computer sciences and mathematics forum

MDPI

Proceeding Paper

General Theory of Information, Digital Genome, Large Language Models, and Medical Knowledge-Driven Digital Assistant †

W. Patrick Kelly [1], Francesco Coccaro [1] and Rao Mikkilineni [2,*]

[1] Barowsky School of Business, Dominican University of California, San Rafael, CA 94901, USA; william.kelly@students.dominican.edu (W.P.K.); francesco.coccaro@students.dominican.edu (F.C.)
[2] Ageno School of Business, Golden Gate University, San Francisco, CA 94105, USA
* Correspondence: rmikkilineni@ggu.edu
† Presented at the Workshop on AI and People, IS4SI Summit 2023, Beijing, China, 14–16 August 2023.

Abstract: Recent advances in large language models, our understanding of the general theory of information, and the availability of new approaches to building self-regulating domain-specific software are driving the creation of next-generation knowledge-driven digital assistants to improve the efficiency, resiliency, and scalability of various business processes while fulfilling the functional requirements addressing a specific business problem. Here, we describe the implementation of a medical-knowledge-based digital assistant that uses medical knowledge derived from various sources including the large language models and assists the early medical diagnosis process by reducing the knowledge gap between the patient and medical professionals involved in the process.

Keywords: digital genome; general theory of information; knowledge-based digital assistant; artificial intelligence; large language models

check for updates

Citation: Kelly, W.P.; Coccaro, F.; Mikkilineni, R. General Theory of Information, Digital Genome, Large Language Models, and Medical Knowledge-Driven Digital Assistant. *Comput. Sci. Math. Forum* **2023**, *8*, 70. https://doi.org/10.3390/cmsf2023008070

Academic Editors: Zhongzhi Shi and Wolfgang Hofkirvhner

Published: 11 August 2023

Copyright: © 2023 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The new science of information processing structures [1–9] derived from our knowledge of genomics, neuroscience, and the general theory of information tells us that human intelligence stems from the knowledge encapsulated in the genome of biological systems and is transmitted from the survivor to the successor.

The genome provides the operational knowledge to implement biological life processes using 30+ trillion cells which are autonomous, collaborate in groups, have the ability to process information, create knowledge, and use replication, and metabolism (conversion of matter and energy). The specification of "life" processes provides full knowledge of functional requirements and their fulfillment, non-functional requirements and their fulfillment, along with best practices passed on from the survivors to their successors including how to fight a virus that affected their ancestors in the past.

As Itai Yanai and Martin Lercher point out in their book "The Society of Genes" [10].

"The single fertilized egg cell develops into a full human being is achieved without a construction manager or architect. The responsibility for the necessary close coordination is shared among the cells as they come into being. It is as though each brick, wire, and pipe in a building knows the entire structure and consults with the neighboring bricks to decide where to place itself".

The genome enables both autopoietic and cognitive behaviors using cell replication and metabolism (using the processes of energy and matter transformations), the mental action or process of acquiring knowledge and understanding through thought, experience, and the senses. It provides a unique capability for the biological systems to make sense of what they are observing fast enough to do something about it while they are still observing it.

The general theory of information (GTI) [1–3] articulated by Prof. Mark Burgin relates the state and the dynamics of physical structures in the material world to the mental structures of biological systems which convert the ontological information carried by the material structures into the epistemic information received by their cognitive apparatuses. According to GTI [1,4], epistemic information has the potential to create or update knowledge in mental structures just as energy has the potential to create or modify material structures. GTI relates the information carried by the material structures to a knowledge representation in the mental structures using the ideal structures in the form of fundamental triads or named sets. The fundamental triads or named sets [3] are used to derive a schema [5,9,11], and operations for knowledge representation in the form of entities with various attributes, their relationships, and their dynamics using the behaviors when events change any of the attributes.

Burgin Mikkilineni's (BM) thesis [6,12] provides a model for the information processing architecture of the mental structures using the schema of knowledge structures and operations executed by a structural machine. Structural machines [4,13] are a new model of computation that can process information in the form of structures, not just sequences of symbols. Structures are more general and expressive than symbols and can represent complex objects such as graphs, networks, images, etc. Structural machines can also manipulate structures using operations such as insertion, deletion, substitution, etc. It explains the autopoietic and cognitive behaviors of biological systems which function at three levels of information processing. We refer the reader to various references cited in this paper for a detailed description of theory and practice and the reason why structural machines are more powerful in information processing than the current symbolic and sub-symbolic computing structures [14]. Suffice it to say that structural machines and the knowledge representation using knowledge structures are a new paradigm in understanding and building new forms of information processing structures that are autopoietic and cognitive [15].

