

# Knowledge Processing as Structure Transformation <sup>†</sup>

Mark Burgin <sup>1</sup>, Rao Mikkilineni <sup>2,\*</sup> and Samir Mittal <sup>3</sup>

<sup>1</sup> University of California, Los Angeles (UCLA), 700 Tiverton Ave Los Angeles, CA 90095, USA; mburgin@math.ucla.edu

<sup>2</sup> C3DNA, 7533 Kingsbury Ct., Cupertino, CA 95014, USA

<sup>3</sup> SCUTI AI, 271 Ely Place Palo Alto, CA 94306, USA; samir@scutiai.com

\* Correspondence: rao@c3dna.com; Tel.: +1-408-406-7639

<sup>†</sup> Presented at the IS4SI 2017 Summit DIGITALISATION FOR A SUSTAINABLE SOCIETY, Gothenburg, Sweden, 12–16 June 2017.

Published: 9 June 2017

**Abstract:** As big enterprises and consumers communicate, collaborate and conduct commerce almost at the speed of light using voice, data and video, information explosion (a term first used in 1941, according to the Oxford English Dictionary) has created a need for its accumulation, processing and integration to create “knowledge.” Knowledge processing, in turn, allows us to use the information to make strategic decisions and improve the efficiency of the processes involved. Therefore, knowledge processing systems, their theory and practice are receiving renewed focus. These systems include processes and activities such as cognition, knowledge production, learning, knowledge acquisition, reasoning, management and application. In this paper we discuss how knowledge processing can be viewed as manipulation of various knowledge structures and their transformation. We argue that efficient organization of knowledge processing has to be based on structure transformations of data represented in a symbolic form.

**Keywords:** knowledge structures; cognition; knowledge transformation; named data networks

---

## 1. Introduction

Knowledge processing is not only a central problem of Artificial Intelligence (AI) but also an urgent dilemma of contemporary society and especially, in business and industry.

As Bray [1] writes, organizational knowledge processes deal with the creation, distribution, use and exchange of knowledge for purposes of value creation. They involve managing the intellectual capital of organizations. These processes are best understood with the ecology and ecosystem metaphors. Performative organizational knowledge is a knowledge ecology—a system consisting of many sources, venues, forms and species of knowledge agents in a symbiotic relationship of productive exchange and value creation.

In order to process knowledge and derive value from it, we need to understand the relationship between data, information and knowledge and create knowledge structures. According to Mark Burgin [2] knowledge is derived from information obtained from data resulting from study, experience or instruction. Information gives knowledge of a specific event or situation; or information provides intelligence to make decisions and take action. In this paper we present various structure transformation techniques for knowledge processing.

## 2. Knowledge Processing

It is possible to understand knowledge processing in three different ways:

1. Transformation of data into knowledge
2. Changing the form of knowledge representation

### 3. Deriving new knowledge from a given knowledge

However, whatever understanding we take, it is evident that knowledge processing involves manipulation with knowledge structures. Thus, to be able to efficiently process knowledge using computers and networks, it is essential to know and properly use knowledge structures. The most basic knowledge structures are described in the *synthetic theory of knowledge* presented in [2]. This theory shows that there are several levels or types of data, which by an enrichment (enhancement) process become knowledge.

The first type is *raw* or *un-interpreted data*. Their first-order structure is represented by Picture (1).

$$U \xrightarrow{q} A \quad (1)$$

In this Picture, U consists of some objects, q is a relation between U and A, while A denotes: (a) attributes of these objects in the case of descriptive data, (b) representations of these objects in the case of representational data, and (c) operations, action and processes related to these objects in the case of operational data. For instance, all three types of data are used in the object-oriented programming for object description. Namely, object-oriented programming (OOP) is a programming paradigm based on the concept of abstract objects represented by data structures that include attributes in the form of descriptive data, characteristics in the form of representational data and methods in the form of operational data.

Raw data correspond to the substantial component of knowledge [2]. The difference is that in contrast to the substantial component of knowledge, raw data are not related to any definite property.

Having raw data, a person or a computer system can transform them into knowledge by means of additional knowledge that this person or computer system already has.

