БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ . BULGARIAN ACADEMY OF SCIENCES

ПРОБЛЕМИ НА ТЕХНИЧЕСКАТА КИБЕРНЕТИКА И РОБОТИКАТА, 52
PROBLEMS OF ENGINEERING CYBERNETICS AND ROBOTICS, 52

София . 2001 . Sofia

Information, Data, Knowledge through the Prism of Two Fictional Experiments

Vassil Sgurev

Institute of Information Technologies, 1113 Sofia

1. Information, data, knowledge

Before the computer-communication technologies era, the notion information has not been very widely used. These technologies have accepted its use worldwide.

Information (from the Latin informatio– explanation, description), has not yet any unique and general definition. It is most often regarded as a notion, derived from the exactly formulated "information processes", "information technologies" and "information theory" [1]. Three points of view can be applied towards information:

a) a human-behaviour point, when semantic information is created and processed;

b) an analytic-linguistic point of view, when parts of the information can be defined not only by their content, but by corresponding structures as well;

c) an engineering-physical point of view, when the information can be exactly described by a corresponding quantitative measure.

The first two points of view are closely connected with the generation and processing of semantic information, when the meaning and the content of the messages is emphasized. There does not exist any general method for exact quantitative measurement of the semantic information. Quite often the meaning in the message is evaluated in comparison with the one, which is in the thesaurus of the receiver - a human and more rarely an abstract agent. When different thesauri are available in the receiver, the message obtained can be loaded with different meaning, i.e., it can possess different semantics. As a rule, the notion semantic information is connected with the information processed by man.

The difficulty degree in the definition and quantitative measurement of the semantic information is identical with the difficulty in proving the basic assumption of the powerful thesis of artificial intelligence (AI) –whether the machine (the computer) can think.

The statistical information theory of Shannon [3] is the basis of the engineering-

physical approach for information estimation. In case the messages are transmitted in binary code, the information unit (bit) denotes the entropy of one digit in the binary number, at equal probability for accepting the value 0 or 1.

This way for evaluation of the information quantity operates better in messages exchange between receiver-transmitting devices.

The data are usually considered as information that can be set in a given format with respect to its specific use [1]. Three features of the data stored and processed in computers, can be noted:

– data contained in the computer memory, which are an object of manipulation and processing by program instructions;

– data, which are a final result of computer processing of any input information;

– and in a more modern aspect, information which is separated from the text, sound and image and serves to record the same.

In such interpretation no semantic loading of the data is supposed.

The notion knowledge, that was recently spread is used as a main instrument in artificial intelligence and expert systems [2, 4]. In knowledge bases the same are recorded by sets of implicative relations "if..., then ..." and by confirmations "A is an axiom". Using a machine for inference from the initial knowledge, new can be obtained. In each separate case external experts assign the appropriate content to each axiom and implicative relation of "if..., then..." type.

In practice when the information technologies are applied in general, and the AI tools – especially, quite often some interpretations of the terms information, data and knowledge appear, that cause dubiousity and contradictions [6, 8, 9]:

– the data are regarded as something objectively existing which is transferred into information after human perception;

– information means just this, which is accepted, processed and created by man, i.e., a sign of equality is put between the notions information and semantic information;

– the narrow scientific notion knowledge, used in AI formalized tools and knowledge in its wide meaning, studied in humanitarian sciences as a purely human phenomenon, are regarded as equivalent.

The results derived from two fictional experiments – of the Chinese room [5] and that of the mobile satellite [7] can bring in certain clarity in the contradictions appearing when interpreting the three terms – information, data and knowledge.

2. Information, data, knowledge and the Chinese room experiment

The debates between the defenders and opponents of AI strong thesis, concerning the fundamental problem whether the machine can think, continue many decades with variable success [8, 9].

This problem is not basic in the present paper, but it has direct relation with the question whether in modern digital computers on the basis of appropriate software, it is possible to achieve the effect of thinking and to generate, esteem and process semantic information.

