 [1].$$

Эти процессы могут продолжаться неограниченно, и сообщение (первичная сущность) может подвергаться многоуровневой обработке:

$$Rpm2: M2 \rightarrow M3; Rpm3: M3 \rightarrow M4; Rpm4: M4 \rightarrow M5 \text{ и т. д.}$$

Такая обработка называется многоуровневой, она характеризует направление обработки «сверху вниз». Возможно другое направление обработки по одному уровню. Первый этап обработки на одном уровне $Rpm1$ представляет собой алгоритмическую обработку вида:

$$Rpm1: M1 \rightarrow D1.$$

В этом случае из первичной информации $M1$ извлекаются сведения и получается новая сущность $D1$, то есть извлекается контент, или контент представляется в новой форме.

Второй этап обработки связан с интерпретацией $M1$ полученных данных $D1$ и получения новой сущности – интерпретированных данных DI (знаний, смысла):

$$Mi1: D1 \rightarrow DI1.$$

Особенность этого преобразования связана с многозначностью результата $DI1$. Как уже отмечалось, это обусловлено языковым окружением, интеллектом и опытом субъекта, осуществляющего интерпретацию.

Например, произведение В. Шекспира может перевести студент, преподаватель английского и специалист в области исследования творчества Шекспира. Однако результат (перевода) будет представлять разные интерпретации одного источника.

Что следует из модели в табл. 2? На каком-то уровне получают новые знания (К). Причём эти знания могут иметь различный смысл, соответствующий уровню обработки. Можно извлекать новые смысловые значения по мере обработки. Кроме того, получение знания, например, $KI5$ на основе первичной информации $M1$ возможно по множеству путей (траекторий).

Таким образом, обработка информации с учётом её семантического содержания влечёт образование нового качества. Это

может быть новое качество исходной сущности, а может быть новая сущность. Это принципиально отличается от подхода, основанного только на кодировании и декодировании.

ЛИТЕРАТУРА:

1. Месарович М., Мако Д., Такахара Н. Теория иерархических многоуровневых систем. – М.: Мир, 1973. – 344 с.
2. Цветков В.Я. Анализ сообщений. – М.: МИИГАиК, 2006. – 140 с.
3. Цветков В.Я. Логика в науке и методы доказательств. – М.: МГОУ, 2012. – 68 с.
4. The works of Aristotle. Volume I / Translated into english under the editorship of W.D. Ross. – Oxford at the Clarendon Press, 1928. – 668 p.