In this paper, we use the structural machines and the knowledge structures to design and implement a digital genome that specifies the life processes of a software system that uses the medical knowledge of an early diagnosis process. Just as the genome specifies and enables the execution of the "life processes" of a biological system, the digital genome enables the specification and implementation of the "life processes" of a software system that defines a process with a purpose and means to execute it [15]. In Section 2, we describe the digital genome design derived from the general theory of information. In Section 3, we describe an implementation of a medical-knowledge-intensive digital assistant with a mission to provide early diagnosis of disease process automation and advise to reduce the knowledge gap between the patient and healthcare providers. The digital assistant is implemented using the genome which specifies functional and non-functional requirements and the best practices based on history and the latest and verified medical knowledge.

2. The Digital Genome

The digital genome specifies the knowledge to execute various tasks that implement functional requirements, non-functional requirements, and the best practices to assure that the process objectives are achieved. Functional requirements deal with domain-specific process execution, components of knowledge acquisition, model creation, the evolution of knowledge structures, analysis of events and actions, etc. Nonfunctional requirements deal with deploying, operating, and managing the compute resources for the various application components, where to set up IaaS and PaaS (containers, databases, etc.), deploy auto-failover, auto-scaling, and live-migration policies. Best practices deal with decision-making by looking at the evolution of the system and predicting risk and suggesting remedies for both maintaining functional and non-functional requirement fulfillment. Figure 1 shows how the digital genome is designed using information from various sources.

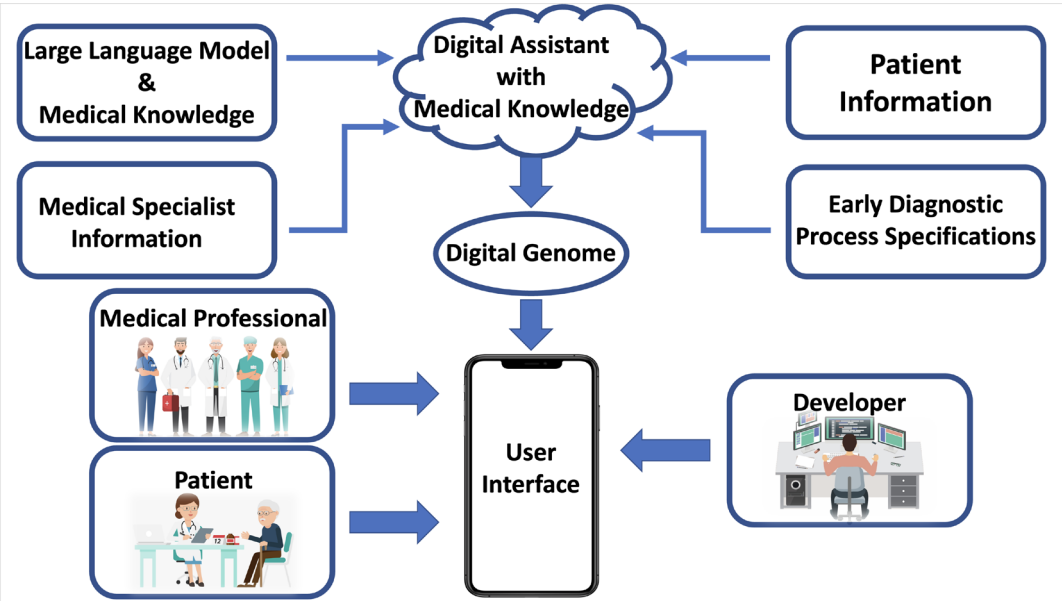


Figure 1. Process to prepare the domain-specific genome.

The digital genome specifies the early diagnosis process derived from various sources of information:

1. Patient information derived from various sources;
2. Medical knowledge about symptoms, diseases, and medical professionals who treat various diseases;
3. The early diagnosis process involves the patient and the healthcare providers. The process involves the early detection of potential diseases causing the symptoms of a patient and providing a detailed analysis of medical information relevant to assist the patient and the healthcare providers in reaching a treatment plan.

The goal is to reduce the knowledge gap between the patient and the healthcare providers by providing accurate and timely information derived from multiple sources. Figure 2 shows the various sources of knowledge.