The answer to this question can be acquired by the famous fictional experiment, suggested by D. Sirl in 1980.

It is assumed, that a man, who knows English only, is placed in an isolated room together with drawers, containing numbered sheets with Chinese hieroglyphs. An instruction in English is provided, in which it is unambiguously said that if a Chinese hieroglyph is handed from outside, the man must take out a corresponding

hieroglyph from a given box and hand it out of the room. When the instruction is correctly set, the behaviour of the "Chinese room" will appear to a high degree intelligent for the Chinese being outside, who lead the dialogue with the help of sheets with hieroglyphs. The man in the room, manipulating the hieroglyphs, does not know Chinese and he cannot learn it by these procedures. His actions are not semantically engaged, they are a simple manipulation of symbols.

The author of the imaginary experiment considers the manipulation of hieroglyphs as symbol processing in a digital computer and this is completely reasonable. Hence, implementing symbols processing only, the computer cannot assign rational meaning to the data, i.e., it cannot create and transform semantic information.

Four rules follow from the Chinese room experiment: the computer programs are formal syntactic objects; the human brain operates with reasonable content, i.e. – with semantics; the syntax does not form semantics and it is not at all sufficient for semantics existence; the brain generates intellect (reason).

The Chinese room experiment dissents the AI powerful thesis only in the type, in which it is recently defined using modern digital computers with Von Noiman's architecture, which manipulate symbols only and cannot produce semantics. But this point of view could be altered when other principally different computer tools and technologies appear in the future.

Another consequence from the fictional experiment described deserves also attention, and it refers to knowledge processing. The relations "if..., then ..." and "A is an axiom", called knowledge, are recorded in modern computers and processed by logical rules for inference using software instructions in the same way, as the data. From the view point of the Chinese room experiment, this is the same symbols manipulation as data processing. Hence formalized knowledge introduced in AI represents symbols of knowledge and being such they are a specific subclass of the wider notion data. These symbols of knowledge do not lead to semantics invention, i.e., they should be regarded as suitable symbolic formalism without any consideration loading.

3. Fictional experiment with the wandering satellite

There is an intuitive attempt among computer sciences specialists to use the notion information for data, meant for man only and to enable him give it any intelligent meaning, i.e., generate semantic information. This narrows the term information to its subclass of semantic information.

The incorrectness of this approach can be demonstrated by the following imaginary experiment with the movable satellite [7].

Let us assume that restricted in volume and not updated data are recorded in a satellite, and they are periodically and continuously transmitted to the environment. In this experiment information means only such data that reach man and are used by him. It is supposed that due to unknown reasons, the satellite is not controllable by the Earth and there is such an orbit, which enables its cyclic entry within the receiver range of the earth stations and its quiting. It follows from this, that one and the same data reach or do not reach the man, i.e., they represent periodically either information or non-information. Nonsense.

In order to avoid the above contradiction, the information must be defined as something which exists objectively and is not obligatory connected with man's perception.

The information should be generally considered as any universal property for

matter reflection, with status analogous to the notions matter and energy. In a narrower scientific-applied aspect the notion information must include different physical signals, as well as data above defined, symbols of knowledge and semantic information, i.e.

(1) {information}={physical signals, data, knowledge symbols, semantic information}.

The sections between the first three subsets in the right side of the upper equality have not to be empty sets obligatory. For example, formalized knowledge can be regarded simultaneously as data and vice versa.

In such an interpretation the information does not depend and is not limited by classes of systems, which acquire, transmit and process it – computer-communication, man-machine or such from persons. In this way the information is exchanged and processed by artificial agents (the contemporary computer-communication systems), by these agents and people, as well as by teams. In the first case only data and formalized knowledge is processed, in the last two cases semantic information is added towards them. In the three cases this is realized within the frames of the general notion information used in [1]. In such an interpretation of the information all the contradictions derived from the experiment of the Chinese room and that of the moving satellite are avoided.