The second type of data is *formally interpreted data*. They are related to the abstract property P, forming a named subset of the information component of a knowledge unit represented by Picture (1), and the first-order structure of this component is represented by Picture (2).

$$N \xrightarrow{p} L \quad (2)$$

Here N consists of the names of objects and L is the set of values (the scale) of the property  $P = (N, p, L)$  on the names of objects that tentatively have these properties, while p is the relation that connects names of considered objects with values of the ascribed properties of these objects, i.e., p is the functional component (evaluation function) of the property P.

Formally interpreted data correspond to the symbolic component of knowledge [2]. The difference is that formally interpreted data are not necessarily included in a knowledge system.

The third type is *attributed data*. They are related not to the abstract property P but to the values of an intrinsic property, i.e., to the attribute A. The first-order structure of attributed data is represented by Picture (3).

$$A \xrightarrow{g} L \quad (3)$$

The fourth type of data is *naming data*, the first-order structure of which is represented by Picture (4).

$$U \xrightarrow{f} N \quad (4)$$

There are two more types of data, which are more enhanced and are closer to knowledge. The fourth type is *object interpreted data*. Their first-order structure is represented by Picture (5).

$$\begin{array}{ccc}
 & & L \\
 & & \uparrow p \\
 U & \xrightarrow{f} & N
 \end{array} \quad (5)$$

The fifth type of data is *object attributed data*. Their first-order structure is represented by Picture (6).

$$\begin{array}{ccc}
 A & \xrightarrow{g} & L \\
 \uparrow q & & \\
 N & &
 \end{array} \quad (6)$$

It is interesting to remark that the statement about the correspondence between linguistic constructions representing knowledge and things in the external world as a necessary component of knowledge, which makes it different from data, was discovered in [3] and then reiterated by Davis [4] and Burgin [5,6].

It is possible to treat all types of data as incomplete knowledge.

Formally, naming data are names of considered objects and correspond to the naming component of knowledge [2]. The difference is that naming data are not necessarily included in a knowledge system.

It is important to understand that data and knowledge can themselves be objects, which have names, as well as intrinsic and ascribed properties. In particular, when the domain  $U$  consists of (some kind of) data, then in this case, we come to *named data*, which play an important role in the recent ideas for the development of the Internet [7,8]. To understand what named data are and why they are so popular, we consider the schema of the data transfer on the Internet.

The contemporary Internet is based on the TCP/IP communication protocol. In it, the TCP (Transmission Control Protocol) part performs separation of the file/message into packets on the source computer and reassembling the received packets at the destination, e.g., at the recipient computer. The IP (Internet Protocol) part handles the address of the destination computer so that each packet is routed (sent) to its proper destination.

In Named Data Networking (NDN) architecture for the future Internet, the transmitted packets of data carry data names rather than source or destination addresses. The developers of this architecture believe that this conceptually simple shift will have far-reaching implications for how people design, develop, deploy, and use networks and applications. The Named Data principle implies that a communication network should allow a user to focus on the data identified by their names he or she needs, rather than having to reference a specific, physical location where that data would be retrieved.

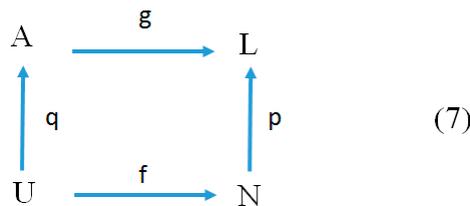
Actually, Internet packets of data are already named by destination addresses and the new approach suggests changing these names to the original data names (identifiers). It is assumed that

such a renaming brings potential for a wide range of benefits such as simpler configuration of network devices, building security into the network at the data level and content caching to reduce congestion and improve delivery speed. In addition, sustained growth in e-commerce, digital media, social networking, and smartphone applications has led to prevailing use of the Internet in the role of a distribution network. Utilization of a point-to-point communication protocol in distribution networks is complex and error-prone, while Named Data Networking better suits distribution environment.