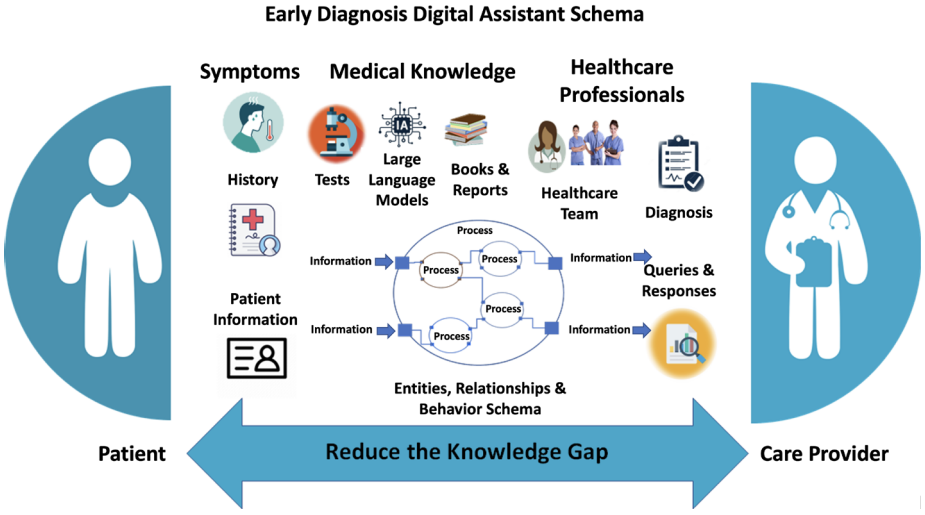


Figure 2. The digital genome schema contains various actors, entities, their relationships, and their behavior when events change the circumstances.

The schema describes the actors (patients and healthcare providers), entities (symptoms, diseases, medical specialties that deal with various diseases), their relationships, and

process flows that define the behaviors when events change the state of the system in the form of knowledge structures [5].

3. Structural Machine Implementation of a Medical-Knowledge-Based Digital Assistant Using the Digital Genome

Figure 3 shows the structural machine implementation of the digital genome consisting of a multi-layer network implementing the digital genome where each node executes a process and shares information with other nodes to execute the autopoietic and cognitive behaviors specified in the digital genome.

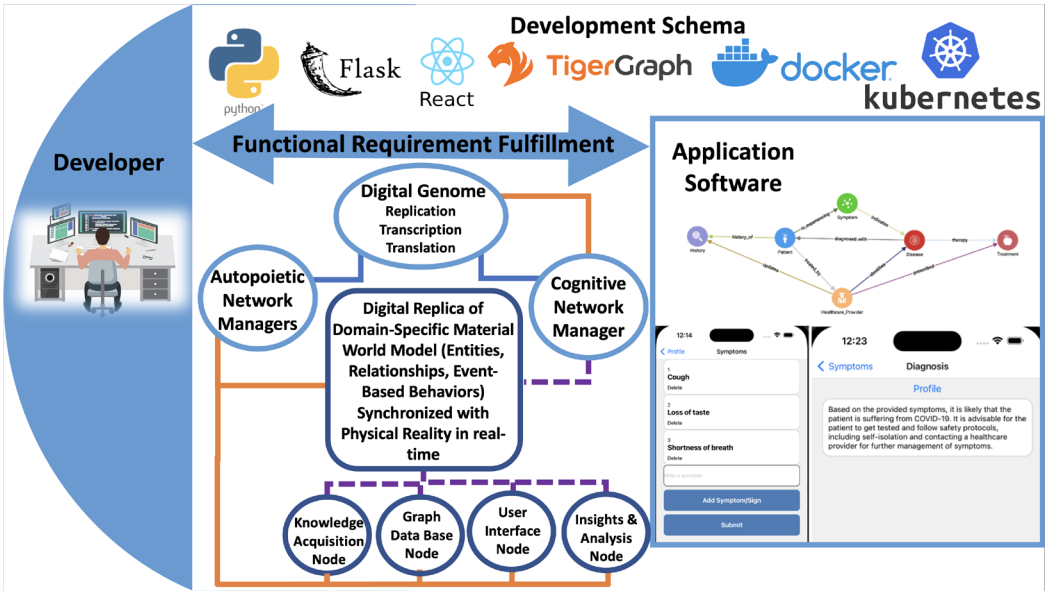


Figure 3. The digital genome implementation using various programming tools.

The autopoietic manager is implemented using Kubernetes and containers. The cognitive network manager sets up connections between various functional modules sharing information to execute the global workflow. A digital replica of the system state and its evolution is stored in a graph database as long-term memory. Various autonomous functional nodes provide knowledge acquisition, user interaction, and various reasoning functions including verification and validation of knowledge. A demonstration of the digital assistant is shown in this video https://youtu.be/vV_38uH0NUk (accessed on 30 July 2023).