It is useful to note that equality (1) has certain historical character and it fixes the recent stage of man-machine and computer-communication technologies.

In future in case it is proved that the adequate modeling of natural intellect, i.e., the generation and processing of semantic information outside man, becomes possible with the help of principally new artificial systems, the relation in (1) can acquire completely another form.

4. Conclusion

In modern information technologies and AI tools quite often there appears ambiguousity and contradictions in the interpretation of the notions information, data and knowledge.

The use of the suggested fictional experiment of the wandering satellite and that of the Chinese room enables the avoiding of some of these contradictions. On the basis of these two imaginary experiments the following inferences and recommendations can be done:

1. The distribution of the notion semantic information on computer-communication systems, working independently on man, is not correct upto now. In the modern state of these systems it is appropriate to use this notion only in man-machine systems, in which the generation, interpretation and processing of the semantic information is not entirely a prerogative of man.

2. The acquisition, processing and transmission of data in computer-communication systems should be regarded as manipulation of symbols only, without any semantic loading.

3. The notion knowledge used in AI system is in its essence a set of knowledge symbols and in this aspect it can be regarded as a specific subclass of the notion data. The notion knowledge spread in AI should be regarded in a limited scientific aspect as a set of symbols without any semantic loading. In wide sense the notion knowledge is appropriately connected with the human way of information processing as it is accepted in humanitarian sciences.

4. The information should be most generally considered as a universal property for matter reflection with status analogous to the notions matter and energy. In a narrow scientific-applied aspect the notion information should include different physical signals, data, symbols of knowledge and semantic information.

References

1. Dictionary of Computing. Oxford University Press, 1986.
2. Shapiro, S. C. Encyclopedia of Artificial Intelligence. V. I, 2. John Willey &Sons, Inc., 1987.
3. Shannon, C. E. A mathematical theory of communication. Pt. I, II – Bell Syst. Techn. J., **27**, 1948, No 3, 379-423.
4. Wielinga, B. J., B. Bredewey, J. Breuker. Knowledge acquisition for expert system. – In: Proceedings of ACAI88, 1988.
5. Scientific American. Nov. 1991, No 11.
6. Sendov, B. Data, nformation, knowledge and wisdom. – In: Proc. SEARCC'92, Cuala Lampur, Malaysia, August, 11-14, I, 1992.
7. Sgurev, V. About artificial intelligence, information and something more. – Automation and informatics. No 5/6, 1998, 7-10.
8. Dreyfus, H. What Computers Can't Do. N.Y., Harper and Row Publ., 1976.
9. Goldkind, S. Machines and Intelligence. N.Y., Greenwood Press, 1987.

Информация, данные, знания в свете двух мисленных экспериментов

Васил Сгурев

Институт информационных технологий, 1113 София

(Резюме)

Рассматриваются некоторые двусмысленности и противоречия, возникающие при применении трех взаимосвязанных понятий – информация, данные, знания, в человеко-машинных системах. На основе предложенного мысленного эксперимента о блуждающем спутнике и эксперимента о китайской комнате даны некоторые выводы, уточнения и рекомендации о применении этих трех понятий.

При современном состоянии человеко-машинных систем целесообразно процессы сотворения, обработки и оценки семантической информации сохранить целиком за человеком. Сбор, обработка и передача данных в компьютерных системах следует рассматривать как манипуляцию символов без какой-либо семантики.

Используемое в системах искусственого интелекта понятие знания является ненагруженной семантикой совокупностью символов знания, и следовательно специфической частью данных. В широком смысле знания необходимо связывать только с человеческим способом переработки информации, как это принято в гуманитарных науках.

В целом информацию следует рассматривать как всеобщее свойство отражения материи со статусом, аналогичном понятиям материи и энергии. В более узком научно-прикладном аспекте, в информацию целесообразно включить физические сигналы, данные, символы знания и семантическую информацию.

Такой подход отражает теперешное конкретно историческое состояние компьютерной обработки информации.