In this context, named sets give a natural mathematical model for named data, which form the naming component of the knowledge quanta with data as their object or domain [9]. Consequently, named set theory provides powerful means for network algorithms and procedures in the form of various operations and correspondences [10].

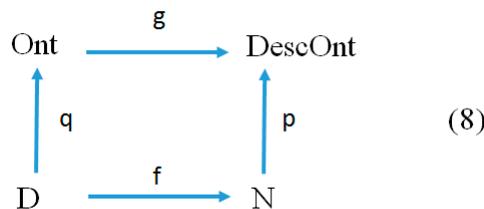
Another example of naming data is named graphs, which are a key structure of the Semantic Web architecture [11]. In it, a set of Resource Description Framework (RDF) statements (a graph) are identified using a URI (Universal Resource Identifier), allowing derivation of descriptions of context, provenance information or other metadata. This shows that named graphs form an extension of the RDF data model giving additional evidence for importance of naming data in contemporary information technology.

Combining all previous Pictures together, we get the quantum knowledge structure or structure of a knowledge quantum.



In Picture (7), U is the knowledge domain (knowledge object), A is an aspect of the domain (object) U, the symbol N, which denotes the name of U or a class of names of the objects from U (a name of the object U), and L is the set of values (the scale) of the property  $P = (N, p, L)$  on the names of objects.

There is a tendency to treat knowledge as data with ontology. This approach is represented by Picture (8).



In Picture (8), D denotes raw data, Ont is an ontology or interpretation of the data D, the symbol N denotes the names of the elements from the raw data D, and DescOnt is a description of the ontology *Ont*, which provides understanding of the ontology Ont.

Thus, we see that efficient organization of knowledge processing has to be based on structure transformations of data represented in a symbolic form.

### 3. Conclusions

Explication of structural features of knowledge processing in this work is aimed at achieving highly efficient semantically based computing and networking in general and at the development of an advanced semantic web, in particular.

**Conflicts of Interest:** The authors declare no conflict of interest.

### References

1. Bray, D. Organizational Dynamics of Technology-Based Innovation: Diversifying the Research Agenda. In *IFIP International Federation for Information Processing*; McMaster, T., Wastell, D., Femeley, E., DeGross, J., Eds.; Springer: Boston, MA, USA, 2007; Volume 235, pp. 457–462.
2. Burgin, M. *Theory of Knowledge: Structures and Processes*; World Scientific: New York, NY, USA; London, UK; Singapore, 2016.
3. Burgin, M. Knowledge in Intelligent Systems. In *Proceedings of the Conference on Intelligent Management Systems, Varna, Bulgaria, 27 June–1 July 1989*; pp. 281–286.
4. Davis, R.; Shrobe, H.; Szolovits, P. What is knowledge representation? *AI Mag.* **1993**, *14*, 17–33.
5. Burgin, M. Data, Information, and Knowledge. *Information* **2004**, *7*, 47–57.
6. Burgin, M. *Theory of Information: Fundamentality, Diversity and Unification*; World Scientific: New York, NY, USA; London, UK; Singapore, 2010.
7. Jacobson, V.; Smetters, D.K.; Thornton, J.D.; Plass, M.F.; Briggs, N.H.; Braynard, R.L. Networking Named Content. *CACM* **2012**, *55*, 117–124.
8. Ntuli, N.; Han, S. Detecting router cache snooping in Named Data Networking. In *Proceedings of the International Conference on ICT Convergence (ICTC), Jeju, Korea, 15–17 October 2012*; pp. 714–718.
9. Burgin, M.; Tandon, A. Naming and its Regularities in Distributed Environments. In *Proceedings of the 2006 International Conference on Foundations of Computer Science, Las Vegas, NV, USA, 26–29 June 2006*; pp. 10–16.
10. Burgin, M. *Theory of Named Sets*; Mathematics Research Developments, Nova Science: New York, NY, USA, 2011.
11. Carroll, J.J.; Bizer, C.; Hayes, P.; Stickler, P. Named graphs, provenance and trust. In *Proceedings of the 14th International Conference on World Wide Web (WWW '05), Amsterdam, The Netherlands, 12–15 December 2005*; pp. 613–622.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>)