4. Conclusions

This paper is the first attempt to demonstrate an implementation of the role of a schema and operations for the knowledge structures that model autopoietic and cognitive behaviors using replication and metabolism. The knowledge structure is like a cell with the knowledge to find and use computing resources to execute the defined processes. The digital genome specifies and executes a society of knowledge structures in the form of a multi-layered knowledge network, where nodes wired together fire together to execute local, clustered, and global autopoietic and cognitive behaviors. The resulting super-symbolic structure integrates the current generation's symbolic and sub-symbolic structures (see Burgin, Mikkilineni [12]) to enhance information processing capabilities.

The digital genome specifies a goal-directed behavior executed by hierarchically nested chains composed of action sequences and checkpoints that glue them. As Riegler [16] points out, the checkpoint character of cognition defines decision making and other cognitive acts as being based upon internal states rather than external states of affairs. For an observer, from a third-person perspective, cognitive systems seem to make the proper decisions in response to the challenges of the environment. However, these checkpoints are a result of the

designer of the digital genome and caution must be exercised in taking them at face value. It is therefore important to use digital assistants to reduce the knowledge gap between various actors participating in decision making instead of giving them autonomous control.

Author Contributions: Conceptualization, R.M.; methodology, R.M., W.P.K. and F.C.; software, W.P.K. and F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not available.

Acknowledgments: The authors are grateful to the Dominican University of California and its graduate program that made this research possible.

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

References

1. Burgin, M. *Theory of Information: Fundamentality, Diversity, and Unification*; World Scientific: Singapore, 2010.
2. Burgin, M. *Theory of Knowledge: Structures and Processes*; World Scientific: Singapore, 2016.
3. Burgin, M.S. *Theory of Named Sets*; Mathematics Research Developments Series; Nova Science Pub Inc.: Hauppauge, NY, USA, 2011; ISBN 978-1-61122-788-8.
4. Mikkilineni, R. The Science of Information Processing Structures and the Design of a New Class of Distributed Computing Structures. *Proceedings* **2022**, *81*, 53. [CrossRef]
5. Burgin, M.; Mikkilineni, R. From Data Processing to Knowledge Processing: Working with Operational Schemas by Autopoietic Machines. *Big Data Cogn. Comput.* **2021**, *5*, 13. [CrossRef]
6. Mikkilineni, R. A New Class of Autopoietic and Cognitive Machines. *Information* **2022**, *13*, 24. [CrossRef]
7. Burgin, M.; Mikkilineni, R.; Phalke, V. Autopoietic Computing Systems and Triadic Automata: The Theory and Practice. *Adv. Comput. Commun.* **2020**, *1*, 16–35. [CrossRef]
8. Burgin, M. Triadic Automata and Machines as Information Transformers. *Information* **2020**, *11*, 102. [CrossRef]
9. Burgin, M.; Mikkilineni, R. General Theory of Information Paves the Way to a Secure, Service-Oriented Internet Connecting People, Things, and Businesses. In Proceedings of the 12th International Congress on Advanced Applied Informatics (IIAI-AAI), Kanazawa, Japan, 2–8 July2022; pp. 144–149. [CrossRef]
10. Yanai, I.; Martin, L. *The Society of Genes*; Harvard University Press: Cambridge, MA, USA, 2016.
11. Mikkilineni, R. Infusing Autopoietic and Cognitive Behaviors into Digital Automata to Improve Their Sentience, Resilience, and Intelligence. *Big Data Cogn. Comput.* **2022**, *6*, 7. [CrossRef]
12. Burgin, M.; Mikkilineni, R. On the Autopoietic and Cognitive Behavior. EasyChair Preprint No. 6261, Version 2. 2021. Available online: https://easychair.org/publications/preprint/tkjk (accessed on 27 April 2023).
13. Mikkilineni, R.; Burgin, M. Structural Machines as Unconventional Knowledge Processors. *Proceedings* **2020**, *47*, 26. [CrossRef]
14. Burgin, M. The Rise and Fall of the Church-Turing Thesis. Manuscript. Available online: http://arxiv.org/ftp/cs/papers/0207/0207055.pdf (accessed on 27 April 2023).
15. Burgin, M.; Mikkilineni, R. Information Theoretical Principles of Software Development. 2022. Available online: https://easychair.org/publications/preprint_open/jnMd (accessed on 8 June 2023).
16. Riegler, A. Superstition in the Machine. In *Anticipatory Behavior in Adaptive Learning Systems: From Brains to Individual and Social Behavior*; Butz, M.V., Sigaud, O., Pezzulo, G., Baldassarre, G., Eds.; Springer-Verlag: Heidelberg, Germany, 2007; pp. 57–72. Available online: https://doi.org/10.1007/978-3-540-74262-3_4 (accessed on 6 June 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.